HAL
open science

Dévoilement de soi et réponses du consommateur face à une sollicitation de ses données personnelles : une application aux formulaires sur Internet

Caroline Lancelot Miltgen

To cite this version:

Caroline Lancelot Miltgen. Dévoilement de soi et réponses du consommateur face à une sollicitation de ses données personnelles : une application aux formulaires sur Internet. Gestion et management. Université Paris Dauphine, 2006. Français. NNT : . tel-01102497

HAL Id: tel-01102497

https://audencia.hal.science/tel-01102497v1

Submitted on 12 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS DAUPHINE

UFR SCIENCES DES ORGANISATIONS
Centre de Recherche DMSP

Nouveau Doctorat en Sciences de Gestion
(Arrêté du 25 avril 2002)

Dévoilement de soi et réponses du consommateur face à une sollicitation de ses données personnelles : une application aux formulaires sur Internet

THESE
Pour l'obtention du titre de Docteur en Sciences de Gestion

Tome I

Présentée et soutenue publiquement par :

Caroline Lancelot Miltgen

Le 01 Décembre 2006

JURY

Directeur de thèse :	**Monsieur Pierre Volle** Professeur à l'Université Paris Dauphine
Rapporteurs :	**Monsieur Pierre Valette-Florence** Professeur à l'IAE de Grenoble
	Monsieur Joël Brée Professeur à l'IAE de Caen
Suffragants :	**Monsieur Eric Vernette** Professeur à l'Université de Toulouse I
	Madame Sandrine Macé Professeur à l'ESCP-EAP

L'Université n'entend donner aucune
approbation ni improbation aux opinions
émises dans les thèses : ces opinions doivent
être considérées comme propres à leurs auteurs.

REMERCIEMENTS

Rédiger ces remerciements est forcément quelque chose d'agréable. D'une part, parce que c'est l'occasion de rendre hommage à toutes les personnes qui ont contribué, d'une manière ou d'une autre, à la réalisation de ce travail. D'autre part, parce que cela marque la fin d'une période et donc le début d'une autre.
C'est donc avec une grande joie que je m'apprête à remercier ici tous mes « bienfaiteurs ».
Je m'excuse par avance auprès des lecteurs du côté « répétitif » des lignes qui vont suivre. Il me semblait néanmoins important de citer l'ensemble des personnes qui m'ont soutenue, aidée et conseillée au cours de ces quatre années. De plus, je tiens à préciser qu'il n'existe pas, dans mon esprit, d'ordre d'importance entre ces différentes personnes, chacune ayant contribué, selon sa disponibilité et ses compétences, à l'aboutissement de ce travail.

Je tiens tout d'abord à remercier mon directeur de thèse, le professeur Pierre Volle, pour la disponibilité dont il a fait preuve et les conseils qu'il m'a prodigués. Je précise aussi qu'il est, en partie, à l'origine de ma décision de faire une thèse et pour cela aussi je l'en remercie. Je fonde l'espoir que cette collaboration se révèlera fructueuse à l'avenir.

J'en profite aussi pour remercier l'ensemble des membres du jury d'avoir accepté d'en faire partie et d'avoir ainsi pris de leur temps pour lire cette modeste contribution de 500 pages.

Pour poursuivre sur cette lignée, je tiens également à dire merci à l'ensemble des professeurs que j'ai eu l'occasion de côtoyer au cours de ces quatre années, que ce soit lors des séminaires du DMSP ou lors de manifestations plus spécifiques comme les congrès de l'Association Française du Marketing. Je pense plus particulièrement aux professeurs Pierre Desmet, Bernard Pras, Raymond-Alain Thiétart, Gilles Laurent, Denis Guiot, Denis Darpy … mais j'en oublie surement beaucoup d'autres. Que ceux-là veuillent bien m'excuser !
Une mention particulière à Claire Gauzente avec qui mes rapports furent autant professionnels qu'amicaux. J'espère que notre collaboration conduira à de nombreuses publications.

Je ne peux me résoudre à tourner cette page « académique » sans remercier du fond du cœur l'équipe administrative du DMSP, en la personne de Muriel Urier, Chantal Charlier et Patricia Lenfant, pour leur soutien sans faille, leur disponibilité et leur générosité.

Enfin, je termine la liste des dauphinois en adressant mes salutations à l'ensemble de mes pairs, anciens doctorants ou futurs docteurs, pour les discussions que nous avons pu avoir et les encouragements qu'ils m'ont toujours prodigués.

Dans un registre plus professionnel mais tout aussi décisif, j'adresse également mes sincères remerciements à Thibaut Munier, Yseulis Costes, Stéphanie Gautier et Anne Guérin de la Société Millemercis, pour m'avoir fait confiance et m'avoir donné la chance de collaborer avec eux. J'espère avoir été à la hauteur de leurs attentes et je crois pouvoir dire que l'étude que j'ai pu mener grâce à eux contribue, en grande partie, à la richesse et à l'originalité de ce travail. Ce partenariat a déjà été l'occasion d'une publication qui sera suivie, je l'espère, de nombreuses autres.

Je remercie également l'ensemble des professionnels avec qui j'ai eu l'occasion de collaborer dans le cadre de ma thèse. Je pense plus particulièrement à Arnaud Belleil qui a su me manifester son enthousiasme et qui s'est toujours montré disponible pour me conseiller. Un grand merci aussi aux membres de la FING, et en particulier à Daniel Kaplan, pour avoir accepté de me confier une étude sur la gestion des données personnelles, laquelle a déjà fait l'objet d'un chapitre de ma thèse, d'une communication académique et d'une soirée-débat auprès de professionnels de l'informatique et de l'Internet.

Même s'ils ne le sauront peut-être jamais, je tiens aussi à dire merci à tous les répondants que j'ai interrogés à l'occasion de ma thèse, pour s'être volontiers prêtés à cet exercice.

Un merci très particulier aux relecteurs, pour s'être pliés à cette tâche ingrate mais néanmoins essentielle. Parents, beaux-parents et amis, angevins et messins, je vous tire mon chapeau !

Enfin, mais peut-être devrais-je dire surtout, un grand merci à François pour avoir toujours été présent à mes côtés et avoir su me supporter, dans tous les sens du terme. Des excuses à mon petit Paul pour n'avoir pas été présente aussi souvent que je l'aurais souhaité, et à Louis pour n'avoir pas pu (ou su) profité pleinement de ces neuf derniers mois.
Un clin d'œil aussi au reste de ma famille pour avoir supporté les conséquences de mon engagement, souvent total, et parfois excessif, dans ce travail.

Une page est désormais tournée et sachez que je reviens vers vous avec plaisir !

A François,

Paul

Louis

Et les autres …

TABLE DES MATIÈRES

Introduction générale

1 CONTEXTE DE LA RECHERCHE 1
2 INTERET DU SUJET DE RECHERCHE 4
3 DELIMITATION DU SUJET ET CHOIX PREALABLES 6
4 OBJECTIF, PROBLEMATIQUE ET QUESTIONS DE RECHERCHE 7
5 CONTRIBUTIONS ATTENDUES DE LA RECHERCHE 9
6 POSITIONNEMENT DE LA RECHERCHE 12
7 PLAN DE LA RECHERCHE 12

Chapitre 1
Les enjeux liés à la collecte de données clients

INTRODUCTION DU CHAPITRE 1 **19**

1 LES FORMES DE COLLECTE DE DONNEES CLIENTS 20
1.1 Les types de données collectées *20*
1.2 Les modes de collecte de l'information client *22*
1.3 Les spécificités de la collecte de données en ligne *24*
2 LA COLLECTE DE DONNEES CLIENTS : UN ENJEU STRATEGIQUE 26
2.1 Le développement du marketing des bases de données *26*
2.2 Du marketing des bases de données au marketing relationnel *30*
2.3 Marketing relationnel et respect de la vie privée *31*
3 COLLECTE DE DONNEES CLIENTS ET RESPECT DE LA VIE PRIVEE : UN ENJEU ETHIQUE 32
3.1 Le respect de la vie privée : une attention croissante sur le sujet *32*
3.2 Une approche par les théories éthiques *32*
3.3 Propriété des données et respect de la vie privée : un conflit d'intérêt *34*
3.4 Les spécificités de l'enjeu éthique sur Internet *35*

CONCLUSION DU CHAPITRE 1 **38**

Chapitre 2
Préambule sur la notion de vie privée

INTRODUCTION DU CHAPITRE 2 **41**

1 DEFINITION ET CARACTERISTIQUES DE LA PROTECTION DE LA VIE PRIVEE 42
1.1 La protection de la vie privée : un concept flou *42*
1.2 Les dimensions de la protection de la vie privée dans la littérature *43*
1.3 Les principales fonctions de la protection de la vie privée *48*
1.4 La distinction avec d'autres concepts proches et/ou liés *50*
2 LA PROTECTION DE LA VIE PRIVEE : UN BESOIN POLYMORPHE 52
2.1 Un besoin différent selon les cultures et les périodes *52*
2.2 Un besoin individuel *54*
2.3 La protection de la vie privée vue comme un processus d'ajustement *55*
3 LA VIE PRIVEE EN TANT QUE PROTECTION DE DONNEES A CARACTERE PERSONNEL 56
3.1 La définition du droit à la protection des données personnelles *56*
3.2 Les dimensions du droit à la protection des données personnelles *57*
3.3 Approche historique et juridique du droit à la protection des données *60*
3.3.1 Approche historique du droit à la protection des données 60
3.3.2 Approche juridique du droit à la protection des données 62

CONCLUSION DU CHAPITRE 2 **66**

Chapitre 3
Dévoilement de soi et divulgation de données personnelles

INTRODUCTION DU CHAPITRE 3 **69**

1 DEFINITION ET ATTRIBUTS DU DEVOILEMENT DE SOI 70
1.1 Définition du dévoilement de soi *70*
1.2 Les attributs du dévoilement de soi *70*
1.3 Les conséquences du dévoilement de soi *71*
2 LE PROCESSUS DE DEVOILEMENT DE SOI 72
2.1 Dévoilement et échange *72*
2.2 Un processus conduit par la motivation *73*
3 LES FACTEURS D'INFLUENCE DU DEVOILEMENT DE SOI 75
3.1 L'influence des facteurs culturels *75*
3.2 L'influence des facteurs individuels *76*
3.2.1 Le dévoilement de soi selon le sexe 76
3.2.2 Dévoilement de soi et personnalité 77
3.3 L'influence des facteurs situationnels *78*
4 DEVOILEMENT DE SOI ET RELATION A L'AUTRE 80
4.1 Dévoilement selon l'avancement de la relation *81*
4.2 Dévoilement et appréciation de l'autre *82*
4.3 Dévoilement et réciprocité *84*

CONCLUSION DU CHAPITRE 3 **86**

Chapitre 4
Sollicitation de données personnelles et réactions des consommateurs

INTRODUCTION DU CHAPITRE 4 **89**

1 SOLLICITATION DE DONNEES PERSONNELLES ET REACTIONS DES CONSOMMATEURS 90
1.1 La mesure des réactions comportementales des consommateurs *90*
1.2 L'impact de la préoccupation sur le comportement des consommateurs *93*
1.3 L'écart entre les convictions personnelles et le comportement réel *95*
2 LES ANTECEDENTS AUX PREOCCUPATIONS ET AUX REACTIONS DES CONSOMMATEURS 98
2.1 Un modèle à 3 entités *98*
2.2 Les antécédents environnementaux *101*
2.3 Les antécédents individuels *102*
2.3.1 Les variables d'ordre sociodémographique 104
2.3.2 La connaissance, l'expérience et l'attitude des consommateurs 108
2.3.3 Les variables d'ordre psychologique 112
2.3.4 Les variables d'ordre idéologique 113
2.4 Les antécédents situationnels *115*
2.4.1 L'importance de la situation en général 115
2.4.2 Les facteurs liés à la politique de Respect de la Vie Privée 119
2.4.3 Les facteurs liés au demandeur 121
2.4.4 Les facteurs liés au dispositif de collecte 123
2.4.5 Les facteurs liés aux « circonstances de collecte » 127
3 LA SOLLICITATION DE DONNEES VUE COMME UN CONTRAT SOCIAL 131
3.1 La théorie du contrat social *131*
3.2 La théorie du contrat social appliquée à la sollicitation de données *133*

CONCLUSION DU CHAPITRE 4 **136**

Chapitre 5
Processus de réponse face à une sollicitation de données personnelles (étude 1)

INTRODUCTION DU CHAPITRE 5 **143**

1 METHODOLOGIE DE L'ETUDE QUALITATIVE 144
1.1 Les objectifs de l'étude *144*
1.2 La méthode de collecte des informations *144*
1.2.1 Le recours à des entretiens semi-directifs 144
1.2.2 Le guide d'entretien 145
1.2.3 Le déroulement de l'entretien 146
1.3 Le recrutement et le profil des participants *147*
1.4 Le choix de la méthode d'analyse : l'analyse de contenu *149*
1.4.1 Justification et principes de l'analyse de contenu 149
1.4.2 Les étapes de l'analyse 150
1.5 Validité et fiabilité de l'étude qualitative *153*
1.5.1 La validité de l'étude 153
1.5.2 La fiabilité de l'étude 154
2 RESULTATS DE L'ETUDE QUALITATIVE 155
2.1 Perceptions et définitions des principaux concepts en jeu *156*
2.2 Les spécificités perçues d'Internet *156*
2.3 L'attitude face au traitement des données personnelles *157*
2.4 Les facteurs influençant le comportement de réponse *158*
2.4.1 L'influence des facteurs environnementaux 158
2.4.2 L'influence des facteurs individuels 159
2.4.3 L'influence des facteurs situationnels 161
2.5 L'évaluation de la sollicitation *163*
2.5.1 La confidentialité perçue des données sollicitées 163
2.5.2 La sensibilité perçue des données sollicitées 164
2.5.3 La pertinence perçue des données sollicitées 164
2.5.4 Les coûts et bénéfices perçus de la divulgation 165
2.6 Les comportements de réponse identifiés *167*
3 MISE EN PERSPECTIVE DES RESULTATS 168
3.1 Les conclusions de l'analyse *168*
3.1.1 Présentation d'un modèle conceptuel provisoire 168
3.1.2 Principaux enseignements de l'étude 170
3.2 Comparaison avec les recherches antérieures *172*

CONCLUSION DU CHAPITRE 5 **174**

Chapitre 6
La gestion des données personnelles chez le consommateur-internaute (étude 2)

INTRODUCTION DU CHAPITRE 6 **177**

1 PRESENTATION DE L'ETUDE 178
1.1 Les objectifs de l'étude *178*
1.2 Présentation du partenariat avec la FING *178*
2 REVUE DE LITTERATURE SUR LES TYPOLOGIES DE PREOCCUPATION POUR LE RVP 180
2.1 Les typologies existantes *180*
2.1.1 La typologie d'origine (Westin, 1996) 180
2.1.2 Les autres typologies existantes 181
2.2 Limites des typologies existantes *183*
2.3 Les variables d'influence *185*
3 METHODOLOGIE DE L'ETUDE 186
3.1 Le recueil des données *186*
3.1.1 Développement du questionnaire 186
3.1.2 Description de l'échantillon 190
3.2 La méthode d'analyse *192*
3.2.1 Le choix des variables à introduire 192
3.2.2 Le choix de la méthode de classification 193

4 RESULTATS DE L'ETUDE 196
4.1 La typologie obtenue *196*
4.2 La description des classes *200*
5 MISE EN PERSPECTIVE DES RESULTATS 205
5.1 Les limites de l'étude *205*
5.2 Les enseignements de l'étude *206*

CONCLUSION DU CHAPITRE 6 **209**

Chapitre 7
Formulation du cadre conceptuel général de la recherche

INTRODUCTION DU CHAPITRE 7 **213**

1 JUSTIFICATIONS ET CHOIX PREALABLES 214
1.1 Le contexte d'analyse *214*
1.2 La phase d'analyse *215*
1.3 Le type de processus considéré : heuristique et/ou systématique *215*
1.4 Un processus guidé par les composantes cognitives et/ou affectives *217*
2 ANCRAGE THEORIQUE 217
2.1 Les théories de l'échange et de l'échange social *218*
2.1.1 Les principes de la théorie de l'échange et de l'échange social 218
2.1.2 La théorie de l'échange social appliquée au dévoilement de soi 219
2.2 La théorie de la motivation ou de l'espérance attendue *221*
2.3 La théorie de l'action raisonnée *223*
3 PRESENTATION DU MODELE CONCEPTUEL GENERAL DE LA RECHERCHE 227
4 FORMULATION DES PROPOSITIONS ET HYPOTHESES DE RECHERCHE 233
4.1 L'influence de l'exposition à la sollicitation *234*
4.2 Les antécédents de l'attitude *236*
4.2.1 L'influence des variables perceptuelles 236
4.2.2 L'influence de la préoccupation pour le RVP 241
4.3 Les conséquences de l'attitude *243*
4.3.1 L'effet de l'attitude sur l'intention de répondre 244
4.3.2 L'effet de l'attitude sur l'intention de mentir 244
4.4 Les conséquences sur l'intention de poursuivre la relation *245*
4.4.1 La théorie de l'engagement 245
4.4.2 L'effet de l'intention de répondre 249
4.4.3 L'effet du comportement de réponse 250

CONCLUSION DU CHAPITRE 7 **252**

Chapitre 8
Méthodologie de la recherche

INTRODUCTION DU CHAPITRE 8 **255**

1 LE CHOIX D'UN DESIGN EXPERIMENTAL 256
1.1 La justification du choix d'un design expérimental *256*
1.2 Le choix des facteurs situationnels étudiés *257*
2 LES REFLEXIONS PREALABLES A LA CONSTITUTION DU PLAN EXPERIMENTAL 261
2.1 Les objectifs du plan expérimental *261*
2.2 Les principaux biais à contrôler *263*
2.2.1 Les biais d'auto-sélection et de non réponse 263
2.2.2 Le biais de couverture 265
2.2.3 Le biais de désirabilité sociale et le biais du répondant 266
2.3 Le choix des échantillons interrogés *269*
3 PRESENTATION DU MODELE A TESTER ET DU PLAN EXPERIMENTAL 272
3.1 Le modèle à tester *272*
3.2 Le plan expérimental *274*

4 LE CONTROLE DES VARIABLES EXTERNES 279
4.1 L'importance de contrôler *279*
4.2 Les variables à contrôler *279*
4.2.1 Le niveau d'implication 280
4.2.2 Les variables sociodémographiques 280
4.2.3 Les variables expérientielles 281
4.2.4 Le comportement passé 282
4.2.5 La qualité perçue de la relation 283
4.3 Les modes de contrôle *284*
5 LES QUESTIONS DE VALIDITE 285
5.1 La validité interne de la recherche *286*
5.2 La validité externe de la recherche *290*

CONCLUSION DU CHAPITRE 8 **291**

Chapitre 9
Choix et développement des instruments de mesure

INTRODUCTION DU CHAPITRE 9 **295**

1 LA DEMARCHE DE DEVELOPPEMENT ET DE VALIDATION DES ECHELLES DE MESURE 296
1.1 Les critères de sélection d'un instrument de mesure *296*
1.1.1 La qualité méthodologique des instruments de mesure 296
1.1.2 La qualité opérationnelle des instruments de mesure 297
1.2 Développement et validation d'un instrument de mesure *298*
1.2.1 Spécification du domaine du construit et génération des items 299
1.2.2 Analyse factorielle exploratoire et purification des mesures 301
1.2.3 Analyse confirmatoire et propriétés psychométriques 302
1.3 Typologie des construits de la recherche et type d'analyse conduite *308*
2 LE DEVELOPPEMENT DES INSTRUMENTS DE MESURE NECESSAIRES A L'EXPERIENCE 1 310
2.1 Spécification du domaine du construit et analyse exploratoire *311*
2.1.1 Les échelles de confidentialité, sensibilité et pertinence 311
2.1.2 L'échelle d'attitude face à la communication des données 316
2.1.3 L'échelle de préoccupation pour le respect de la vie privée 317
2.1.4 L'échelle d'implication envers la catégorie de produits 320
2.2 L'analyse confirmatoire *322*
2.2.1 La réplication de l'analyse exploratoire 322
2.2.2 L'analyse confirmatoire et la qualité des échelles 324
3 LE DEVELOPPEMENT DES INSTRUMENTS DE MESURE NECESSAIRES A L'EXPERIENCE 2 329
3.1 Spécification du domaine du construit et analyse exploratoire *330*
3.1.1 L'échelle de valeur perçue 330
3.1.2 L'échelle d'intention de poursuivre la relation 334
3.2 L'analyse confirmatoire *336*
3.2.1 La réplication de l'analyse exploratoire 337
3.2.2 L'analyse confirmatoire et la qualité des échelles 340

CONCLUSION DU CHAPITRE 9 **346**

Chapitre 10
Expliquer l'attitude face à la communication de données personnelles (étude 3)

INTRODUCTION DU CHAPITRE 10 **353**

1 OBJECTIFS ET APPORTS DE L'EXPERIENCE 1 355
2 LE DESIGN EXPERIMENTAL 356
2.1 Les facteurs manipulés *357*
2.1.1 Type de manipulation et niveau des facteurs manipulés 357
2.1.2 Facteurs manipulés et opérationnalisation 359
2.2 Les variables contrôlées *370*
2.2.1 Le type de variable à contrôler 370
2.2.2 Le type de contrôle 371

3 PRESENTATION DU MODELE A TESTER ET DES HYPOTHESES DE RECHERCHE 372
3.1 Présentation du modèle à tester *372*
3.2 Formulation des hypothèses de recherche *374*
3.2.1 L'influence des mentions de RVP sur la confidentialité et la sensibilité perçues 374
3.2.2 L'influence de la familiarité sur la confidentialité et la pertinence perçues 376
3.2.3 L'influence de la quantité sur la sensibilité et la pertinence perçues 377
4 RECUEIL DES DONNEES ET CARACTERISTIQUES DES REPONDANTS 378
4.1 L'instrument de collecte et le recueil des données *378*
4.2 Les caractéristiques de l'échantillon *380*
5 L'ANALYSE DES DONNEES 382
5.1 La validité interne du plan d'expérience *382*
5.1.1 Le contrôle des variables exogènes 382
5.1.2 L'égalité des traitements 382
5.1.3 L'homogénéité des variances 383
5.1.4 La normalité des variables 383
5.1.5 L'indépendance des observations 383
5.1.6 L'orthogonalité des variables manipulées 384
5.2 La vérification des manipulations expérimentales *385*
5.3 La qualité psychométrique des instruments de mesure *388*
5.4 Le plan de traitement *389*
6 PRESENTATION DES RESULTATS 391
6.1 L'impact de l'exposition à la sollicitation sur l'évaluation de celle-ci *391*
6.1.1 L'impact des mentions de RVP 394
6.1.2 L'impact de la Familiarité 395
6.1.3 L'impact de la Quantité de données demandées 397
6.1.4 L'impact des variables exogènes 398
6.2 L'impact de l'évaluation de la sollicitation et de la préoccupation pour le RVP sur l'attitude face à la communication des données *400*
7 RECAPITULATIF DES RESULTATS ET LIMITES INHERENTES A CETTE EXPERIENCE 403
7.1 Récapitulatif des résultats et commentaires *403*
7.2 Les limites de l'expérience 1 *406*

CONCLUSION DU CHAPITRE 10 **408**

Chapitre 11
Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)

INTRODUCTION DU CHAPITRE 11 **411**

1 OBJECTIFS ET APPORTS DE L'EXPERIENCE 2 412
2 LE DESIGN EXPERIMENTAL 414
2.1 Les facteurs manipulés *415*
2.1.1 Opérationnalisation du facteur « Quantité » 417
2.1.2 Opérationnalisation du facteur « Contexte de sollicitation » 418
2.2 Les variables contrôlées *423*
2.2.1 Le type de variable à contrôler 423
2.2.2 Le type de contrôle 423
3 PRESENTATION DU MODELE A TESTER ET DES HYPOTHESES DE RECHERCHE 425
3.1 Présentation du modèle à tester *425*
3.2 Formulation des hypothèses de recherche *427*
3.2.1 L'influence de la quantité sur la valeur perçue à répondre 427
3.2.2 L'influence du contexte sur la valeur perçue à répondre 428
4 RECUEIL DES DONNEES ET CARACTERISTIQUES DES REPONDANTS 429
4.1 L'instrument de collecte et le recueil des données *429*
4.2 Les caractéristiques de l'échantillon *431*
5 L'ANALYSE DES DONNEES 434
5.1 La validité interne du plan d'expérience *434*
5.1.1 Le contrôle des variables exogènes 434
5.1.2 L'égalité des traitements 434
5.1.3 L'homogénéité des variances 435
5.1.4 La normalité des variables 436
5.1.5 L'indépendance des observations 436
5.1.6 L'orthogonalité des variables manipulées 437

5.2 La vérification des manipulations expérimentales *437*
5.3 La qualité psychométrique des instruments de mesure *439*
5.4 Le plan de traitement *441*
6 PRESENTATION DES RESULTATS 444
6.1 L'impact de l'exposition à la sollicitation sur l'évaluation de celle-ci *445*
6.1.1 L'impact de la Quantité de données demandées 446
6.1.2 L'impact du Contexte de sollicitation 447
6.1.3 L'impact des variables exogènes 450
6.2 L'impact de l'évaluation de la sollicitation et de la préoccupation pour le RVP *450*
sur l'attitude face à la communication des données *450*
6.3 L'impact de l'attitude sur l'intention de répondre et de mentir *452*
6.3.1 L'impact de l'attitude sur l'intention de répondre 453
6.3.2 L'impact de l'attitude sur l'intention de mentir 454
6.4 L'impact de l'intention de répondre sur celle de poursuivre la relation *457*
6.5 Le test des effets médiateurs de l'attitude et de l'intention de répondre *459*
7 RECAPITULATIF DES RESULTATS ET LIMITES INHERENTES A CETTE EXPERIENCE 462
7.1 Récapitulatif des résultats et commentaires *462*
7.2 Les limites de l'expérience 2 *464*

CONCLUSION DU CHAPITRE 11 **467**

Chapitre 12
Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

INTRODUCTION DU CHAPITRE 12 **471**

1 OBJECTIFS ET APPORTS DE L'EXPERIENCE 3 472
2 LE DESIGN EXPERIMENTAL 473
2.1 Les facteurs manipulés *473*
2.2 Les variables contrôlées *476*
2.2.1 Le type de variables à contrôler 476
2.2.2 Le type de contrôle 477
2.3 La mise en œuvre du design expérimental *479*
2.3.1 La phase 1 de l'expérimentation (dite de « requalification ») 479
2.3.2 La phase 2 de l'expérimentation : le post-test 484
3 PRESENTATION DU MODELE A TESTER ET DES HYPOTHESES DE RECHERCHE 488
3.1 Présentation du modèle à tester *488*
3.2 Formulation des hypothèses de recherche *491*
3.2.1 L'influence de l'exposition à la sollicitation sur la réponse 491
3.2.2 L'influence de la préoccupation pour le RVP sur la réponse 495
3.2.3 L'influence sur l'intention de poursuivre la relation 496
4 LE RECUEIL DES DONNEES ET LES INSTRUMENTS DE COLLECTE 499
4.1 L'instrument de collecte de la phase 1 *499*
4.2 L'instrument de collecte de la phase 2 *500*
4.3 Les données à exploiter *501*
5 CHOIX ET CARACTERISTIQUES DES REPONDANTS 504
5.1 Les caractéristiques de l'échantillon de requalification *504*
5.2 Comparaison des échantillons des phases 1 et 2 *506*
6 L'ANALYSE DES DONNEES 509
6.1 La validité interne du plan d'expérience *509*
6.1.1 Le contrôle des variables exogènes 510
6.1.2 L'indépendance des observations 510
6.2 La qualité psychométrique des instruments de mesure. *511*
6.2.1 La validité et la fiabilité des échelles de mesure multi-items 512
6.2.2 La validité discriminante des instruments de mesure 515
6.3 Le plan de traitement *515*
6.4 L'adéquation des données à des tests statistiques *518*
6.4.1 Les conditions d'utilisation de la régression 518
6.4.2 La validation des hypothèses 519

7 PRESENTATION DES RESULTATS 521
7.1 Préambule concernant la participation à l'opération de requalification *522*
7.2 Résultats concernant la probabilité de communiquer des données *523*
7.3 Résultats concernant la probabilité d'abandonner *526*
7.4 Résultats concernant la probabilité de fournir une donnée sensible *529*
7.5 Résultats concernant la poursuite de la relation avec l'entreprise *532*
7.6 Test du rôle médiateur de la réponse à la sollicitation *535*
8 RECAPITULATIF DES RESULTATS 536

CONCLUSION DU CHAPITRE 12 **541**

Discussion et Mise en perspective des Résultats de la Recherche

INTRODUCTION DE LA DISCUSSION **545**

1 MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE 546
1.1 Le processus de décision face à une sollicitation de données personnelles *546*
1.2 L'impact des facteurs situationnels sur le processus de réponse *553*
1.2.1 L'impact des facteurs sur l'évaluation de la sollicitation 553
1.2.2 L'impact des facteurs sur le reste du processus de réponse 557
1.3 L'impact des variables individuelles sur le processus de réponse *559*
1.3.1 L'impact de la préoccupation pour le respect de la vie privée 560
1.3.2 L'impact des variables exogènes 562
2 LES APPORTS DE LA RECHERCHE 567
2.1 Les apports théoriques de la recherche *567*
2.1.1 Une compréhension élargie de la décision du consommateur 567
2.1.2 Une approche novatrice de l'étude du comportement de réponse 568
2.1.3 La mise en avant de l'effet de la situation de sollicitation 569
2.2 Les apports empiriques de la recherche *571*
2.3 Les implications managériales de la recherche *571*
2.3.1 L'identification des critères de décision des consommateurs 572
2.3.2 L'identification des mécanismes de confiance et de réassurance 572
2.3.3 Les moyens de se prémunir d'une rupture de la relation 574
2.4 Les contributions méthodologiques *579*
2.4.1 Les leçons tirées de la mise en œuvre du plan expérimental 579
2.4.2 Les leçons tirées du développement d'échelles 581
3 LIMITES ET VOIES DE RECHERCHE 584
3.1 Limites et voies de recherche liées à la validité interne *584*
3.1.1 La mise en situation des sujets expérimentaux 584
3.1.2 Les biais liés au sujet de recherche et au dispositif expérimental 586
3.2 Limites et voies de recherche liées à la validité externe *587*
3.2.1 Le choix et l'opérationnalisation des variables étudiées 588
3.2.2 La robustesse et la généralisation des résultats obtenus 592
3.3 Voies de recherche complémentaires *596*
3.3.1 Le recours à des méthodologies plus sophistiquées 596
3.3.2 La prise en compte de nouvelles variables 597
3.3.3 Vers un modèle intégrateur 600
3.3.4 La réalisation d'études complémentaires 603

CONCLUSION DE LA DISCUSSION **606**

Conclusion générale **609**

Bibliographie **i**

LISTE DES TABLEAUX

Chapitre 1

Tableau 1.1 Principales classifications de données personnelles proposées dans la littérature 21
Tableau 1.2 Les modes de collecte de l'information client 25

Chapitre 2

Tableau 2.1 Les trois dimensions majeures du droit à la protection de la vie privée 48
Tableau 2.2 Les principales dimensions du droit à la protection des données dans la littérature 59
Tableau 2.3 Les principes de base de la Directive Européenne de 1995, comparés aux FIP 64

Chapitre 3

Tableau 3.1 Facteurs d'influence du dévoilement de soi identifiés dans la littérature 79

Chapitre 4

Tableau 4.1 Principales variables comportementales identifiées dans la littérature 92
Tableau 4.2 Impact des variables sociodémographiques dans le cadre d'une sollicitation de données personnelles 107
Tableau 4.3 Principaux facteurs situationnels influençant la réaction à une sollicitation 129

Chapitre 5

Tableau 5.1 Profil sociodémographique des répondants 148
Tableau 5.2 Expérience des répondants en matière d'achat à distance et d'Internet 149
Tableau 5.3 Liste des catégories et thèmes retenus 151
Tableau 5.4 Typologie des attitudes face au traitement des données personnelles 157
Tableau 5.5 Typologie des stratégies de réponse 167
Tableau 5.6 Liens entre préoccupation pour le RVP et stratégies de réponses 168

Chapitre 6

Tableau 6.1 Synthèse des principales typologies de consommateurs dans la littérature 184
Tableau 6.2 Description de l'échantillon interrogé pour l'étude 2 190
Tableau 6.3 Chaîne d'agrégation (derniers niveaux de regroupement) 194
Tableau 6.4 Répartition des internautes français en 4 classes 197
Tableau 6.5 Tableau de bord des classes 197
Tableau 6.6 Tests de chi-deux et différences de moyenne sur les variables signalétiques 200
Tableau 6.7 Test de moyenne sur le comportement de protection et le comportement habituel 203

Tableau 6.8 Test de moyenne sur le comportement de refus selon le type d'interlocuteur 204
Tableau 6.9 Test de moyenne sur le comportement de refus selon le type de contexte 205

Chapitre 7

Tableau 7.1 Récapitulatif des propositions et hypothèses de recherche 251

Chapitre 8

Tableau 8.1 Justification du choix des facteurs à manipuler 260
Tableau 8.2 Principaux biais à maîtriser et moyens de limiter ces biais 268
Tableau 8.3 Sélection des échantillons à interroger et types de biais à maîtriser 271
Tableau 8.4 Ensemble des effets testés dans le cadre du plan expérimental 277
Tableau 8.5 Plan expérimental 278
Tableau 8.6 Variables contrôlées dans chacune des expériences 283
Tableau 8.7 Les menaces à la validité interne et les moyens de s'en prémunir 289

Chapitre 9

Tableau 9.1 Etapes de développement et de validation d'un instrument de mesure 299
Tableau 9.2 Etapes suivies lors de l'analyse factorielle exploratoire (AFE) 301
Tableau 9.3 Interprétation du Coefficient Alpha de Cronbach selon Peterson (1995) 302
Tableau 9.4 Etapes de réalisation d'une analyse confirmatoire (AFC) 302
Tableau 9.5 Les indices d'ajustement retenus : caractéristiques et valeurs critiques 304
Tableau 9.6 Identification des concepts de la recherche selon plusieurs critères 308
Tableau 9.7 Démarche adoptée pour chacun des construits de la recherche 310
Tableau 9.8 Echelle de mesure initiale de la confidentialité perçue (CONFID) 311
Tableau 9.9 Résultat de l'analyse factorielle exploratoire sur l'échelle de confidentialité 312
Tableau 9.10 Echelle de mesure initiale de la sensibilité perçue (SENS) 313
Tableau 9.11 Résultat de l'analyse factorielle exploratoire sur l'échelle de sensibilité 314
Tableau 9.12 Echelle de mesure initiale de la pertinence perçue (PERT) 315
Tableau 9.13 Résultat de l'analyse factorielle exploratoire sur l'échelle de pertinence 315
Tableau 9.14 Echelle de mesure initiale de l'attitude (ATT) 317
Tableau 9.15 Résultat de l'analyse factorielle exploratoire sur l'échelle d'attitude 317

Tableau 9.16 Echelle de mesure de la préoccupation pour le respect de la vie privée (PREOC) 319
Tableau 9.17 Résultat de l'analyse factorielle exploratoire sur l'échelle de préoccupation 320
Tableau 9.18 L'échelle d'implication (PIA) originale (Strazzieri 1994) et l'échelle simplifiée 321
Tableau 9.19 Résultat du pré-test de l'échelle PIA (IMPL) sur la « téléphonie mobile » 321
Tableau 9.20 Résultat de la réplication de l'analyse factorielle exploratoire 323
Tableau 9.21 Résultat de l'analyse factorielle multivariée sur les trois variables de perception 323
Tableau 9.22 Etapes préalables au test du modèle de mesure sous AMOS 324
Tableau 9.23 Vérification de la normalité des données 324
Tableau 9.24 Qualités psychométriques des échelles de perceptions et d'attitude 328
Tableau 9.25 Echelle de mesure initiale de la valeur perçue (VAL) 332

Tableau 9.26 Résultat de l'analyse factorielle exploratoire sur l'échelle de valeur perçue 333
Tableau 9.27 Echelle de mesure de l'intention de poursuivre la relation (IPR) 335
Tableau 9.28 Résultat de l'AFE sur l'échelle d'intention de poursuivre la relation (IPR) 336
Tableau 9.29 Coefficients d'asymétrie et d'aplatissement pour les items de valeur perçue 337
Tableau 9.30 Résultat de la réplication de l'analyse factorielle sur l'échelle de valeur perçue 338
Tableau 9.31 Résultat de la dernière analyse factorielle sur l'échelle de valeur perçue 339
Tableau 9.32 Réplication de l'analyse factorielle sur l'intention de poursuivre la relation 339
Tableau 9.33 Qualités psychométriques des dimensions de l'échelle de valeur perçue 343
Tableau 9.34 Qualités psychométriques de l'échelle d'intention de poursuivre la relation 345

Chapitre 10

Tableau 10.1 Opérationnalisation des Mentions de Respect de la Vie Privée (Facteur 1) 363
Tableau 10.2 Opérationnalisation de la Familiarité avec l'entreprise sollicitante (Facteur 2) 366
Tableau 10.3 Opérationnalisation de la Quantité de données demandées (Facteur 3) 368
Tableau 10.4 Le design expérimental en plan factoriel complet 369
Tableau 10.5 Instruments de mesure utilisés pour les variables externes contrôlées 371
Tableau 10.6 Récapitulatif des hypothèses à tester dans l'expérience 1 378
Tableau 10.7 Caractéristiques de l'échantillon interrogé dans le cadre de l'expérience 1 381
Tableau 10.8 Indépendance des observations en fonction des traitements expérimentaux 384
Tableau 10.9 Orthogonalité des facteurs manipulés (corrélations de Pearson) 385
Tableau 10.10 Analyse factorielle exploratoire sur les « *Manipulation Checks* » 386
Tableau 10.11 Vérification des manipulations expérimentales 386
Tableau 10.12 Niveau de réalisme perçu du scénario proposé 387
Tableau 10.13 Réalisme perçu selon les caractéristiques individuelles et les facteurs manipulés 388
Tableau 10.14 Qualités psychométriques des instruments de mesure utilisés 388
Tableau 10.15 Plan de traitement des hypothèses correspondant à l'expérience 1 390
Tableau 10.16 Influence des facteurs manipulés sur les trois variables de perception 393
Tableau 10.17 Régression visant à expliquer l'attitude face à la communication des données 401
Tableau 10.18 Récapitulatif des hypothèses à tester et des résultats obtenus (Expérience 1) 403

Chapitre 11

Tableau 11.1 Opérationnalisation de la Quantité de données demandées (Facteur 1) 417
Tableau 11.2 Opérationnalisation du Contexte de sollicitation (Facteur 2) 420
Tableau 11.3 Design expérimental complet de l'expérience 2 422
Tableau 11.4 Instruments de mesure utilisés pour les variables externes contrôlées 424
Tableau 11.5 Récapitulatif des hypothèses à tester dans l'expérience 2 429
Tableau 11.6 Caractéristiques de l'échantillon interrogé dans le cadre de l'expérience 2 432
Tableau 11.7 Répartition des effectifs par traitement 435
Tableau 11.8 Répartition des effectifs par traitement et par facteur bloqué (sexe) 435
Tableau 11.9 Indépendance des observations en fonction des traitements expérimentaux 437
Tableau 11.10 Analyse factorielle exploratoire sur les « *Manipulation Checks* » 438
Tableau 11.11 Vérification des manipulations expérimentales 439
Tableau 11.12 Analyse factorielle sur les variables Attitude, Préoccupation et Implication 440
Tableau 11.13 Qualités psychométriques des autres échelles 441
Tableau 11.14 Plan de traitement des hypothèses correspondant à l'expérience 2 443

Tableau 11.15 Influence des facteurs manipulés sur la valeur perçue (VAL) 445
Tableau 11.16 Régression visant à expliquer l'attitude face à la communication des données 451
Tableau 11.17 Régression visant à expliquer l'intention de fournir les données (IR) 453
Tableau 11.18 Régression visant à expliquer l'intention de fournir des données erronées (IM) 454
Tableau 11.19 Régression visant à expliquer l'intention de poursuivre la relation (IPR) 458
Tableau 11.20 Test de l'effet médiateur de l'attitude (ATT) et de l'intention de répondre (IR) 460
Tableau 11.21 Récapitulatif des hypothèses à tester et des résultats obtenus (Expérience 2) 462

Chapitre 12

Tableau 12.1 Facteurs manipulés dans l'expérience 3 474
Tableau 12.2 Instruments de mesure utilisés pour les variables externes contrôlées 478
Tableau 12.3 Répartition des individus selon les traitements et les facteurs manipulés 482
Tableau 12.4 Récapitulatif des hypothèses à tester dans l'expérience 3 498
Tableau 12.5 Types de données selon la longueur du questionnaire (Facteur Quantité) 500
Tableau 12.6 Profil sociodémographique des répondants et des non répondants 505
Tableau 12.7 Description du profil des individus appartenant aux différents échantillons 506
Tableau 12.8 Profil sociodémographique des répondants et des non répondants au post-test 508
Tableau 12.9 Principe de pondération appliqué dans l'échantillon mixte 509
Tableau 12.10 Résultats des tests de chi-deux pour tester l'indépendance des observations 510
Tableau 12.11 Echelle de mesure du comportement habituel 512
Tableau 12.12 Validité et cohérence interne de l'échelle d'intention de poursuivre la relation 513
Tableau 12.13 Validité et cohérence interne de l'échelle de préoccupation pour le RVP 513
Tableau 12.14 Echelle de mesure de la qualité perçue du programme Elisa 514
Tableau 12.15 Validité et cohérence interne de l'échelle de qualité perçue du programme 514
Tableau 12.16 Les différentes techniques d'analyse selon la nature des variables testées 516
Tableau 12.17 Plan de traitement des hypothèses correspondant à l'expérience 3 516

Tableau 12.18 Impact des facteurs manipulés sur la communication de données personnelles 524
Tableau 12.19 Tableau croisé Communication des données / Familiarité et Contexte 524
Tableau 12.20 Résultat concernant la probabilité de communiquer des données (CDP) 525
Tableau 12.21 Impact des facteurs manipulés sur la probabilité d'abandonner 527
Tableau 12.22 Tableau croisé Abandon / Contexte et Quantité 527
Tableau 12.23 Résultat concernant la probabilité d'abandonner (Abandon) 528
Tableau 12.24 Impact des facteurs manipulés sur la communication de données sensibles 530
Tableau 12.25 Tableau croisé Communication de données sensibles / Familiarité et Contexte 530
Tableau 12.26 Résultat concernant la probabilité de communiquer des données sensibles (ND) 531
Tableau 12.27 Impact de la réponse sur la probabilité de poursuivre la relation 532
Tableau 12.28 Impact de l'abandon sur la probabilité de poursuivre la relation 533
Tableau 12.29 Test de l'effet médiateur des variables de réponse (CDP et Abandon) 535
Tableau 12.30 Récapitulatif des hypothèses à tester et des résultats obtenus (Expérience 3) 539

Discussion et Mise en perspective des Résultats de la Recherche

Tableau D.1 (partie 1) Récapitulatif des résultats de la recherche concernant la validation du processus de réponse 547
Tableau D.1 (partie 2) Récapitulatif des résultats de la recherche concernant l'impact des facteurs situationnels 548
Tableau D.2 Mise en perspective des résultats de la recherche 565
Tableau D.3 Modélisation explicative du processus de réponse à une sollicitation de données 568
Tableau D.4 Effectifs des groupes de préoccupation pour le RVP dans les études 2 à 5 576
Tableau D.5 Tests de chi-deux sur le niveau de préoccupation pour le RVP 577
Tableau D.6 Solutions envisagées pour pallier les risques de refus et de rupture 578
Tableau D.7 Effets de l'Ordre sur la communication de données et l'abandon (EXP3) 592
Tableau D.8 Les éléments de la prise de décision susceptibles d'être « altérés » 602

Conclusion générale

Tableau C.1 Récapitulatif des études réalisées dans le cadre de ce travail doctoral 613

LISTE DES FIGURES

Introduction

Figure I.1 Démarche de la recherche 14

Chapitre 1

Figure 1.1 Acteurs concernés par la collecte et l'utilisation de données consommateurs 28

Chapitre 4

Figure 4.1 Catégories de variables influençant la réaction à une sollicitation de données 100
Figure 4.2 Les antécédents individuels étudiés dans la littérature 103
Figure 4.3 Typologie des facteurs situationnels influents 118

Chapitre 5

Figure 5.1 Modèle processuel de réponse à une sollicitation de données personnelles 171

Chapitre 6

Figure 6.1 Variance intraclasse / variance interclasse en fonction du nombre de classes 195
Figure 6.2 Cadre d'analyse en vue d'une typologie des internautes français 196

Chapitre 7

Figure 7.1 Analyse coûts/bénéfices dans le cadre d'une sollicitation de données personnelles 220
Figure 7.2 Processus d'évaluation dans le cadre de la théorie de la motivation 222
Figure 7.3 Modèle de structure cognitive de Olson et Dover (1978) 223
Figure 7.4 Théorie de l'action raisonnée de Fishbein (1967) et Fishbein et Ajzen (1975) 223
Figure 7.5 Processus de réponse du consommateur face à une sollicitation de données 226
Figure 7.6 Modèle conceptuel général de la recherche 230

Chapitre 8

Figure 8.1 Modèle conceptuel simplifié 262

Figure 8.2 Modèle général à tester : Attitude et comportement du consommateur face à une sollicitation de ses données personnelles 273
Figure 8.3 Facteurs manipulés dans chacune des expériences 276

Chapitre 9

Figure 9.1 Modèle de mesure des trois variables de perception et indices d'ajustement 326
Figure 9.2 Impact des trois variables de perception sur l'attitude et indices d'ajustement 327
Figure 9.3 Impact de la préoccupation pour le RVP sur l'attitude et indices d'ajustement 329
Figure 9.4 Modèle de mesure des deux dimensions de valeur perçue et indices d'ajustement 341
Figure 9.5 Modèle de mesure de l'échelle de valeur perçue et indices d'ajustement 342
Figure 9.6 Modèle de mesure de l'attitude et de l'intention de poursuivre la relation 344

Chapitre 10

Figure 10.1 Phases suivies pour mener l'expérimentation 354
Figure 10.2 Modèle simplifié testé dans l'expérience 1 355
Figure 10.3 Modèle à tester dans le cadre de l'expérience 1 373
Figure 10.4 Impact des Mentions de RVP sur la confidentialité et la sensibilité perçues 394
Figure 10.5 Effets d'interaction Familiarité x Quantité 396
Figure 10.6 Impact de la Quantité sur la sensibilité et la pertinence perçues 397
Figure 10.7 Récapitulatif des hypothèses validées dans le cadre de l'expérience 1 405

Chapitre 11

Figure 11.1 Modèle simplifié testé dans l'expérience 2 412
Figure 11.2 Modèle à tester dans le cadre de l'expérience 2 426
Figure 11.3 Les relations de médiation dans le cadre de l'expérience 2 443
Figure 11.4 Impact de la Quantité sur la valeur perçue 446
Figure 11.5 Impact du Contexte et du Contexte x Quantité sur la valeur perçue 448
Figure 11.6 Impact du Contexte et du Contexte x Quantité sur l'attitude 448
Figure 11.7 Récapitulatif des hypothèses validées dans le cadre de l'expérience 2 465

Chapitre 12

Figure 12.1 Démarche expérimentale de l'expérience 3 (phase 1) 483
Figure 12.2 Démarche expérimentale de l'expérience 3 (phase 2) 487
Figure 12.3 Modèle à tester dans le cadre de l'Expérience 3 489
Figure 12.4 Processus de sélection des données 503
Figure 12.5 Les relations de médiation dans le cadre de l'expérience 3 517
Figure 12.6 Récapitulatif des hypothèses validées dans le cadre de l'expérience 3 540

Discussion et Mise en perspective des Résultats de la Recherche

Figure D.1 Ensemble des hypothèses validées dans le cadre de cette recherche 561

LISTE DES ENCADRÉS

Introduction générale

Encadré I.1 Exemple de formulaire Internet permettant le recueil de données personnelles 3

Chapitre 10

Encadré 10.1 Copie écran de la page « Nos engagements » sur le site de SFR 362
Encadré 10.2 Copie écran du bas d'une page formulaire d'inscription sur le site de SFR 362
Encadré 10.3 Extrait du formulaire pour le traitement « Mentions de RVP limitées » 364
Encadré 10.4 Extrait du formulaire pour le traitement « Mentions de RVP nombreuses » 364
Encadré 10.5 Extrait du formulaire pour le traitement « Familiarité nulle » 366
Encadré 10.6 Extrait du formulaire pour le traitement « Familiarité élevée » 366
Encadré 10.7 Extrait du formulaire pour le traitement « Quantité faible » 368
Encadré 10.8 Extrait du formulaire pour le traitement « Quantité élevée » 369

Chapitre 11

Encadré 11.1 Extrait du formulaire pour le traitement « Quantité faible » 417
Encadré 11.2 Extrait du formulaire pour le traitement « Quantité élevée » 418
Encadré 11.3 Exemple de procédure de sollicitation de données dans le cadre de l'inscription à un Espace Personnel sur le site de l'opérateur Orange 419
Encadré 11.4 Page 6 du questionnaire correspondant au traitement « Contexte Jeu-concours » 421
Encadré 11.5 Page 7 du questionnaire correspondant au traitement « Contexte Jeu-concours » 421

LISTE DES ANNEXES

Chapitre 5
Processus de réponse face à une sollicitation de données personnelles

Annexe 5.1	Guide d'entretien pour l'étude qualitative	A1
Annexe 5.2	Consignes pour le déroulement de l'entretien	A4
Annexe 5.3	Fiche d'identification de la personne interviewée	A5
Annexe 5.4	Les étapes de l'analyse de contenu	A6
Annexe 5.5	Tableaux présentant les principaux résultats de l'étude qualitative	A9

Chapitre 6
La gestion des données personnelles chez le consommateur-internaute

Annexe 6.1	Résultats des études réalisées par Westin entre 1996 et 2003	A14
Annexe 6.2	Pages écrans du questionnaire (Etude 2)	A15
Annexe 6.3	Variables mesurées pour l'étude 2	A23
Annexe 6.4	Résultats de la classification sur la totalité et sur deux moitiés de l'échantillon	A24
Annexe 6.5	Résultats de la classification par la procédure « TwoStep Cluster »	A25

Chapitre 7
Formulation du cadre conceptuel général de la recherche

Annexe 7.1	Processus conduisant à une sollicitation de données personnelles sur Internet	A26

Chapitre 9
Choix et développement des instruments de mesure

Annexe 9.1	Estimation du modèle comprenant les 3 variables de perceptions (Figure 9.1)	A27
Annexe 9.2	Estimation du modèle évaluant l'impact des perceptions sur l'attitude (Figure 9.2)	A31
Annexe 9.3	Coûts et bénéfices perçus à fournir des données personnelles à une entreprise	A36
Annexe 9.4	Estimation du modèle de l'échelle de valeur perçue (Figure 9.5)	A38
Annexe 9.5	Estimation du modèle mesurant l'attitude et de l'intention de poursuivre la relation (Figure 9.6)	A42

Chapitre 10
Expliquer l'attitude du consommateur face à la communication de ses données personnelles

Annexe 10.1 Questionnaire final (Expérience 1, traitement 1) A46

Annexe 10.1 bis Extrait du questionnaire final (Expérience 1, traitement 8) A55

Annexe 10.2 Vérification de la validité interne de l'expérience 1 A57

Annexe 10.3 Analyse factorielle et fiabilité des échelles de « *Manipulation Checks* » A63

Annexe 10.4 Normalité des distributions pour les variables de Préoccupation et d'Implication A69

Annexe 10.5 Résultats de la MANCOVA réalisée dans le cadre de l'expérience 1 A70

Annexe 10.6 Vérification des conditions d'application de la régression multiple A74

Chapitre 11
Sollicitation de données personnelles et intentions comportementales de réponse

Annexe 11.1 Pages écrans correspondant aux six formulaires pour l'opérateur Orange A75

Annexe 11.2 Pages écrans du questionnaire (Expérience 2) A78

Annexe 11.3 Répartition des répondants selon le traitement, l'âge et le sexe A94

Annexe 11.4 Répartition par jour des questionnaires administrés pour l'expérience 2 A95

Annexe 11.5 Vérification de la validité interne de l'expérience 2 A96

Annexe 11.6 Analyse factorielle et fiabilité des échelles de « *Manipulation Checks* » A98

Annexe 11.7 Normalité des échelles d'implication, de préoccupation pour le RVP et de comportement habituel de réponse, tous traitements confondus A102

Annexe 11.8 Résultats de l'ANCOVA réalisée dans le cadre de l'expérience 2 A103

Annexe 11.9 Vérification des conditions d'application de la régression multiple A105

Chapitre 12
Réponses du consommateur face à une sollicitation de ses données personnelles

Annexe 12.1	Répartition des individus par cible (traitement) en %	A109
Annexe 12.2	Contenu des emails envoyés dans le cadre du jeu-concours et de l'enquête	A111
Annexe 12.3	Résultats de l'envoi des 2 emails (requalification puis post-test)	A112
Annexe 12.3 bis	Bilan des résultats concernant l'email de requalification	A116
Annexe 12.3 ter	Bilan des résultats concernant l'email post-test	A117
Annexe 12.4	Première page du site expérimental dans le cadre de l'enquête	A119
Annexe 12.5	Quelques pages du questionnaire administré dans le cadre du jeu-concours	A120
Annexe 12.6	Emails envoyés dans le cadre du post-test (version brandée Dauphine)	A124
Annexe 12.7	Questionnaire phase 1 (version Long-Sensible)	A125
Annexe 12.8	Questionnaire Post-Test (version Jeu-concours)	A128
Annexe 12.9	Comparaison des profils des individus à qui est envoyé l'email post-test	A131
Annexe 12.10	Validité discriminante des échelles multi-items	A133
Annexe 12.11	Résultats des tests de corrélations	A135
Annexe 12.12	Vérification de la normalité des variables métriques et assimilées	A138
Annexe 12.13	Courbes d'attrition eu égard à la progression au sein du questionnaire	A144

INTRODUCTION GÉNÉRALE

« Chaque individu doit pouvoir,
quels que soient sa culture, son statut et sa situation personnelle,
réaliser un ajustement continu entre ses besoins de solitude et de partage,
d'intimité ou d'intégration sociale,
d'anonymat ou de participation responsable dans la Société,
de réserve ou de divulgation de soi. »

Westin (1970)

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

Pour bâtir des relations fortes et durables avec leurs clients, de plus en plus d'entreprises mettent aujourd'hui en place des programmes relationnels ou de fidélité. Pour être efficaces, ceux-ci nécessitent de collecter de nombreuses données (notamment à caractère personnel[1]) concernant ces consommateurs, afin de proposer des offres adaptées aux goûts et besoins de chacun. La littérature, tout comme l'examen des pratiques, montre en effet que, plus l'entreprise détient d'informations sur ses clients, plus elle pourra leur apporter de la valeur (notamment sous la forme d'offres personnalisées), et donc mieux elle pourra les fidéliser (Hugues 1994, Jackson et Wang 1994). Ainsi, le succès des stratégies marketing dépendrait de plus en plus de l'utilisation efficace d'un grand nombre d'informations touchant les consommateurs (Blattberg et Deighton 1991, Glazer 1991), à la condition expresse que celles-ci soient justes, à jour et complètes.

Même en l'absence de toute démarche de fidélisation, force est de constater que l'information client est aujourd'hui une source de valeur majeure pour les entreprises. Comme le constate Bell (1973), cité par Glazer (1991), la connaissance (entendue comme le prolongement de la détention d'informations) a en effet remplacé l'énergie comme ressource première de la société. La collecte et l'utilisation efficaces des données consommateurs représentent donc aujourd'hui l'un des plus grands défis du management (Mason 1986), puisque c'est l'un des principaux moyens d'obtenir des avantages en termes de compétitivité (Glazer 1991). De fait, la composante majeure des échanges entre une entreprise et ses clients est devenue l'échange d'informations, bien avant l'échange de produits/services.

1 Contexte de la recherche

Si pendant longtemps, seules les firmes adoptant des pratiques de marketing direct ou ayant mis en place des politiques relationnelles[2] abouties ont cherché à obtenir des informations sur les consommateurs, tel n'est plus le cas aujourd'hui. On peut en effet penser qu'actuellement la quasi-totalité des organismes faisant du commerce collecte des données sur leurs clients, à

[1] Comme nous le reverrons au chapitre 1, une donnée personnelle peut se définir comme toute donnée concernant un individu identifié ou identifiable (par un ou plusieurs éléments qui lui sont propres). Il s'agit donc d'informations de toute nature qui sont directement reliées à la personne qui les a fournies.

[2] A l'instar de O'Malley, Patterson et Evans (1997), nous proposons la définition suivante du marketing relationnel : « l'identification, la spécification, l'initiation, le maintien (et éventuellement la dissolution) de relations de long terme avec le consommateur à travers un échange mutuel, la réalisation des promesses et l'adhésion à certaines normes visant à satisfaire les objectifs et à améliorer l'expérience des parties concernées ».

des degrés d'implication et d'efficacité divers cependant. Ainsi, selon Jones (1991), les compagnies d'assurance américaines disposent de données sur plus de 9 Américains sur 10.

Parallèlement, la manière dont les données sont collectées a fortement évolué. Tout d'abord, alors que les entreprises se sont longtemps contentées de récolter des informations liées au marché ou à des segments de consommateurs (i.e. caractéristiques de la cible, audience média, parts de marché, …), elles focalisent aujourd'hui leur attention sur des données spécifiques à chaque individu (i.e. de nature personnelle) comme ses caractéristiques sociodémographiques, ses goûts et préférences et ses habitudes en matière de consommation. Ensuite, alors que la collecte se faisait traditionnellement en face à face, par téléphone ou par courrier, l'arrivée des NTIC (i.e. Nouvelles Technologies de l'Information et de la Communication) est venue bouleverser le paysage. L'essor de ces technologies a ainsi permis une augmentation de la vitesse avec laquelle les données sont collectées, stockées et transmises, mais également du montant d'informations susceptibles d'être récoltées et utilisées. Helms (2001), cité par Ingargiola (2002), parle ainsi de « *cyber panopticon* » pour évoquer la facilité avec laquelle les renseignements peuvent aujourd'hui être collectés et partagés d'une part et les consommateurs surveillés d'autre part, depuis l'arrivée d'Internet.
A titre d'information, si dans les années 90, 68% des sites américains (tous acteurs confondus, y compris les associations et les administrations) cherchaient à collecter des données sur leurs visiteurs (Pardun et Lamb 1999), 97 à 99% des sites à vocation commerciale ont désormais l'habitude de le faire (étude FTC, 1998). Ces chiffres seraient d'ailleurs comparables d'un pays à l'autre. Ainsi, une étude menée sur les 80 sites français les plus fréquentés (d'après le classement Cybermétrie de juillet 2001) montre que la quasi-totalité demande des renseignements personnels à leurs visiteurs : des données d'identification[3] (comme le nom, le prénom, l'adresse postale, l'email, …) pour 97% d'entre eux et de nature sociodémographique pour 91% (Gauzente, Benetteau et Dubreuil, 2002).
L'encadré I.1 ci-après offre un exemple de formulaire trouvé sur Internet (plus précisément sur le site de l'agence de voyages Expedia) et recueillant des données de nature personnelle.

Toutefois, cet essor des pratiques de collecte et d'utilisation des données clients ne s'opère pas sans que, parallèlement, les consommateurs y trouvent à redire. Les enquêtes d'opinion réalisées sur le sujet, principalement aux Etats-Unis, révèlent ainsi que les citoyens sont de plus en plus sensibles aux risques que font peser ces pratiques.

[3] Nous reviendrons sur les différents types de données collectées par les entreprises dans le chapitre 1.

Encadré I.1 Exemple de formulaire Internet permettant le recueil de données personnelles

La littérature et l'observation des pratiques confirment que les consommateurs se méfient généralement des situations de sollicitation de leurs données personnelles. D'une part, parce qu'ils y voient une intrusion dans leur vie privée et qu'ils n'aiment pas forcément se dévoiler ; d'autre part, parce qu'ils ont peur des conséquences possibles d'un tel dévoilement, et notamment d'une utilisation ultérieure abusive des données qu'ils auront bien voulu fournir. Dans un article publié dans une revue juridique en 1890, le droit d'une personne au respect de sa vie privée (en anglais *privacy*) était défini comme le droit « d'être laissé tranquille » ("*the right to be let alone*") (Warren et Brandeis 1890). Plus d'un siècle plus tard, force est de constater que la définition de ce terme est toujours aussi ambigüe[4]. De plus, ce droit semble s'être à ce point détérioré que Scott McNealy[5], Directeur Général de Sun Microsystems, disait récemment "*you already have zero privacy, get over it*" (The Industry Standard 1999).

[4] Nous reviendrons sur les différentes définitions de ce terme dans le chapitre 2.
[5] Cité par Milne et Rohm (2000).

Dans l'ensemble, les sondages témoignent que les individus sont autant préoccupés par ce que les entreprises peuvent savoir d'eux que par la manière dont elles ont obtenu et utilisent ces informations (Harris Equifax 92/95/96, Katz et Tassone 1990). En général, les personnes interrogées estiment qu'on leur demande trop de données personnelles et qu'elles ont peu de contrôle sur ce qu'il en advient (Nowak et Phelps 1992). Ceci est loin d'être sans conséquence. Il apparaît en effet que beaucoup de consommateurs auraient pris des mesures de protection, y compris des restrictions sur leurs échanges d'informations avec les entreprises (Nowak et Phelps 1992, Wang et Petrison 1993).

Dès lors, alors que les entreprises font de plus en plus appel aux techniques impliquant la collecte et l'utilisation d'informations personnelles (Culnan et Armstrong 1999), elles doivent désormais faire face au dilemme de respecter et de protéger la vie privée des consommateurs, tout en utilisant les données qu'elles pourront obtenir sur eux avec profit. La collecte de données auprès de prospects et/ou de clients représente ainsi, comme nous le verrons par la suite, un triple défi pour les managers : éthique, réglementaire et stratégique.

2 Intérêt du sujet de recherche

La collecte des données personnelles des consommateurs tout en respectant leur vie privée est un sujet, si ce n'est de préoccupation, au moins d'attention, dans un grand nombre de communautés, et notamment pour les groupes d'activistes qui cherchent à protéger les droits de l'individu, et pour les médias qui veulent informer l'opinion publique sur ces questions. L'importance du phénomène se voit aussi dans l'acharnement des gouvernements (notamment européens) à promouvoir des lois informant et protégeant les consommateurs à ce sujet.

La question concerne toutefois plus particulièrement les entreprises, dont les besoins commerciaux doivent être confrontés au besoin de rassurer les clients et à l'obligation de se conformer à la législation en vigueur, et la communauté académique qui souhaite mieux comprendre comment on peut aboutir à un tel équilibre.

Afin d'aider les entreprises à bâtir un design de collecte efficace et respectueux des lois en vigueur et des préoccupations individuelles en matière de vie privée, il apparaît important d'étudier et de comprendre la façon dont les consommateurs sont susceptibles de réagir face aux sollicitations de données à caractère personnel dont ils peuvent faire l'objet. Il s'agit notamment de répondre aux questions que de nombreux managers se posent, à savoir :

A quelles conditions les individus acceptent-ils de fournir des informations personnelles ? Comment présenter la sollicitation pour qu'elle ne soit pas perçue comme une invasion de la vie privée ? Existe t-il un seuil de tolérance à ne pas dépasser ? Ce seuil est-il identique pour tous les individus ?

Malgré l'ensemble des études et des recherches déjà menées sur le sujet, on sait encore peu de choses sur les raisons qui poussent les individus à répondre, favorablement ou non, à une sollicitation de données personnelles. De plus, si plusieurs travaux académiques s'intéressent aux techniques visant à améliorer le taux de réponse, aucun ne s'intéresse réellement au processus ou aux critères qui guident le choix de la personne sollicitée.

Notre travail s'inscrit donc dans une perspective théorique qui vise à palier l'absence, au sein de la littérature, de modèle susceptible d'expliquer la réaction des consommateurs. En effet, bien qu'offrant un point d'ancrage intéressant, les recherches menées sur le sujet souffrent de trois lacunes majeures. Tout d'abord, la quasi-totalité adopte une logique de sondage et mesure l'attitude des individus en matière de respect de la vie privée, sans toujours évaluer son impact sur le comportement de réponse face à une sollicitation de données personnelles. De plus, lorsque l'analyse porte sur les comportements, il s'agit le plus souvent de comportements déclarés, qui sont probablement différents de ceux adoptés en situation réelle. Enfin, l'étude des facteurs situationnels (liés aux caractéristiques de la sollicitation), qui semblent pourtant particulièrement influents dans ce domaine, est encore limitée.

Dans cette perspective, cette recherche vise donc à élaborer un modèle permettant d'expliquer les réponses, attitudinales et comportementales (notamment l'acceptation ou, au contraire, le refus de répondre), des consommateurs français à différents dispositifs de sollicitation de données personnelles. Il s'agira dès lors, après avoir identifié les principaux facteurs d'influence, d'expliquer de quelle manière ils s'inscrivent dans le processus de décision, puis de mesurer leur impact sur le taux et la qualité des renseignements obtenus par les entreprises.

L'objectif de cette recherche sera donc d'identifier les situations dans lesquelles les individus sont favorables ou au contraire défavorables à la divulgation de leurs données personnelles. Le cadre conceptuel que nous serons amenés à tester doit ainsi permettre aux managers de tenir compte non seulement de leurs besoins en informations, mais aussi des attentes des consommateurs concernant la manière dont les données pourront être utilisées par la suite. Il s'agira donc de mesurer les perceptions des individus sur ces questions et de voir en quoi cela affecte la façon dont ils se comportent face à ces demandes (i.e. réponse ou non) d'une part, et leur confiance dans les entreprises qui les auront sollicités, d'autre part.

3 Délimitation du sujet et choix préalables

Un individu peut se trouver confronté à une sollicitation de données personnelles en de multiples occasions, aussi bien dans le cadre de ses relations interpersonnelles, qu'à son travail (par exemple, lors d'un entretien d'embauche), ou dans ses rapports avec l'Etat, lors des différentes démarches administratives qu'il peut être amené à effectuer. Nous nous intéresserons, pour notre part, aux sollicitations de nature commerciale, c'est-à-dire celles réalisées par toute entité ayant vocation à vendre des produits/services. Nous exclurons toutefois le cadre bancaire et médical, étant donné la spécificité des liens qui unissent les deux entités (i.e. client/conseiller et patient/médecin) d'une part, et la nature des données sollicitées d'autre part.

Comme nous le verrons au chapitre 1, les données clients peuvent être obtenues de différentes manières. On distingue traditionnellement trois modes de collecte : la collecte discrète (sans que le consommateur soit informé et/ou conscient que des données sont collectées), la collecte directe (qui implique une sollicitation de la part de l'entreprise et une forme de dévoilement de soi de la part du consommateur) et la collecte indirecte (informations contenues dans des bases de données louées ou achetées par l'entreprise à des tiers). Dans le cadre de ce travail, nous nous intéresserons exclusivement aux collectes directes, pour lesquelles le consommateur fait le choix de fournir ou non les données demandées.

De plus, à des fins d'application, nous étudierons plus spécifiquement le cas des sollicitations de données personnelles sur Internet, par le biais d'un formulaire à remplir en ligne. On sait en effet que, de nos jours, cet outil est de plus en plus utilisé par les entreprises pour collecter des données auprès de leurs prospects/clients. Parallèlement, les caractéristiques spécifiques de ce média tendent à exacerber les menaces qui pèsent sur la vie privée et donc les craintes de se dévoiler. S'intéresser à la sollicitation de données personnelles dans ce contexte doit donc permettre d'étudier une situation *a priori* paradoxale qui veut que l'individu soit peu enclin à fournir des données en ligne, tout en étant davantage sollicité par ce biais. En outre, les particularités d'Internet n'empêcheront pas, le moment venu, de comparer voire d'étendre nos résultats aux sollicitations de données opérées sous d'autres formes (i.e. par courrier, téléphone ou face à face notamment), tout en prenant cependant les précautions de rigueur.

4 Objectif, problématique et questions de recherche

L'objectif premier de cette recherche sera donc d'identifier l'ensemble des variables ayant une influence sur l'attitude et le comportement des consommateurs, confrontés à une sollicitation d'informations personnelles.

Il s'agira ensuite de tester, plus particulièrement, l'impact des caractéristiques de la sollicitation (variables situationnelles) sur la décision de l'individu de donner (ou non) les renseignements demandés. Le cas des formulaires en ligne sera, plus spécifiquement, étudié.

Enfin, le fait, pour un individu, d'être sollicité par une entreprise pour fournir des données personnelles ne serait pas sans incidence sur la nature de la relation qui unit ces deux entités (Schoenbachler et Gordon 2002). Ainsi, au-delà de la réponse à la sollicitation elle-même, il est également intéressant de voir si la collecte est susceptible d'influencer la pérennité de la relation que le consommateur entretient avec l'entreprise ayant sollicité les données.

Notre problématique peut donc se formuler de la manière suivante :

Comment expliquer la réponse du consommateur
face à une sollicitation de ses données personnelles ?

Autrement dit, nous nous intéresserons, plus particulièrement, à deux points principaux :

- Qu'est-ce qui influence la décision d'un individu de fournir (ou non) des données personnelles à une entreprise ?
- Comment (i.e. en fonction de quels critères) cette décision se prend elle ?

Pour traiter cette problématique, nous avons dû répondre aux questions formulées ci-dessous :

1. Comment le consommateur perçoit-il la sollicitation de ses données personnelles ?

La décision d'un individu de fournir des données personnelles à une entreprise repose essentiellement sur la manière dont il va percevoir la sollicitation. L'étude de cette perception représente donc la première étape dans la compréhension des mécanismes conduisant à la décision. L'exploration qualitative que nous serons amenés à réaliser (présentée au chapitre 5) nous permettra ainsi d'identifier quatre critères sur lesquels repose prioritairement l'évaluation (ou perception) de la sollicitation. Il s'agit de la confidentialité, de la sensibilité et de la pertinence perçues des données sollicitées auxquelles s'ajoute la perception de la valeur (ou du gain net) à répondre.

2. Quels sont les effets de ces perceptions sur le comportement du consommateur ?

La manière dont l'individu perçoit la sollicitation aura une incidence majeure sur sa façon de réagir. Il s'agira donc, dans un second temps, d'élaborer un modèle visant à identifier et à évaluer le processus de décision adopté par le consommateur, en réponse à la sollicitation dont il fait l'objet. La théorie de l'action raisonnée (Fishbein et Ajzen 1975) sera mobilisée ici pour aboutir au modèle conceptuel général de cette recherche, qui sera présenté au chapitre 7.

3. En quoi et de quelle manière la situation de collecte peut-elle influencer la réponse ?

La perception de la sollicitation est, elle-même, fonction de la manière dont celle-ci se présente. La situation de collecte aurait ainsi une incidence majeure sur la décision. Nous tenterons, dès lors, de répondre à la question suivante : quel est l'effet de la situation, telle que perçue par l'individu, sur sa décision de communiquer (ou non) les données demandées ? En cela, nous considèrerons majoritairement l'effet des facteurs situationnels sur le processus de décision. De plus, comme nous nous intéressons aussi à la pérennité de la relation avec l'entreprise ayant sollicité les données, nous étudierons également l'effet des caractéristiques de la situation sur l'intention du consommateur de continuer à commercer avec elle.

4. Quel est l'effet des facteurs individuels sur la réponse à la sollicitation de données ?

La littérature sur le respect de la vie privée met en exergue l'influence des variables individuelles sur la décision du consommateur de fournir les données demandées. Il s'agit alors de répondre à la question suivante : existe-t-il des différences individuelles dans l'attitude et le comportement de réponse face à une sollicitation de données personnelles ? Et si oui, de quelle nature ? Si une majorité de recherches s'intéresse à l'effet de ce type de facteurs, force est de constater qu'elles sont loin d'aboutir pour autant à un consensus avéré. Nous évaluerons, pour notre part, l'influence des facteurs individuels à travers une variable fréquemment mobilisée dans les recherches académiques portant sur le sujet, à savoir la préoccupation pour le respect de la vie privée (en anglais, *privacy concern*).

De plus, comme nous étudions également l'impact des facteurs situationnels sur la réponse du consommateur face à une sollicitation de ses données personnelles, nous chercherons aussi (et peut-être surtout) à mesurer l'impact respectif de ces deux catégories de facteurs (i.e. individuels et situationnels) sur le processus de réponse.

L'importance de cette problématique est justifiée, tant par le contexte dans lequel évoluent actuellement les entreprises (et, plus spécifiquement, celles qui sont présentes sur Internet), que par les lacunes identifiées dans la littérature. Sa portée se matérialise aussi à travers les contributions attendues de cette recherche.

5 Contributions attendues de la recherche

Les contributions attendues de cette recherche sont à la fois théoriques, empiriques et managériales. Elles peuvent être complétées par des apports d'ordre méthodologique.

D'un point de vue théorique, notre contribution réside principalement en trois points.
Tout d'abord, l'objectif principal de cette recherche est d'élaborer un modèle visant à expliquer le comportement de réponse d'un individu qui se trouve sollicité pour fournir des données à caractère personnel ainsi que le processus permettant d'y aboutir. Etant donnée l'absence d'un tel modèle dans la littérature, il s'agit donc là d'une contribution majeure.

La deuxième contribution d'ordre théorique réside dans l'adoption d'une perspective processuelle et réaliste. En effet, la majorité des études menées jusqu'ici s'intéresse essentiellement à la mesure de l'attitude des individus face aux invasions de leur vie privée, sans aller jusqu'à évaluer leur comportement quand une entreprise leur demande de fournir des données personnelles. Dans ce domaine comme dans beaucoup d'autres, des relations semblent exister entre les cognitions, les attitudes et les comportements (Regan 1995), mais aucune recherche ne les a jusqu'à présent réellement étudiées. De plus, quand le comportement est mesuré, il s'agit rarement de comportement réel mais plutôt de comportement déclaré. Nous chercherons, pour notre part, tout en découvrant les liens entre les différentes étapes du processus de réponse, à évaluer le comportement de l'individu lorsqu'il se trouve réellement confronté à une sollicitation de ses données personnelles.

Enfin, les recherches portant sur les facteurs encourageant le dévoilement de soi (notamment de données personnelles à une entreprise qui en fait la demande) sont encore limitées. De plus, la majorité de ces travaux se focalisent sur l'étude des facteurs individuels alors que celle des facteurs situationnels nous semble tout aussi, si ce n'est plus, prometteuse. Nous intéresser plus particulièrement à l'impact de ces derniers constitue donc un apport significatif. Comme nous le verrons au chapitre 8, nous étudierons ainsi, plus

particulièrement, quatre facteurs situationnels dont l'influence a été très peu (voire jamais) étudiée dans la littérature alors même qu'ils présentent un intérêt managérial certain, en tant qu'éléments sur lesquels une entreprise peut facilement agir (aspect opérationnel).

La principale contribution d'ordre empirique, tient à l'étude du comportement des internautes français. En effet, la quasi-totalité des recherches menées sur le sujet concerne le public américain et, de façon anecdotique, le public européen (anglais et allemand). A notre connaissance, aucune étude empirique n'a jusqu'à présent étudié, de façon approfondie, la manière dont les consommateurs français considèrent le respect de la vie privée en général, sur Internet en particulier, ni comment leurs opinions affectent leurs comportements. Seule la recherche menée par Gauzente (2004) constitue un premier pas en ce sens. Elle ne mesure toutefois pas les facteurs susceptibles d'influencer le comportement de réponse, ce dernier étant en outre, évalué dans un contexte général, sans que l'on puisse savoir si les individus interrogés réagiraient systématiquement de cette façon. Evaluer ce phénomène dans un contexte français se révèle donc une priorité, d'autant que la différence de développement économique et de pénétration du média Internet, tout comme les écarts culturels et de style de vie, laissent présager que les résultats seront différents de ceux obtenus outre-Atlantique.

Si les chercheurs sont soucieux de comprendre la manière dont les consommateurs prennent leur décision quand ils se trouvent confrontés à une sollicitation de données personnelles, les managers cherchent également à mieux appréhender le phénomène et à savoir comment optimiser, à la fois, les coûts de la collecte et la satisfaction des clients sollicités. Précisément, il s'agit de répondre aux questions suivantes : *que peut-on demander ? et sous quelle forme ?*

Pour les entreprises, la compréhension des préoccupations et des réactions des consommateurs en matière de sollicitation d'informations peut les aider à distinguer les usages autorisés, des utilisations souhaitables, des données à caractère personnel. Il s'agit, en effet, de faire la différence entre ce qu'elles souhaitent et ce qu'elles peuvent réaliser en la matière. Notre analyse peut ainsi contribuer à identifier les pratiques susceptibles d'entraîner des réactions négatives de la part des consommateurs sollicités et à développer ou maintenir des avantages concurrentiels dans ce domaine, par la mise en place de politiques proactives en matière de protection des données personnelles en particulier, et de la vie privée, en général.

Sur le plan managérial, nos résultats doivent donc permettre aux responsables d'identifier les « mécanismes » à mettre en place, avant ou au moment de la sollicitation, pour accroître la probabilité de réponse des clients et ainsi obtenir des informations fiables. Ils doivent aussi les amener à réfléchir aux conséquences possibles d'une sollicitation jugée trop intrusive. Nos conclusions doivent en effet être l'occasion de montrer que si les données consommateurs sont nécessaires au développement de relations, elles ont aussi le pouvoir de les détruire.

Alors que les informations clients deviennent un facteur clé de succès dans l'économie actuelle, la question du respect de la vie privée reste donc d'une importance stratégique. En identifiant les facteurs affectant la volonté des consommateurs de partager leurs données, les managers pourront alors concevoir des offres les encourageant à profiter de la personnalisation. Notre travail aboutira donc à la formulation d'un ensemble de recommandations concrètes afin de rendre les dispositifs de collecte de données clients plus efficaces, tant d'un point de vue économique que relationnel.

Enfin, sur le plan méthodologique, notre apport tient essentiellement à la validation d'un modèle conceptuel par le biais d'un plan expérimental comprenant trois phases complémentaires. Si le recours à une logique expérimentale est désormais usuel, y compris en France, la forme sous laquelle nous l'envisageons ici constitue, selon nous, une nouveauté.
Le plan d'expérience imaginé découpe en effet le modèle en blocs successifs (correspondant à une ou plusieurs étapes majeures du processus de réponse), et le teste en trois fois, chaque expérimentation se focalisant plus particulièrement sur l'un des blocs.
De plus, afin d'améliorer la validité de la recherche, tout en limitant les différents biais susceptibles de la réduire, nous avons choisi de faire varier, dans chacune des expériences, le type d'échantillon interrogé et le contexte d'interrogation. L'expérience 1 (présentée au chapitre 10) consistera ainsi à interroger des étudiants, par questionnaire papier-crayon, et à mesurer leur attitude face à la communication de données personnelles. Dans l'expérience 2 (présentée au chapitre 11), nous soumettrons des internautes panélistes à un questionnement en ligne et mesurerons leurs intentions comportementales de réponse. Enfin, l'expérience 3 (présentée au chapitre 12) sera l'occasion de nous situer dans un contexte réel, en observant la manière dont les clients d'une entreprise partenaire répondent à une sollicitation de données.

6 Positionnement de la recherche

Notre approche du sujet et la manière dont nous avons choisi de le traiter ont fait l'objet d'un raisonnement itératif, enrichi à la fois par l'étude de la littérature, les premiers résultats empiriques trouvés (issus de deux études exploratoires, l'une à vocation qualitative, l'autre de portée descriptive) et nos relations avec différents professionnels s'intéressant à ces questions. Plusieurs partenariats passés avec des acteurs expérimentés dans ce domaine (présentés respectivement dans les chapitres 6 et 12) ont ainsi autant conditionné nos analyses, que contribué à consolider nos résultats.

La posture générale de cette recherche est hypothético-déductive et s'inscrit donc dans le paradigme positiviste. Nos hypothèses de recherche ont été formulées sur la base d'un corpus théorique varié (littérature sur le respect de la vie privée, le dévoilement de soi et la méthodologie d'enquête d'une part, théories de l'échange social, de la motivation et de l'action raisonnée, d'autre part), enrichi de deux études exploratoires. Nous rejoignons ainsi les exigences de Sheth, Gardner et Garrett (1988), qui militent pour une approche scientifique et une définition rigoureuse des concepts en marketing. De plus, la dialectique entre le terrain et la littérature a été constante, avant d'aboutir au cadre conceptuel de la recherche.

7 Plan de la recherche

Nous avons organisé ce document en trois grandes parties, composées de trois à cinq chapitres chacune.

La première partie (intitulée « Comprendre »), présente une revue de la littérature sur le respect de la vie privée et la réponse à une sollicitation de données personnelles.
Dans le chapitre 1, après avoir présenté les différentes formes de collecte de données clients, nous en rappelons les principaux enjeux.
Le chapitre 2 est l'occasion de revenir sur la notion de vie privée, de la définir et d'en donner les principales caractéristiques. Nous insistons également, plus particulièrement, sur la vie privée en tant que protection des données à caractère personnel, concept qui sera majoritairement mobilisé dans cette recherche.
Dans le chapitre 3, nous étayons le principe du dévoilement de soi qui, à l'origine étudié dans le contexte des relations interpersonnelles, peut aussi s'appliquer à un contexte commercial.

Enfin, dans le chapitre 4, nous détaillons les résultats des principales recherches déjà menées au sujet des réactions des consommateurs face à une sollicitation de leurs données personnelles.

La deuxième partie (intitulée « Explorer et Bâtir ») indique la démarche suivie pour élaborer notre modèle conceptuel puis le tester.
Les chapitres 5 et 6 exposent respectivement les résultats de deux études exploratoires (l'une d'ordre qualitatif, l'autre à portée descriptive) qui nous ont permis d'enrichir la revue de littérature et de préparer la formulation du cadre conceptuel et des hypothèses de recherche. Ces derniers sont présentés au chapitre 7, qui s'appuie, en outre, sur un ancrage théorique varié, depuis la théorie de l'échange social jusqu'à celle de l'action raisonnée en passant par la théorie de la motivation (ou de l'espérance attendue).
Dans le chapitre 8, nous exposons la démarche méthodologique suivie pour cette recherche, sous la forme d'un plan expérimental.
Enfin, dans le chapitre 9, nous détaillons la procédure et les études menées pour développer les échelles de mesure nécessaires au test et à la validation de notre cadre conceptuel.

La troisième partie (intitulée « Expliquer ») vise à tester le modèle conceptuel proposé.
Dans les chapitres 10, 11 et 12, nous exposons successivement les résultats des trois dernières études menées, qui correspondent aux phases du plan expérimental présenté au chapitre 8. Ces études se distinguent, notamment, par la variété des publics interrogés et par la nature des variables dépendantes étudiées.

Enfin, nous terminons ce document en soulignant les contributions, les limites et les voies de recherche que ce travail soulève.

La structure générale qui guide l'exposé de cette thèse est présentée à la page suivante.

Figure I.1 Démarche de la recherche

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

PARTIE I

COMPRENDRE

Collecte de données clients et Respect de la vie privée

Les apports de la littérature

INTRODUCTION DE LA PARTIE I

L'étude des réactions des consommateurs face à une sollicitation de leurs données personnelles soulève, dans un premier temps, plusieurs interrogations d'ordre théorique.

Cette première partie (intitulée « COMPRENDRE ») a, en particulier, pour objectif de présenter les connaissances établies sur le sujet et de synthétiser l'ensemble des conclusions issues de la littérature.

Il s'agit, plus précisément, de répondre aux questions suivantes :

1. Quels sont, pour une entreprise, les enjeux liés à la collecte de données clients ?
Sont-ils compatibles avec les attentes des consommateurs en matière de respect de la vie privée ?

2. Comment peut-on définir la notion de vie privée ?
En quoi la protection des données à caractère personnel peut-elle s'y rattacher ?

3. Quel est le processus mis en œuvre en cas de dévoilement de soi ?
Quels sont les facteurs susceptibles d'influencer un tel processus ?

4. Quels sont les résultats établis dans la littérature concernant la réaction des consommateurs face à une sollicitation de leurs données personnelles ?

Nous répondrons successivement à ces questions à travers les quatre chapitres qui composent cette première partie.

Partie I - COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

CHAPITRE 1
LES ENJEUX LIÉS A LA COLLECTE DE DONNÉES CLIENTS

« Information technology, infrastructure and the information it contains, particularly customer information, will be a critical success factor for all e business initiatives ».

Weil and vitale (2001)

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 1

La question du respect de la vie privée[6] des consommateurs, associée au marketing des bases de données et au marketing direct, déjà largement débattue en France à la fin des années 70 – comme en atteste la promulgation de la Loi dite « Informatique et Libertés » du 6 janvier 1978[7] - continue d'être un sujet de discussion important, qui reçoit une attention croissante de la part des autorités de régulation.
Si, comme le soulignent Milne et Boza (1998), les pratiques de l'ensemble des professionnels des bases de données doivent être surveillées, la discussion s'est focalisée récemment sur la collecte et l'utilisation d'informations personnelles via Internet. En effet, si l'attitude des consommateurs vis-à-vis de telles pratiques va « de la tolérance à un dégoût résigné » (Hoffman, Novak et Peralta 1999) lorsqu'elles passent par les médias traditionnels - courrier et téléphone notamment - l'attente de respect de la vie privée semble accrue en ligne.

Comment les entreprises collectent-elles des données auprès de leurs clients ?
En quoi Internet pose t-il des problèmes spécifiques ?
En quoi consiste le marketing des bases de données ?
Est-il compatible avec le respect de la vie privée des consommateurs ?
Voici, résumés en quelques points, les principales questions qui se posent concernant la collecte de données consommateurs et auxquelles nous chercherons à répondre.

Nous verrons tout d'abord les différentes formes de collecte de données pouvant être utilisées par les entreprises, en insistant notamment sur les spécificités de la collecte en ligne. Nous expliquerons ensuite que, malgré l'attrait de ces méthodes, la démarche de collecte est loin d'être évidente. La sollicitation de données auprès de prospects et/ou de clients représente ainsi un triple enjeu pour les managers : un enjeu stratégique, un enjeu éthique et un enjeu réglementaire. Nous revenons dans les sections 2 et 3 sur ces deux premiers types de défis, le dernier étant abordé plus spécifiquement dans le chapitre suivant (i.e. chapitre 2).

[6] Par souci de simplification, nous utiliserons régulièrement dans la suite de ce document le « sigle » RVP pour désigner le respect de la vie privée.
[7] Nous reviendrons sur les principes inhérents à cette Loi dans le chapitre 2 (paragraphe 2.3)

1 Les formes de collecte de données clients

Nous feront le point, dans ce qui suit, sur les types de données susceptibles d'être collectées par les entreprises (1.1), avant d'aborder les différents modes de collecte de données clients (1.2). Nous terminerons en évoquant les spécificités de la collecte en ligne (1.3).

1.1 Les types de données collectées

Les entreprises cherchent à collecter de multiples renseignements sur leurs clients afin de mieux les connaître et ainsi leur faire des offres adaptées. Constitue, à cet égard, une donnée à caractère personnel (que nous appellerons par la suite donnée personnelle), « toute information relative à une personne physique identifiée ou identifiable par un numéro d'identification ou un élément qui lui est propre »[8] (comme son nom, son numéro de sécurité sociale ou son adresse électronique …). Une donnée d'ordre sociodémographique n'est donc pas personnelle en tant que telle mais le devient dès lors qu'elle est associée à des données permettant d'identifier clairement l'individu auquel elle est reliée. On parle de traitement de données à caractère personnel pour désigner toute opération portant sur de telles données, et ce, quel que soit le procédé utilisé. On distingue ainsi entre autres procédés : la collecte, la conservation, la consultation, l'utilisation, la communication, la diffusion ou toute autre forme de mise à disposition, l'effacement, la destruction ainsi que le rapprochement ou l'interconnexion entre différents fichiers. Dans le cadre de cette recherche, nous nous intéressons plus particulièrement à la collecte, par une organisation, des données personnelles de ses clients - actuels ou potentiels - à des fins d'utilisation commerciale.

On peut classer les données personnelles des consommateurs en différentes catégories. La littérature offre à cet égard une multitude de classifications possibles, chaque chercheur ou organisme travaillant sur le sujet ayant d'ailleurs tendance à développer la sienne. On trouve ainsi des classifications allant de 2 à 8 catégories possibles. Parmi les catégories de données personnelles qui reviennent le plus souvent, on distingue notamment les données d'identification (nom, adresse, numéro de sécurité sociale …), les caractéristiques sociodémographiques (sexe, âge, niveau d'études …) et les données financières (niveau des revenus, montant des crédits …). Les deux dernières catégories les plus couramment citées concernent le « style de vie » de la personne (toutes les informations qui décrivent les

[8] Définition donnée en France par la Loi relative à l'informatique, aux fichiers et aux libertés du 06 janvier 1978

hobbies, les préférences, les valeurs et la personnalité de l'individu), son comportement ainsi que ses habitudes en matière d'achat (transactions passées, type de magasin fréquenté, …).

Le tableau 1.1 ci-dessous identifie les principales classifications proposées dans la littérature.

Tableau 1.1 Principales classifications de données personnelles proposées dans la littérature

Auteur	Catégories identifiées
Classifications en 2 catégories	
Nyshadham (2000)	- données d'identification (nom, email, adresse, n°téléphone, n°CB, n°sécurité sociale, ..) - données sociodémograhiques (âge/date de naissance, informations sur la famille, sexe, niveau d'études, code postal, revenus, préférences et centres d'intérêt, profession, ..)
Gurau et al. (2003)	- données « personnelles » (nom, email, adresse, n°téléphone, n°CB, ..) - données sociodémograhiques (âge, taille de la famille, sexe, niveau d'études, revenus, profession, hobbies, ..)
O'Connor (2003)	- données « personnelles » (nom, prénom, email, adresse, pays, n°téléphone, n°CB, date de naissance, ..) - données sociodémograhiques (nationalité, nombre d'enfants, préférences, centres d'intérêt, ..)
Classifications en 3 catégories	
Chellappa et Sin (2005)[9]	- données identifiantes (nom, adresse, n°sécurité sociale, ..) - données non identifiantes (sexe, code postal, âge, ..) - données anonymes[10] (données liées à la connexion et au comportement de navigation de l'internaute : adresse IP, navigateur, heure de connexion, ..)
Nowak et Phelps (1992)	- données « personnelles » (âge, religion, n°téléphone, n°sécurité sociale, revenus, informations médicales, informations financières, ..) - habitudes de médias (programmes TV regardés, magazines lus, ..) - données comportementales (marques achetées, supermarchés et vépécisites fréquentés, ..)
Classification en 5 catégories	
Phelps et al. (2000)	- données d'identification (nom, n°téléphone, n°CB, n°sécurité sociale, ..) - données sociodémograhiques (âge/date de naissance, statut marital, niveau d'études, profession, ..) - données de « style de vie » (hobbies, loisirs, ..) - données comportementales (supermarchés fréquentés, produits/services achetés à distance, achats avec CB, ..) - données financières (revenu annuel, ..)
Classification en 8 catégories	
Weible (1993)	- informations générales (nom, prénom, adresse, ..) - données d'identification (n°CB, n°sécurité sociale, ..) - données sur l'éducation et la profession - loisirs et activités - données relatives aux normes sociales (handicaps, traits particuliers, ..) - données médicales - données financières - données relatives aux jeux d'argent et aux pensions alimentaires

[9] Ils reprennent ici la classification préconisée par le FTC (*Federal Trade Commission*), organisme américain chargé notamment des questions relatives au respect de la vie privée des citoyens américains.

[10] Catégorie spécifique à Internet

Nous retiendrons, pour notre part, la classification en 5 catégories proposée par Phelps, Nowak et Ferrell (2000) qui offre l'avantage d'être ni trop étroite (cas des classifications en 2 catégories) ni trop large (cas de la classification en 8 catégories).

Le fait, pour un consommateur, de fournir des données personnelles à une entreprise est un acte risqué, notamment en cas de divulgation de ces informations à des tiers non directement impliqués dans l'échange. Les données diffèrent cependant par le niveau de risque auquel les individus s'exposent (Milne 1997, Nowak et Phelps 1992). Ce dernier semble d'ailleurs d'autant plus important que les données sont considérées comme sensibles[11]. Plusieurs recherches suggèrent ainsi que les consommateurs protègent davantage les informations financières, les données d'identification et, de manière générale, toute donnée susceptible de conduire à des offres commerciales (Nowak et Phelps 1992, Vidmar et Flaherty 1985, Wang et Petrison 1993). Les données financières représentent, à cet égard, un cas particulier, dans la mesure où les individus n'acceptent de les fournir que si cela est nécessaire à la réalisation de la transaction, par exemple à l'occasion d'une demande de crédit auprès d'un organisme financier. La capacité de l'organisme à obtenir de telles données dépendra alors - entre autres - de la nature de la relation qu'il entretient avec le consommateur.

Dans le cadre de ses échanges avec une entreprise, le consommateur peut donc être amené à fournir des données à caractère personnel, en de multiples circonstances. Voyons à présent quels sont les différents modes de collecte susceptibles d'être utilisés.

1.2 Les modes de collecte de l'information client

Les données spécifiques à l'individu peuvent être collectées à diverses occasions : lors de l'inscription à un programme de fidélité ou d'une prise de commande, par le biais de sondages ou de jeux-concours, etc … Les informations recueillies peuvent ainsi être classées en 4 catégories (Long et al. 1999), correspondant à l'objectif qu'elles permettent d'atteindre : données d'enregistrement (liste des produits achetés lors du passage en caisse par exemple), d'ordre administratif (généralement les données d'identification comme le nom et l'adresse du client), de facilitation (données fournies pour obtenir un plus en échange, par exemple lors de l'inscription à un programme de fidélité) et données de sondage.

[11] Nous reviendrons ultérieurement sur la notion de sensibilité d'une donnée personnelle

Plus spécifiquement, les données consommateurs peuvent être obtenues de deux manières :

- de manière directe : elles sont fournies de façon volontaire par le consommateur au cours de l'échange (par exemple, lors d'une commande ou de la participation à un club) ;
- de manière indirecte : il s'agit d'informations clients contenues dans des bases de données louées ou achetées à des tiers.

Au-delà de ces deux modes de collecte, il faut aussi tenir compte du type de média utilisé (téléphone, courrier, face à face et Internet notamment) qui influence les coûts et bénéfices de la collecte aussi bien pour l'entreprise que pour le consommateur. En matière de sondage par exemple, les études permettant aux répondants de choisir leur mode de réponse montrent que les individus ont généralement un média préféré. Ainsi, pour Tse (1998), les personnes à fort désir de confidentialité pourraient préférer un sondage par courrier plutôt que par email dans la mesure où les données y sont moins facilement « traçables ».

Si pour l'entreprise, certaines données sont plus faciles à obtenir que d'autres d'un point de vue technique, la méthode utilisée n'est cependant pas neutre pour le consommateur. Dans le cas d'une vente de fichiers par exemple, les informations clients sont obtenues de manière non obtrusive. Ces derniers n'ont pas d'effort à faire pour les fournir, mais dans le même temps ils n'ont pas le choix de les transmettre ou non. D'autres données sont moins directement accessibles et obligent alors à utiliser des méthodes plus intrusives, qui impliquent de demander au client de s'engager dans une forme de dévoilement de soi[12]. C'est ce type de démarche que nous étudierons ici : sollicitation d'informations de la part d'une entreprise (approche directe) et réponse du consommateur prenant la forme d'un dévoilement de soi (plus précisément de données à caractère personnel).

Si, jusqu'à présent, ces sollicitations d'informations se faisaient souvent en face à face, par courrier ou par téléphone, aujourd'hui les entreprises ont de plus en plus recours à Internet pour recueillir des données sur leurs clients. L'efficacité de la démarche dépend alors, notamment, de la facilité avec laquelle les consommateurs acceptent de révéler des informations par le biais de l'ordinateur (Weisband et Kiesler 1996). Ce travail doctoral s'intéresse précisément à ces sollicitations de données sur Internet, celles-ci faisant face à des spécificités qu'il conviendra de prendre en compte et que nous présentons maintenant.

[12] En anglais, *self disclosure*. Cette notion, qui concerne à l'origine les relations interpersonnelles, sera largement développée au chapitre 3

1.3 Les spécificités de la collecte de données en ligne

On retrouve sur Internet les mêmes mécanismes que ceux utilisés hors ligne, avec cependant quelques spécificités liées aux particularités de ce média. De plus, aux deux modes de collecte identifiés précédemment, il faut en ajouter un troisième : la collecte « discrète ».

Internet donne la possibilité aux entreprises de recueillir plus de données personnelles qu'il n'est possible de le faire dans le monde traditionnel. En plus des données habituelles (de type sociodémographique par exemple), les sites peuvent aussi collecter des informations sur le comportement de navigation et d'achat de l'internaute. De plus, les champs à compléter sur les formulaires sont souvent plus nombreux et plus « personnels »[13]. Plusieurs consommateurs indiquent ainsi qu'il n'est pas rare de devoir indiquer sa date de naissance pour pouvoir commander en ligne alors qu'il n'y a pas besoin de le faire à la caisse d'un supermarché[14].

En ligne, la collecte directe (encore appelée « demande - réponse ») repose sur l'idée de solliciter l'internaute afin qu'il donne des informations, par le biais d'un formulaire à remplir. Il s'agit donc d'une technique « ouverte » puisque, d'une part, l'individu sait que des données sont collectées et, que d'autre part, il a la possibilité de s'y soustraire. Les internautes peuvent ainsi fournir des informations lorsqu'ils commandent en ligne, lorsqu'ils accèdent à un site par le biais d'un processus d'enregistrement ou quand ils remplissent un questionnaire sur leurs habitudes de consommation. Ces démarches se font parfois en échange de produits ou de services gratuits ou en contrepartie d'avantages promotionnels spécifiques. Il existe aussi une forme de collecte plus sournoise qui consiste à récupérer des informations sans que le consommateur ne soit directement informé et surtout sans qu'il ait la possibilité de s'y opposer[15] (Mabley 2000). La transmission des données s'opère alors passivement, par le biais de technologies dites « discrètes » (Caudill et Murphy 2000). Il s'agit notamment du suivi du

[13] Au sens de plus sensibles, plus privés, plus intimes.

[14] Ceci est cependant moins vrai si l'on considère l'importance prise par les cartes de fidélité mises en place, notamment, par les grandes enseignes de distribution. Ces cartes permettent en effet aux entreprises de relier le profil de l'individu (et donc un ensemble de données de nature personnelle fournies à l'occasion de l'ouverture de la carte) à son comportement d'achat. Nous reviendrons sur cette opposition en et hors ligne à plusieurs reprises dans ce document, et notamment à l'occasion de la discussion finale. La différence de perception des consommateurs à ce sujet constitue, en effet, une piste de recherche prometteuse.

[15] Il existe néanmoins aujourd'hui des solutions techniques permettant de réduire la portée de ces collectes dites « discrètes ». Nous ne développerons pas ce point dans la mesure où nous n'étudions pas ce mode de collecte dans le cadre de cette thèse et renvoyons donc les lecteurs intéressées aux études spécifiques sur le sujet.

comportement d'achat et de navigation par l'intermédiaire de cookies[16] ou encore de l'extraction d'adresses email auprès de listes de serveurs ou sur des forums de discussion.
Des études réalisées sur le sujet montrent que plus de la moitié des consommateurs américains (52%) n'apprécient guère les cookies (Cranor et al. 1999), ce qui indique leur faible attrait pour cette forme de collecte. Ce chiffre est d'ailleurs probablement plus important dans la réalité, sa faiblesse relative s'expliquant, notamment, par l'ignorance d'une majorité d'internautes, certains ne sachant pas vraiment de quoi il s'agit, tandis que d'autres ne mesurent pas les conséquences de ce type d'outils. Cette forme de collecte discrète, bien que non spécifique à Internet[17], a tout de même vu son utilisation s'accroître de façon exponentielle avec l'avènement de ce média. Au total, il existerait donc, tous médias confondus, trois modes de collecte de l'information client, comme l'atteste le tableau suivant. Nous nous intéresserons uniquement, dans le cadre de cette thèse, au mode de collecte direct.

Tableau 1.2 Les modes de collecte de l'information client

Modes de collecte	Explications et exemple	Types de données concernées
direct	Implique une sollicitation de la part de l'entreprise et un accord de la part du consommateur	Tous types possibles
indirect	Informations obtenues sous la forme de fichiers achetés ou loués à des tiers	Surtout des données de profil (sociodémographiques) et des données sur le comportement et les habitudes d'achat
discret	Informations recueillies sans que le consommateur n'en soit informé et sans qu'il ait la possibilité de s'y opposer (sauf à prendre des précautions particulières)	Surtout des données comportementales (comportement d'achat, comportement de navigation) et des données liées à la connexion (adresse IP, navigateur, …)

Plusieurs auteurs expliquent que la relation de confiance entre l'entreprise et le consommateur peut être affectée par le mécanisme de collecte utilisé pour obtenir les données. Ainsi, quand les consommateurs divulguent des informations de manière volontaire, cela n'aurait pas le même effet sur la confiance que lorsque la firme obtient les renseignements de manière discrète (Nowak et Phelps 1995, Caudill et Murphy 2000). Pourtant, quel que soit le mode de collecte envisagé, le recueil de telles données représente un enjeu stratégique majeur pour les entreprises. Car l'information client constitue aujourd'hui le nerf de la guerre.

[16] les cookies sont des petits programmes qui sont implantés sur les disques durs des internautes afin de pouvoir les identifier facilement quand ils reviendront sur le site.
[17] Elle est également utilisée, notamment, par la Grande Distribution qui recueille un grand nombre de données comportementales lors du passage en caisse auxquelles peuvent être associées des données liées à l'utilisation de cartes de fidélité.

2 La collecte de données clients : un enjeu stratégique

Nous verrons, dans un premier temps, en quoi consiste le marketing des bases de données et les raisons pour lesquelles il s'est ainsi développé (2.1), avant d'indiquer les liens existants entre ce type de marketing et le marketing relationnel (2.2). Nous terminerons en évoquant la question de la compatibilité entre le marketing relationnel et le respect de la vie privée des consommateurs (2.3).

2.1 Le développement du marketing des bases de données

Devant l'extrême compétitivité des marchés depuis les années 90, de nombreuses entreprises ont désormais compris que rien ne donne autant de pouvoir que de connaître les pratiques et préférences de chaque consommateur. Une grande majorité d'entre elles (du fabricant de produits agro-alimentaires au concessionnaire automobile, en passant par le supermarché de proximité) se sont ainsi mises à développer des bases de données constituées de l'ensemble des informations collectées sur leurs clients, actuels ou potentiels.
Selon Donnelley Marketing Inc.[18], au début des années 90, plus de la moitié des fabricants et distributeurs américains développaient ainsi une base de données et 10% envisagaient de le faire. De plus, 85% des managers interrogés dans cette étude estiment qu'il est nécessaire de posséder une base de données pour rester compétitif. Enfin, selon Stan Rapp de Cross Rapp Associates[17], il s'agit du développement marketing le plus important des années 90.

Même si cela nous permettrait de mieux cerner l'ampleur du phénomène, identifier la taille du marché des bases de données (ou de la collecte et de l'utilisation de données consommateurs), que ce soit au niveau français, européen ou mondial, est délicat (voire quasi-impossible), essentiellement pour des problèmes de définition et de délimitation de son périmètre. Plusieurs interrogations doivent ainsi être soulevées :

De quel marché parle-t-on réellement ?
S'agit-il uniquement de celui de la vente et de la location de fichiers ? Ne doit-on pas aussi tenir compte du marché des systèmes de gestion de bases de données relationnelles[19] ?

[18] Cité par Berry et al. (1994)
[19] Plus connus sous le sigle de SGBDR, il s'agit de logiciels qui permettent à l'entreprise de mettre en relation différentes bases de données et de disposer ainsi d'une vision complète de ses clients.

Que faut-il comptabiliser ?

Ne faut-il prendre en compte que les bases de données donnant lieu à une transaction (vente ou location), alors même que la création d'une base en interne constitue une source de valeur majeure pour l'entreprise ?

De même, doit-on seulement comptabiliser ce qui a trait au secteur marchand quand les associations, les organisations non gouvernementales et même l'Etat possèdent, chacun à leur niveau, des bases de données souvent plus riches que celles détenues par les entreprises ?

Autant de questions auxquelles il est bien difficile de répondre et qui conduisent à un certain flou quand il s'agit de déterminer le chiffre d'affaires dégagé par les organismes présents sur ce marché. Seul celui des SGBDR semble clairement défini[20], dominé il est vrai par des mastodontes tels Oracle (leader avec 6,5 milliards de dollars en 2005 soit environ 45% de parts de marché), IBM (21%) et Microsoft (17%).

Aussi, plutôt que d'avancer un chiffre qui serait loin de refléter la réalité, nous préférons présenter ci-après (cf. Figure 1.1) une cartographie de l'ensemble des acteurs concernés[21], de près ou de loin, par la constitution et le développement de bases de données.

Deux axes d'analyse peuvent alors être utilisés. Le premier distingue les acteurs marchands et non marchands. Le second concerne la « nature » du phénomène, suivant que l'on se place du point de vue du consommateur (i.e. la protection des données) ou de celui des entreprises (i.e. l'exploitation des données). C'est ce dernier axe qui est privilégié ici, le premier donnant lieu à une différence de couleur sur la figure (i.e. acteurs non marchands représentés en grisé). De plus, afin de différencier ces acteurs, nous les identifions par un numéro qui reflète, par ordre décroissant, leur « importance » au sein du processus. Nous faisons aussi apparaître l'ensemble des liens susceptibles d'exister entre ces différents acteurs. L'acteur n°1 (les entreprises) se situe ainsi au cœur du processus puisqu'il est celui qui entretient le plus de relations avec les autres.

Nous présentons brièvement ci-dessous les différents acteurs qui se retrouvent à la figure 1.1.

1) les entreprises : il s'agit de l'ensemble des organismes à vocation commerciale cherchant à collecter des données sur leurs clients dans le but de les utiliser en interne (i.e. pour leur propre compte) ou de les revendre ;

[20] Il était ainsi estimé à près de 15 milliards de dollars pour l'année 2005 (source IDC et Gartner Group), avec une croissance de 10 à 15% par an depuis le début des années 2000.

[21] Pour plus d'informations sur les différents acteurs présents sur ce marché, nous renvoyons à notre mémoire de DEA (p 33-40) et à l'ouvrage d'Arnaud Belleil (2001), intitulé « *e-Privacy* ».

2) les consommateurs : personnes sollicitées (directement ou non) pour fournir des données ;
3) les organismes de régulation : tels l'Etat, les organismes de contrôle (la CNIL en France) et les organisations professionnelles (comme la FEVAD), dont l'objectif est d'informer les différents acteurs et de réguler la collecte et l'utilisation des données ;
4) les groupes de pression : ensemble des « défenseurs de la vie privée » dont l'objectif est de sensibiliser l'opinion publique sur ces questions et de dénoncer les principaux abus ;
5) les tiers certificateurs : organismes chargés de vérifier que les entreprises respectent les règles en vigueur et d'approuver leurs pratiques par l'apposition d'un sceau ou d'un label ;
6) les fournisseurs de bases de données ou de solutions techniques : nous faisons référence ici aux organismes qui louent ou revendent des bases de données et à ceux qui offrent des solutions visant à les exploiter (comme les SGBDR) ;
7) les fournisseurs de solutions de protection : acteurs proposant aux consommateurs des moyens de « protéger » leurs données (logiciels anti-virus, solutions de cryptage, …).

Figure 1.1 Acteurs concernés par la collecte et l'utilisation de données consommateurs

Le fait que la création et le développement de bases de données concerne aussi bien les grands groupes que les petites entreprises montre à quel point le phénomène est important. Une des explications à l'intérêt croissant pour cette pratique serait liée à l'arrivée des nouvelles technologies de l'information et de la communication - plus connues sous le sigle de NTIC - qui rend la collecte, le stockage et l'utilisation des données plus aisés et moins coûteux. Le coût de la collecte et du stockage des données consommateurs a ainsi été divisé par 1000 depuis 1970 et continue de diminuer rapidement. De plus, une base de données de qualité fournit à l'entreprise qui la possède de nombreux avantages sur ses concurrents.

Selon Blattberg et Deighton (1991), elle offre ainsi différentes possibilité :

1) calculer la valeur du client (ou du prospect)
2) établir le dialogue avec les consommateurs
3) gagner du pouvoir sur les distributeurs
4) intégrer les campagnes marketing
5) améliorer la productivité du marketing

En résumé, le recours à une base de données permet de faire des économies, voire de dégager un bénéfice, à condition d'en faire bon usage. La valeur d'une telle base dépend ainsi à la fois de la qualité des informations qui s'y trouvent et de la manière dont elles sont utilisées. La majorité des données contenues dans ces bases sont issues des transactions - vues comme tout échange - entre l'entreprise et ses clients. La qualité des informations obtenues s'évalue alors à l'aune de plusieurs critères (Weible 1993) : exactitude, juste-à-temps, pertinence, complétude et fiabilité[22]. Ces données dégagent ensuite de la valeur à trois niveaux :

1) les revenus des transactions futures sont plus élevés (notamment de par la possibilité de vendre des produits complémentaires à ceux issus de la précédente transaction) ;
2) les coûts des transactions futures sont réduits (les informations collectées permettent de faire des offres mieux ciblées) ;
3) les données obtenues peuvent être compilées puis louées ou revendues à des tiers.

La valeur d'une base de données est donc particulièrement élevée quand les informations obtenues sur les consommateurs sont utilisées pour concevoir des offres en collaboration avec eux. En effet, ces offres auront d'autant plus de chance d'être acceptées qu'elles répondent spécifiquement aux besoins de ces clients.

[22] En anglais : *accuracy, timeliness, relevance, completeness and reliability.*

Cette connaissance précise des préférences et des caractéristiques des consommateurs, permise par le marketing des bases de données, apparaît ainsi comme l'un des fondements majeurs du marketing relationnel, paradigme émergent en marketing ces dernières années.

2.2 Du marketing des bases de données au marketing relationnel

De nombreuses entreprises reconnaissent aujourd'hui les données consommateurs comme le pilier du marketing des bases de données et du marketing direct, qui sont deux éléments clés du marketing relationnel. Ce dernier repose sur l'idée que tous les consommateurs ne sont pas égaux et qu'il est préférable - et plus profitable - de retenir les clients actuels plutôt que d'en acquérir de nouveaux. Ceci représente une évolution majeure pour de nombreux professionnels. Pendant longtemps en effet, les entreprises se sont focalisées sur l'acquisition de nouveaux clients en s'appuyant sur une segmentation du marché. Les bases de données étaient alors analysées dans le but de déterminer des profils de clients à cibler. Aujourd'hui, les nouvelles technologies permettent de s'adresser à des micro-segments, voire à des individus, en particulier et de calculer leur valeur client (*life time value*). Il est donc possible de classer les consommateurs en termes de profitabilité et de s'adresser en priorité aux clients les plus rentables. Il est donc primordial d'obtenir un maximum de données les concernant pour que le calcul soit le plus proche possible de la réalité.

Etant fondé sur l'idée de servir chaque consommateur individuellement, le marketing relationnel suppose de connaître parfaitement les goûts, besoins et préférences de celui-ci. Plusieurs chercheurs soulignent ainsi l'importance de collecter des données sur les clients au sein du processus CRM[23] (Zhang, Wang et Chen 2000). Ces informations sont en effet considérées aujourd'hui comme l'élément le plus différenciateur pour des firmes en concurrence (Meltzer 2000). De nombreux managers insistent aussi sur la nécessité de disposer de données fiables et complètes pour pouvoir mettre en place une démarche CRM avec succès. Or, outre les erreurs introduites par les salariés qui saisissent les données et manipulent ces bases, l'absence de fiabilité peut provenir des consommateurs eux-mêmes qui, de mieux en mieux informés sur les avantages et inconvénients liés aux pratiques de collecte de leurs données personnelles, se montrent parfois réticents à donner des informations vraies et complètes. Ce type de comportement peut conduire l'entreprise à une situation délicate. Elle peut ainsi manquer d'informations lui permettant de prendre des décisions pertinentes ou

[23] Pour *Customer Relationship Management*, en français : Management de la Relation Client (MRC)

prendre de mauvaises décisions sur la base d'informations erronées délibérément maquillées par le consommateur. La préoccupation de ce dernier en termes de respect de la vie privée semble ainsi constituer un obstacle majeur à l'obtention d'informations fiables à son sujet (Caudill et Murphy 2000) et doit donc être prise en compte lors de la collecte des données.

2.3 Marketing relationnel et respect de la vie privée

Si les efforts des entreprises pour développer des relations avec leurs clients les ont amenées à considérer la base de données comme un instrument permettant d'identifier et de servir au mieux leurs besoins, la croissance des pratiques de collecte et d'utilisation des informations a multiplié les inquiétudes des consommateurs quant au respect de leur intimité[24]. Dans un sondage, Punch (1996) trouve que 85% des répondants se sentent concernés par les menaces vis-à-vis du respect de la vie privée. D'autres études sur le sujet, réalisées principalement aux USA, soulignent que les consommateurs sont préoccupés par ce que les entreprises peuvent savoir d'eux et par la manière dont elles ont obtenu ces informations et les utilisent par la suite (sondages Harris Equifax 1992/95/96, Katz et Tassone 1990). Beaucoup d'entre eux auraient ainsi pris des mesures de protection, y compris des restrictions sur leurs échanges de données (Nowak et Phelps 1992, Wang et Petrison 1993).

L'un des aspects les plus controversés du marketing relationnel en B to C[25] semble donc concerner le respect de la vie privée. En effet, si l'objectif de cette méthode est de se rapprocher du consommateur, de mieux le connaître et d'engager avec lui une véritable relation, elle implique aussi de capturer et de manipuler des montants croissants d'informations le concernant. Son utilisation a donc généré une préoccupation croissante pour les questions relatives à la protection de la vie privée, notamment concernant les fichiers qui sont loués ou vendus sans le consentement des personnes concernées (Culnan 1995), mais aussi à propos des moyens mis en œuvre pour collecter les données.

D'une question stratégique, le recueil d'informations clients devient alors un enjeu éthique : comment, en effet, concilier cet objectif de collecte avec les réticences qu'il occasionne auprès des consommateurs ?

[24] Les notions d'« intimité », de « vie privée » et les différences entre les deux seront expliquées au chapitre 2.
[25] Sigle utilisé pour indiquer les relations entre les entreprises (*Business*) et les consommateurs (*Consumers*)

3 Collecte de données clients et respect de la vie privée : un enjeu éthique

3.1 Le respect de la vie privée : une attention croissante sur le sujet

Pour Westin (2003), quatre facteurs sont à l'origine des préoccupations croissantes du public vis-à-vis du respect de la vie privée : 1) le recours aux nouvelles technologies, 2) le climat social, 3) les groupes d'intérêt et les débats politiques, 4) les politiques organisationnelles et les lois en vigueur. Le premier facteur cité semble être celui qui trouve le plus d'écho au sein de la communauté scientifique. Ainsi, le fait que les entreprises soient aujourd'hui largement dépendantes des communications électroniques fait craindre que le droit au RVP soit de plus en plus difficile à maintenir[26].

A ces quatre facteurs, on peut en ajouter deux autres qui contribuent aussi à l'attention grandissante sur ce sujet. Le premier concerne la valeur des informations recueillies. Celle-ci semble en effet augmenter du fait de l'avantage compétitif qu'elles sont censées conférer. Or, les consommateurs sont de plus en plus conscients des bénéfices que les entreprises retirent de l'utilisation de leurs données. Selon Hagel et Rayport (1997), les cris émanant de l'opinion publique à propos de l'utilisation non autorisée des informations sont avant tout le résultat d'une connaissance accrue de la valeur de données divulguées autrefois gratuitement. Le second est lié à une confusion quasi permanente entre ce qui est éthiquement correct ou non. Le recours aux théories éthiques pourrait permettre de limiter une telle confusion.

3.2 Une approche par les théories éthiques

Selon Caudill et Murphy (2000), les théories éthiques peuvent se montrer utiles, notamment pour mieux comprendre les différences entre les recommandations édictées par les pouvoirs publics et les comportements effectifs des entreprises.

La première question qui se pose à ce sujet concerne la nécessité (ou non) de respecter la vie privée des consommateurs. Fried (1968) indique ainsi que la protection de la vie privée est instrumentalement utile car elle est nécessaire pour développer l'intimité et la confiance dans les relations. Rachels (1975) explique que les gens ont besoin de contrôler les informations les concernant pour maintenir une diversité dans leurs relations. Pour Nissenbaum (1998), les

[26] Cette question concerne notamment les droits des salariés à utiliser la messagerie de l'entreprise à des fins privées et le droit de regard qui peut être accordé à l'employeur à cet égard.

efforts entrepris par les entreprises pour collecter des informations sur leurs clients constituent alors une « violation morale de la vie privée ». Enfin, pour Mason (1986), les entreprises ont l'obligation morale de protéger les informations personnelles de leurs clients, notamment quand celles-ci circulent sur Internet.

Les hommes de marketing discutent souvent d'éthique en ayant recours aux théories téléologiques ou déontologiques. Les premières ont pour objet d'évaluer la moralité des actions en examinant leurs conséquences (Buchholz 1989). L'égoïsme et l'utilitarisme sont deux théories téléologiques couramment utilisées pour justifier les actions marketing. Selon l'égoïsme éthique, les conséquences d'un acte sont uniquement évaluées en termes de coûts et de bénéfices pour l'individu qui considère l'action. Si une action lui est bénéfique mais qu'elle « blesse » les autres, alors l'égoïste éthique choisit de la mener tout de même. L'utilitarisme évalue la moralité d'un acte en considérant les conséquences pour chacune des parties impliquées. Il s'agit ici d'obtenir le plus grand bien pour tous. Le côté positif et négatif de chaque acte est alors évalué et on choisira de mener l'action qui apporte le plus grand bénéfice net global.
De leur côté, les théories déontologiques n'évaluent pas les conséquences, mais plutôt la valeur morale de l'action dans la mesure où celle-ci met en jeu certains droits ou devoirs universels. Selon l'éthique déontologique de Kantian, un acte est moral s'il : 1) est universel (s'applique à toutes les personnes, dans toutes les situations, sans effets négatifs) ; 2) traite les êtres humains comme des fins et non comme des moyens (suppose que les personnes soient traitées avec respect et que leurs droits ne soient pas violés) ; 3) respecte l'autonomie des êtres humains (préserve le droit de chacun de décider librement).

Les entreprises justifient souvent le traitement des données personnelles des consommateurs sur la base de la théorie utilitaire. Elles reconnaissent que cela peut avoir des inconvénients mais elles estiment que, globalement, il en résulte un meilleur service qui profite à l'ensemble de la communauté. En réalité, cette justification semble davantage correspondre aux théories égoïstes, dans la mesure où elles ont généralement du mal à estimer les conséquences pour les consommateurs. De plus, elles ne répondent pas aux exigences des théories déontologiques. Ainsi, le fait pour une entreprise de retirer bénéfice de la vente des données à d'autres sociétés, sans avoir obtenu le consentement préalable des consommateurs, constitue une violation des trois principes de Kantian.

Deux conflits éthiques majeurs opposant les entreprises et les consommateurs à propos du respect de la vie privée se dessinent alors : 1) des vues différentes à propos de qui est propriétaire et/ou contrôle l'information ; 2) des frictions entre les droits du consommateur en matière de respect de la vie privée et d'autres droits. Ces conflits sont exacerbés par le manque d'informations dont dispose le consommateur quant à l'utilisation qui est faite de ses données et aux moyens de recours à sa disposition en cas d'utilisation abusive.

3.3 Propriété des données et respect de la vie privée : un conflit d'intérêt

Pour toute transaction impliquant la fourniture de données personnelles, consommateurs et entreprises sont souvent en désaccord en ce qui concerne la jouissance des informations échangées et le type d'échange acceptable, compte tenu de la situation. Alors que la majorité des managers pensent que ces données leur appartiennent et qu'ils peuvent les utiliser comme bon leur semble, de nombreux consommateurs considèrent au contraire que ce droit de propriété leur revient (Nowak et Phelps 1992). De plus, ces derniers sont également à la recherche de mesures de protection contre les utilisations non désirées de leurs données, sans pour autant compromettre les gains de choix et de flexibilité qu'offrent, notamment, les techniques d'achat à distance (Cespedes et Smith 1993).

Parmi les explications à ce désaccord profond, on trouve une différence de point de vue quant à l'origine des informations collectées. Ainsi, Davis (1997) considère que les consommateurs ont un droit de propriété sur les données achetées par les entreprises et vendues à des fins commerciales suggérant ainsi, qu'étant - à l'origine - détenteurs de l'information, ils doivent être dédommagés du bénéfice qu'elles en retirent. De même, pour Westin (2003) « le droit de décision envers ses propres données doit être vu comme un droit de propriété ». Cette conception peut se justifier si l'on considère les données personnelles comme une extension de la personnalité de l'individu. Le problème vient du fait que ces informations sont souvent formées en relation avec les autres. Ainsi, le comportement de navigation sur Internet est le produit de l'interaction entre l'individu et les sites qu'il visite. Dès lors, Pour Foxman et Kilcoyne (1993), si les entreprises estiment posséder l'information collectée auprès des consommateurs c'est parce qu'elles considèrent que celle-ci dérive - au moins en partie - de la transaction, dont elles sont à l'origine.

Dans les faits, la frontière entre le moment où l'information est détenue par l'individu et où elle devient la possession de l'entreprise est souvent floue. Milne et Gordon (1993) indiquent

ainsi que la propriété de l'information n'est pas absolue. L'enjeu dans ce domaine est donc de trouver un équilibre entre différents droits en opposition. A ce sujet, au moins quatre sources de conflits ont d'ores et déjà été identifiées (Goodwin 1991). Il faut ainsi trouver un compromis entre le droit au respect de la vie privée du consommateur et : 1) d'autres attentes en termes de services ; 2) d'autres droits du consommateur (en termes d'information, de sécurité, …) ; 3) les coûts résultant de la protection ; 4) d'autres valeurs sociétales. Cependant, le conflit majeur réside dans le droit du consommateur à protéger sa vie privée et le droit à l'appropriation des informations recueillies par l'entreprise. On se trouve en effet à la croisée de plusieurs libertés qui se trouvent être en contradiction : la liberté du commerce et des échanges d'un côté et la liberté individuelle de l'autre (Rigaux 2000). De plus, on touche ici un point particulièrement sensible dans la mesure où, de tous les droits de l'individu, celui lié au respect de la vie privée apparaît comme l'un des plus importants. Il s'agit en effet d'un droit basique autour duquel de nombreuses libertés se fondent (Rotemberg 1993). Or, celui-ci est moins bien pris en compte sur Internet, puisqu'il est plus facile de recueillir des données par ce biais que par les moyens traditionnels.

3.4 Les spécificités de l'enjeu éthique sur Internet

La nature interactive d'Internet présente un environnement unique dans lequel la préoccupation pour la vie privée semble renforcée (Korgaonkar et Wolin 1999). Ainsi, dans une étude menée sur le sujet, Jacobs (1997) trouve que 70% des répondants se disent davantage concernés par les problèmes de respect de la vie privée sur Internet que sur les médias traditionnels. Cette question est donc primordiale pour les internautes. Pour Hoffman, Novak et Peralta (1999) la raison pour laquelle les consommateurs sont plus préoccupés par le RVP sur Internet est liée au fait que le besoin de contrôle et de protection de leurs données personnelles est exacerbé sur la toile. Dans leur étude en effet, 87% des internautes interrogés (de nationalité américaine) estiment avoir droit à un contrôle complet des informations démographiques que les sites capturent et 71% pensent qu'il devrait y avoir de nouvelles lois pour protéger leur vie privée en ligne.

Pour Petty (2000), chaque canal de communication et/ou de vente implique différents coûts (en termes de temps, d'argent, etc …) et types d'intrusion pour le consommateur. Dès lors, si Internet a engendré de nouvelles préoccupations par rapport aux autres médias, cela est notamment dû à la facilité avec laquelle les informations peuvent être collectées, stockées et échangées par ce biais (Caudill et Murphy 2000, Richards 1997, Sheehan et Hoy 2000). C'est

aussi lié au nombre élevé de sollicitations commerciales envoyées par email, grâce au faible coût de l'envoi (Shiman 1996) et à la plus grande capacité à cibler l'offre. On peut y ajouter une autre raison liée aux particularités de l'environnement en ligne. Ce dernier s'ajoute en effet à l'aspect contextuel de la protection de vie privée, de par la facilité avec laquelle un internaute peut devenir lié à une entité. On pense ici à la possibilité, en un seul clic, d'accéder à un site dont le nom était précédemment inconnu.

En résumé, Burgoon (1982) estime que les menaces que font peser les nouvelles technologies sur le respect de la vie privée des consommateurs sont liées à cinq facteurs principaux :
1) le degré de contrôle que les individus exercent sur la divulgation et l'utilisation ultérieure de leurs données personnelles ;
2) la quantité d'informations possédées par les autres : quand les consommateurs ont l'impression que cette quantité est minime, la vulnérabilité perçue est limitée ;
3) le nombre d'entités et/ou de personnes possédant ces informations : plus il est élevé, plus l'individu court le risque qu'elles soient « détournées » ;
4) le type d'informations possédées : plus l'information est considérée comme sensible, plus l'individu est vulnérable si elle est dévoilée publiquement ;
5) la nature de la relation avec l'entité possédant ces informations. Or, l'identification des entreprises collectant les informations est de plus en plus difficile. De plus, connaître l'entreprise les ayant collectées ne signifie pas que l'on connaît celle qui les réutilisera.

L'ensemble de ces facteurs conduit à une préoccupation accrue pour le respect de la vie privée sur Internet, qui aboutit, d'une part, à une réticence à fournir des informations personnelles sur le web et a, d'autre part, des conséquences non négligeables sur le commerce en ligne. Tout ceci a incité managers et chercheurs à s'intéresser au problème. En effet, aujourd'hui encore, et même si cela est de moins en moins vrai[27], un certain nombre de consommateurs ne font pas suffisamment confiance aux e-commerçants ou aux e-fournisseurs de services pour s'engager dans des relations d'échange avec eux. Si les individus lisent les journaux, recherchent un emploi ou comparent les prix sur Internet, une partie d'entre eux n'achète toujours pas en ligne ou se refuse à délivrer des informations personnelles, à cause des craintes relatives au respect de leur vie privée (McGraw 1999, Hershel et Andrews 1997). Pour Wang et Petrison (1993) et Lohse et al. (2000), cette peur concernant les questions de RVP est même l'une des principales raisons expliquant la réticence à commercer sur le Net.

[27] La progression croissante du commerce en ligne, notamment en France, vient en effet quelque peu démentir ce propos.

Pour Hoffman, Novak et Peralta (1999), ce manque de confiance proviendrait de la faiblesse de contrôle perçu par les internautes concernant l'accès à leurs données personnelles, durant le processus de navigation. D'après leurs résultats, la première préoccupation des internautes lorsqu'ils surfent sur le Net est liée au contrôle de leurs informations personnelles, plutôt qu'aux risques de l'achat à distance. De plus, ces préoccupations semblent influencer le désir d'achat en ligne. Les principales raisons données par les internautes non acheteurs en ligne ne sont en effet pas fonctionnelles (i.e. liées au prix, au temps de livraison, etc …) mais liées à ces questions de contrôle sur les données.

Si la question du contrôle des données personnelles est essentielle pour les consommateurs, c'est à cause de la vulnérabilité qui en découle, ces données pouvant être dévoilées quasi publiquement. Au cœur du problème, se pose donc la question de savoir ce que représente réellement le respect de la vie privée, notion que nous allons définir au sein du chapitre 2.

Conclusion du chapitre 1

Il existe plusieurs modes de collecte de l'information client. Chacun d'entre eux possède des spécificités que ne sont pas sans conséquence quant à la manière dont ils sont perçus par les individus. Dans cette recherche, nous nous intéresserons plus particulièrement aux collectes directes, c'est-à-dire faisant l'objet d'une sollicitation de la part de l'entreprise et d'une réponse de la part du consommateur sollicité. De plus, nous focaliserons sur les collectes d'informations en ligne, qui tendent à exacerber les craintes des consommateurs (internautes) quant à l'utilisation ultérieure de leurs données personnelles, une fois celles-ci divulguées.

Face à la valeur croissante de l'information client, les entreprises doivent faire face à deux défis majeurs lorsqu'elles souhaitent solliciter les consommateurs : un défi stratégique et un défi éthique.

Le défi stratégique concerne la mise en place et le développement d'un marketing des bases de données qui permet, non seulement de calculer la valeur du client mais aussi d'améliorer la productivité des opérations marketing. Cet élément permet ensuite à l'entreprise de déployer une politique relationnelle efficace, visant à servir chaque consommateur individuellement. Ceci nécessite alors de connaître les goûts et préférences de chacun et donc de collecter, à chaque fois qu'elle en a l'occasion, de nombreuses informations. Une telle pratique s'avère cependant difficilement compatible avec le droit au respect de la vie privée de chaque citoyen.

La nécessaire conciliation entre les objectifs des managers et les réticences des consommateurs représente le deuxième défi auquel les entreprises doivent faire face : le défi éthique. L'attention croissante du public pour les questions liées au respect de la vie privée ont amené professionnels et chercheurs à réfléchir aux moyens à mettre en œuvre pour respecter les droits de chacun. Les théories éthiques représentent, à cet égard, une explication possible aux différences de points de vue entre les principales parties à l'échange. Toutefois, un conflit d'intérêt majeur se dessine quant à savoir qui est réellement propriétaire de l'information. Ce conflit tend d'ailleurs à s'exacerber sur Internet. Les spécificités de ce média semblent en effet renforcer les préoccupations des consommateurs quant au respect de leur vie privée, notion que nous allons à présent définir plus précisément.

Partie I - COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

CHAPITRE 2
PRÉAMBULE SUR LA NOTION DE VIE PRIVÉE

« La vie privée est comme l'oxygène, nous ne l'apprécions réellement que lorsque nous en sommes privés » ; « Sans espace privé, la vie publique serait impossible ».

Charles Sykes (1999)

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

Chapitre 1 : Les enjeux liés à la collecte de données clients

Chapitre 2 : Préambule sur la notion de vie privée

Chapitre 3 : Dévoilement de soi et divulgation de données personnelles

Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)

Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)

Chapitre 7 : Formulation du cadre conceptuel général de la recherche

Chapitre 8 : Méthodologie de la recherche

Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)

Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)

Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 2

Nous avons tous l'idée intuitive que certains aspects de notre vie sont privés. L'idée de séparer le côté privé du côté public existe d'ailleurs depuis les temps anciens, sans que le concept de protection de la vie privée n'ait été pour autant utilisé. Aujourd'hui encore, ce concept n'existe pas dans de nombreuses cultures (i.e. arabe, allemande, japonaise et russe notamment), même si l'on en trouve l'idée sous-jacente sous d'autres termes.

L'idée de protéger les aspects privés de notre vie est donc, dans l'ensemble, largement répandue. Dans beaucoup de pays, notamment ceux de l'Union Européenne, cela se manifeste d'ailleurs par la reconnaissance d'un droit au respect de sa vie privée, qui est fondamental et inaliénable. Pour de nombreux auteurs, l'existence d'un tel droit est le fondement même de la démocratie car, sans vie privée, l'homme serait constamment surveillé et perdrait donc sa liberté (Sykes 1999[28]). Warren et Brandeis (1890) considèrent ainsi le droit au Respect de la Vie Privée comme « le plus valorisé par les hommes civilisés ». Il s'agit donc d'un besoin humain majeur, qui se manifeste de différentes façons selon les situations. Les recherches menées sur le sujet, dans le domaine de l'anthropologie, de l'histoire du droit, de la philosophie, de la sociologie ou de la psychologie, se sont d'abord focalisées sur l'étude de ce concept, dans le cadre de relations interpersonnelles. Or, force est de constater que les menaces qui pèsent aujourd'hui sur la vie privée sont largement le fait des relations commerciales qu'entretiennent individus et entreprises. Une partie significative de ces menaces concernent ainsi la collecte et l'utilisation des données à caractère personnel des clients et/ou prospects.

Nous verrons, dans une première partie, la définition et les caractéristiques de la protection de la vie privée. Nous soulignerons ensuite l'aspect polymorphe du besoin de protection de la vie privée. Nous terminerons en évoquant un des aspects essentiels de la protection de la vie privée et sur lequel nous nous focaliserons dans le reste de ce document, à savoir la protection des données à caractère personnel.

[28] Cité par Ingargiola (2002), Realizing privacy in the information age : an exercise in uncovering the values shaping the online privacy debate, Doctoral Dissertation, Georgetown University, Washington DC

1 Définition et caractéristiques de la protection de la vie privée

Nous donnons tout d'abord les raisons pour lesquelles la protection de la vie privée est si difficile à définir (1.1), avant d'identifier les principales définitions et dimensions identifiées au sein de la littérature (1.2). Nous évoquons ensuite les principales fonctions de la protection de la vie privée (1.3) puis terminons par une distinction entre ce concept et d'autres concepts proches et/ou liés (1.4).

1.1 La protection de la vie privée : un concept flou

La protection - ou le respect - de la vie privée (appelée *privacy* chez les anglo-saxons) est une notion large qui comprend entre autres : la liberté de pensée et d'action, le contrôle de son corps, la solitude dans sa maison, le contrôle des informations personnelles, l'absence de surveillance, la protection de sa réputation. Ce concept est couramment utilisé pour décrire à la fois un besoin d'espace, un besoin de séparation visuelle, physique ou psychologique, et le contrôle sur les possessions et les informations personnelles.

Etablir une définition précise de ce concept est particulièrement délicat, car il n'y a pas d'accord général dans ce domaine. Il existe en effet de multiples conceptualisations, selon la considération - légale, politique, sociale - retenue. De nombreux auteurs soulignent cette difficulté de parvenir à une définition claire de la notion de protection de la vie privée. Westin (1967) fait d'ailleurs remarquer qu'il s'agit là d'une situation quasi unique dans la mesure où « peu de valeurs aussi fondamentales dans la Société que celle de respect de la vie privée sont restées sans définition au sein de la théorie sociale »[29]. Il est possible d'avancer plusieurs explications à cela.

La première catégorie d'explications est liée au fait que ce concept souffre de ce que Margulis (1977) qualifie de « *definitional ambiguity and vagueness* »[30]. Pour Miller (1971), cette notion est difficile à appréhender car elle est « désespérément vague et évanescente ». On a donc du mal à la définir car elle est floue et ambigüe.

[29] La phrase d'origine est : « *few values so fundamental to society as privacy have been left so undefined in social theory* » (Westin 1967), cité dans Solove (2002)
[30] Qu'on peut traduire par : « ambiguïté et imprécision définitionnelle »

Pour d'autres auteurs, dont Judith Thomson (1984), le problème vient du fait qu'il ne s'agit pas d'un concept « distinct » mais qu'il est compris dans plusieurs autres concepts ou droits[31]. Robert Post (2001) indique à cet égard qu'« il s'agit d'une valeur si complexe, si engluée dans des dimensions concurrentes et contradictoires, si pleine de sens variés et distincts que [je] désespère parfois de pouvoir en parler de manière utile » [32].

Une dernière explication à la difficulté de saisir ce concept de protection de la vie privée est liée à son aspect largement pluridisciplinaire. Cette notion apparaît en effet dans de nombreuses disciplines incluant l'anthropologie (Roberts et Gregor 1971), l'histoire (Flaherty 1984), la sociologie (Schwartz 1968), les ressources humaines (Schein 1977) et beaucoup plus récemment le marketing. Elle se trouve, dès lors, associée à de multiples sens, incluant nombre d'antécédents et de contextes ainsi qu'une variété de motivations et de réponses. Ainsi, quand les philosophes voient la protection de la vie privée comme une fonction nécessaire de l'existence humaine, les psychologues soulignent son importance pour le développement de la personne. De même, alors que les juristes la considèrent comme un droit fondamental, les sociologues se focalisent sur son importance pour l'établissement des relations humaines. A cela, s'ajoute le fait que la signification du concept varie aussi à l'intérieur d'une même discipline. Tel est le cas par exemple de la psychologie, au sein de laquelle les définitions diffèrent sensiblement. Si certains auteurs mettent en effet l'accent sur l'isolement, le fait d'éviter le contact avec autrui, d'autres insistent plus particulièrement sur le contrôle de l'accessibilité à soi (Jourard 1966 ; Shils 1966 ; Westin 1970).

Malgré ces différences, il existe un point commun à toutes ces littératures, celui du contrôle de l'individu sur la divulgation de ses informations personnelles qui est, pour la majorité des auteurs – dont Shils (1966) et Westin (1967) - et ce, quelle que soit leur discipline d'origine, une des dimensions majeures de la protection de la vie privée. D'autres dimensions existent cependant, qu'il convient à présent d'identifier.

1.2 Les dimensions de la protection de la vie privée dans la littérature

Johnson (1992) et Solove (2002) ont cherché à synthétiser les conceptualisations de la vie privée trouvées dans la littérature. De ces travaux, ressortent six définitions principales du

[31] Voir paragraphe 1.4 de ce chapitre

[32] La phrase d'origine est : « *Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether I can usefully addressed at all* » (Post 2001), cité dans Solove (2002).

concept de protection de la vie privée, que nous présentons ci-dessous, avant de souligner les limites inhérentes à chacune d'elle.

Le concept de vie privée correspond ainsi, selon les auteurs, à l'une des acceptions suivantes :

1) le droit d'être laissé seul et/ou tranquille [33]
2) un accès limité à soi (pouvoir empêcher l'accès des autres lorsqu'il n'est pas désiré)
3) le secret (pouvoir cacher certains faits ou informations aux autres)
4) le contrôle sur ses informations personnelles
5) la protection de sa personnalité, de son individualité et de sa dignité
6) l'intimité (contrôle ou limitation de ses relations avec les autres)

Malgré son intérêt, et notamment son souci d'exhaustivité, deux remarques doivent être faites à propos de cette classification. Tout d'abord, certaines définitions concernent davantage les moyens d'obtenir la protection (conceptions 3 et 6 notamment) alors que d'autres s'intéressent plutôt aux objectifs sous-jacents (conception 5). Ensuite, on s'aperçoit que toutes ces conceptions ne sont pas indépendantes ; parfois, elles se chevauchent.

Ainsi, la conceptualisation d'un accès limité à soi (2) correspond à une formulation plus sophistiquée et plus globale du droit à être laissé seul (1), fondement originel de la protection de la vie privée aux USA. En effet, l'accès limité à soi ne réside pas uniquement dans l'aspect solitude, dans le fait d'être séparé des autres. Il comprend aussi la protection contre toute intrusion non désirée, que celle-ci émane de la presse, du gouvernement ou d'entreprises.

La conception de la protection de la vie privée comme droit au secret (3) évoque la possibilité de dissimuler certains faits pour éviter qu'ils ne soient ensuite dévoilés publiquement. Jourard (1966) reconnaît ce principe du secret dans sa définition du concept. Pour lui, la protection de la vie privée est le résultat de « la volonté d'une personne de cacher aux autres des informations sur ses actions, ses expériences passées et présentes et ses intentions futures ». D'autres auteurs considèrent cependant cette conceptualisation du secret comme trop étroite. Pour eux en effet, la protection implique davantage que le simple fait de ne pas se dévoiler, elle contiendrait aussi la capacité de s'assurer que les informations sont utilisées dans un but acceptable pour l'individu. Comme l'indique Inness (1992), la protection de la vie privée, conceptualisée comme un secret, omet la dimension de contrôle. Or, cette dernière est à l'origine de la théorie prédominante en matière de protection de la vie privée, celle qui touche au contrôle de ses données personnelles[34] (4).

[33] En anglais : "*the Right to be let alone*" (Warren et Brandeis, 1890)
[34] Un paragraphe entier étant consacré ultérieurement à cette notion, nous ne nous étendrons pas dessus ici

L'avant dernière approche s'intéresse à la protection de l'individu, de son individualité, de sa dignité et de son autonomie (5). Cette conception diffère des précédentes en ce sens qu'elle est construite autour d'une idée normative de la vie privée, liée à l'intégrité de la personnalité. La protection correspond ici au respect de la personne en tant qu'individu capable de faire des choix. Certains auteurs critiquent cette conception qui confond la vie privée avec d'autres concepts proches, comme la liberté et l'autonomie. En réalité, cette conception s'attache davantage à expliquer pourquoi nous valorisons la vie privée (i.e. pour protéger son individualité, sa dignité et son autonomie) plutôt que d'en donner une définition.
La dernière conception (6) reconnaît que la protection de la vie privée n'est pas seulement essentielle à la définition du soi – qui caractérise la conception précédente – mais qu'elle l'est aussi pour le développement des relations humaines. Selon cette définition, nous valorisons la vie privée afin de pouvoir maintenir le niveau désiré d'intimité dans nos relations avec les autres. Pour Fried (1968), l'intimité représente « le droit de ne pas partager d'informations personnelles sur ses actions, ses croyances et ses émotions avec tout le monde ».

Si l'on écarte les notions de solitude (1), de secret (3) et de dignité (5), la première parce qu'elle est comprise dans la dimension d'accès limité à soi (2), la seconde parce qu'elle est trop étroite et la troisième parce qu'elle est plus une raison à la valorisation de l'intimité, les réponses les plus courantes à la définition de la protection de la vie privée tombent alors dans 3 catégories principales : le droit d'être laissé seul, le contrôle de l'accès à ses données personnelles et le contrôle des interactions avec autrui. Celles-ci correspondent aux dimensions majeures de la protection de la vie privée telles qu'identifiées par un grand nombre de chercheurs (Stone et Stone 1990 ; Evans, O'Malley et Patterson 1998 ; Charters 2002) et représentent autant de préoccupations possibles pour un individu, en cas de non respect. Il faut souligner que le fait de retenir une classification en 3 catégories plutôt que quatre, comme le font certains auteurs (Wolfe et Golan 1976 ; Petrison et Wang 1995), offre l'avantage de limiter les imbrications entre les différentes dimensions, tout en restant exhaustif. En outre, il est possible de caractériser ces catégories en insistant sur le type de protection concerné. Trois aspects majeurs ressortent ainsi de la littérature - un aspect physique, un aspect informationnel et un aspect psychologique et relationnel, correspondant parfaitement à chacune des 3 conceptions retenues.

Nous présentons plus précisément ci-dessous les trois dimensions majeures du Respect de la Vie privée identifiées dans la littérature, en nous basant sur la classification proposée par Stone et Stone (1990). Par ailleurs, nous insistons sur l'aspect concerné par la protection de

l'individu, qui rend la classification plus explicite, notamment managérialement parlant. L'ensemble est résumé dans le tableau 2.1 ci-après.

- « le droit d'être laissé seul » (aspect physique)

Nous pourrions aussi appeler cette dimension « retrait », dans la mesure où elle touche à l'aspect physique de la protection. Ainsi, un individu qui refuse d'ouvrir sa porte à un démarcheur, peut le faire parce qu'il considère son domicile comme un lieu privé, un endroit de retrait pour la vie familiale et le développement personnel. La maison est d'ailleurs considérée par de nombreux chercheurs en psychologie comme une métaphore des aspects publics et privés du soi (Goodwin 1992).

Cette première dimension correspond donc à la fois au droit d'être laissé seul et au contrôle de l'accès à soi (définitions 1 et 2 de Solove 2002). Elle reflète la possibilité de s'isoler, de placer une barrière physique entre soi et les autres et repose sur l'idée qu'on ne doit pas agir d'une manière qui constitue une intrusion dans la vie de quelqu'un (McCloskey 1980). Cet aspect de la protection est donc fréquemment confondu avec le concept de liberté qui, bien que proche, n'en est pas moins différent. La liberté est en effet le droit d'être « dégagé » de toute interférence physique et de toute coercition (McCloskey 1980). Une perte de liberté n'est donc ni un précurseur ni une condition d'invasion de la vie privée (Boatright 2000).

Le fait de se mettre à distance des autres, par le biais de barrières physiques, fait aussi écho aux notions d'espace personnel et de territorialité, cette dernière correspondant à la défense des propriétés et possessions (Brown 1987). Cet aspect spatial de la protection est lui-même axé sur l'idée de séparation entre sphère publique et sphère privée. Edney et Buda (1976) insistent toutefois sur la nécessité de distinguer les concepts de territorialité et de vie privée en expliquant que le premier contribue à l'identité de soi alors que le second concerne plutôt les sentiments de liberté et d'autonomie.

Pour résumer cette dimension, nous retiendrons la phrase de Schwartz (1968) qui définit la vie privée comme un « mode institutionnalisé de retrait ». Selon lui, protéger sa vie privée revient à contrôler le nombre de contacts avec autrui. Le contrôle correspond alors à un « réseau de mécanismes comportementaux utilisés par les individus pour parvenir au niveau désiré d'interaction sociale » (Altman 1977).

- « le contrôle de l'information sur soi » (aspect « informationnel »)

Nous pourrions aussi, à l'instar de nombreux chercheurs (Petrison et Wang 1995, Charters 2002), appeler cette dimension « contrôle de l'accès aux informations personnelles », dans la mesure où elle concerne l'aspect « informationnel » de la protection. C'est, par exemple, le

cas d'un individu qui refuse de délivrer des informations sur sa santé lors d'un entretien d'embauche ou d'un consommateur qui ne souhaite pas donner son âge à une entreprise de crédit ou d'assurance, de peur qu'elle ne lui refuse le contrat.

De nombreux théoriciens réduisent le concept de protection de la vie privée au droit de contrôler l'accès à ses données personnelles. Celui-ci est alors conçu comme le droit pour un individu de déterminer dans quelle mesure des informations le concernant peuvent être révélées aux autres (McCloskey 1980). Pour cet auteur, les personnes peuvent cependant renoncer à ce contrôle dans deux cas précis : quand elles y accordent peu de valeur (importance accordée aux normes) ou lorsqu'elles sont devenues résignées, du fait d'un non respect continuel de ce droit (effet de l'expérience). Selon cette perspective, les individus sont donc protégés quand ils sont capables de contrôler les impressions suscitées par les données personnelles qu'ils fournissent. Cela s'opère par le biais d'une procédure de divulgation sélective d'informations (Jourard 1966) - qui empêche autrui d'acquérir des données fiables et complètes (Derlega et chaikin 1977) - et par la régulation de la dissémination de ces données. C'est cette dimension qui est majoritairement concernée lors de la sollicitation de données personnelles des consommateurs par les entreprises.

- « la liberté vis-à-vis du contrôle des autres » (aspect psychologique et relationnel)

Cette dimension correspond à la notion de limitation et de contrôle de ses interactions sociales ainsi qu'à la protection de son individualité (définitions 5 et 6 de Solove 2002). Elle inclut donc, au-delà de l'aspect relationnel, un aspect psychologique. On fait référence ici à la pression pouvant être exercée par l'autre - tout mécanisme utilisé pour déterminer nos pensées et nos attitudes - et aux moyens de s'en protéger. La protection peut ainsi prendre la forme d'une discussion avec un inconnu, sous couvert d'anonymat.

Cette approche se base notamment sur les travaux de Goffman (1959), qui explique que lorsque les autres ont des informations sur vous, ils deviennent capables de vous contrôler. Pour Kelvin (1973), le respect de la vie privée vise donc à assurer l'indépendance des individus, dans toute situation qui les rendraient vulnérables au pouvoir des autres. Tolchinsky et al. (1981) estiment que la possibilité de contrôler l'accès à soi permet à l'individu de se protéger du pouvoir des autres à entreprendre une action aux conséquences potentiellement négatives pour lui. A travers cette protection, l'individu restreint le degré auquel les autres peuvent avoir accès à lui et ainsi la possibilité pour ces derniers d'observer et de contrôler ses pensées, ses attitudes et comportements. Elle permet donc d'améliorer son autonomie et de réduire sa vulnérabilité (Jourard 1966 ; Margulis 1977).

Tableau 2.1 Les trois dimensions majeures du droit à la protection de la vie privée

Dimensions	Explications	Auteurs de référence
« Droit d'être laissé seul » ou « Droit de retrait »	Possibilité de s'isoler, de placer une barrière physique entre soi et l'extérieur et de limiter ainsi le nombre de contacts avec les autres	McCloskey (1980) Schwartz (1968) Altman (1977)
« Contrôle de l'information sur soi » ou « Contrôle de l'accès à ses informations personnelles »	Droit de déterminer dans quelle mesure les informations nous concernant peuvent être révélées aux autres	Westin (1967) Stone et al. (1983)
« Liberté vis-à-vis du contrôle des autres »	Contrôle de ses interactions sociales et protection de son individualité	Goffman (1959) Kelvin (1973)

Bien que distinctes, ces 3 dimensions n'en sont pas moins complémentaires. Laufer et Wolfe (1977) indiquent en effet que des individus décrivent leur expérience d'invasion de la vie privée comme des situations dans lesquelles ils ont perdu à la fois le contrôle de leurs interactions avec les autres et le contrôle de leurs informations personnelles. Pour être complète, chaque définition conceptuelle de la protection de la vie privée doit donc contenir des éléments issus de chacune des perspectives évoquées ci-dessus. La définition que nous proposons ci-après répond à cette exigence. Nous considérons en effet ce concept comme « *un contrôle de l'accès à sa personne et/ou à son environnement personnel, à travers la possibilité de s'isoler, de contrôler son propre désir de se dévoiler ainsi que l'utilisation que les autres peuvent faire des informations nous concernant, à travers, enfin, la possibilité de contrôler les intrusions et le pouvoir exercé par les autres* ». Cette définition offre en outre l'avantage de reprendre une notion considérée comme essentielle par la majorité, si ce n'est la totalité, des auteurs étudiant ce phénomène, la notion de contrôle.

Après avoir présenté les principales dimensions liées à ce concept et en avoir donné une définition, il convient à présent de s'interroger sur les différentes fonctions qui incombent à la protection de la vie privée.

1.3 Les principales fonctions de la protection de la vie privée

La protection de la vie privée est un besoin humain essentiel : elle permet de gérer à la fois ses activités personnelles et ses interactions sociales. Dans l'ensemble, le terme de vie privée est d'ailleurs considéré, soit comme un élément majeur de la dignité humaine (Benn 1971,

Bloustein 1964), soit comme un « élément sacré de la personne » (Simmel 1971), ce dernier terme impliquant l'existence de l'individu en dehors du groupe.

Plusieurs chercheurs ont tenté d'énumérer les fonctions que permet d'atteindre le respect de la vie privée. Parmi les classifications les plus abouties, on trouve celle de Margulis (1977) et celle de Westin (1970) qui, tout en étant complémentaires, disposent de certains éléments en commun. Le premier distingue ainsi 3 fonctions principales que sont : 1) le management des interactions avec les autres (chaque individu recherche le niveau d'interaction sociale qui offre le plus grand bénéfice) ; 2) la définition du soi (en rapport avec les autres) ; 3) l'identification de soi (mécanisme permettant de se distinguer des autres). Le second énumère pour sa part 4 fonctions majeures : 1) l'autonomie personnelle ; 2) le « dégagement émotif »[35] ; 3) l'évaluation de soi ; 4) la mise en place de communications limitées et protégées.

Pour ces deux auteurs, une des fonctions principales de la protection de la vie privée est donc le management des interactions avec son environnement social qui implique, selon Westin (1970), la mise en place de communications limitées et protégées. Ceci suppose l'existence de barrières entre soi et autrui et la recherche d'un équilibre entre l'ouverture et le fait de cacher des informations sur soi. Cette régulation constitue d'ailleurs un élément essentiel à la définition de soi, autre fonction majeure de la protection. En effet, la capacité à réguler les frontières autour de soi (ou l'incapacité de le faire) fournit à la personne de précieuses indications quant à la définition de son identité et lui permet de développer son sens de l'individualité. Il faut cependant noter que ce n'est pas tant le processus d'inclusion ou d'exclusion des autres mais la capacité de le faire (ou non) qui contribue à la définition de soi. Westin (1970) indique aussi qu'un des buts de la protection est l'auto-évaluation. Elle offrirait en effet l'opportunité pour la personne d'assimiler ses expériences passées et d'examiner les possibilités de relations futures avec les autres. Enfin, la protection de la vie privée serait au service de l'affirmation de son identité. Westin (1970) souligne en effet que l'autonomie, une des fonctions majeures de la protection selon lui, inclut un sens personnel de l'intégrité et de l'indépendance ainsi que la capacité d'éviter de se faire manipuler par autrui. Elle est donc un instrument permettant l'« identification de soi » (Margulis 1977). Plusieurs chercheurs décrivent ainsi l'invasion de la vie privée comme particulièrement blessante parce qu'elle détruit l'autonomie individuelle, le respect de soi et la dignité.

[35] En anglais : *emotional release*

Dans l'ensemble donc, la perception de vie privée s'avère essentielle pour le concept de soi, le développement et le maintien de son identité et de sa responsabilité personnelle. Différentes études soulignent d'ailleurs l'importance de ce concept dans le cadre du processus de développement de l'enfant. Plusieurs auteurs, dont Laufer et Wolfe (1977), soulignent ainsi l'importance de fournir aux enfants le contrôle de leur environnement, de manière à faciliter le développement de leur identité et de leur autonomie.

Dans les modèles de motivation fondés sur la théorie de la valeur attendue[36], les individus sont présumés se comporter de manière à maximiser les conséquences positives et à minimiser les conséquences négatives de leurs comportements. En écho à ce principe, la littérature sur la protection de la vie privée indique qu'à travers celle-ci, les individus peuvent obtenir des résultats positifs comme : un meilleur bien-être physique et psychologique (Jourard 1966), la protection de soi et de l'identité (Derlega et Chaikin 1977 ; Laufer et Wolfe 1977 ; Margulis 1977), des avantages compétitifs dans l'atteinte de buts économiques et sociaux, un meilleur sens de l'individualité et de l'autonomie (Derlega et Chaikin 1977), un sentiment de sécurité et de bonheur ainsi que d'autres états affectifs positifs. Ils peuvent aussi éviter certaines conséquences négatives comme une blessure physique ou psychologique, une moindre capacité à accomplir une tâche, une plus forte vulnérabilité vis-à-vis de l'exploitation des autres (Derlega et Chaikin 1977), une perte d'estime de soi (Laufer et Wolfe 1977), des réactions émotionnelles telles que la colère ou le ressentiment (Chapman 1985), qui peuvent elles-mêmes avoir des répercussions majeures. Il est ainsi admis que le fait de ne pas atteindre le niveau d'intimité souhaité peut conduire un individu à des comportements anti-sociaux (Glazer 1964 ; Heffron 1972[37]). C'est d'ailleurs la raison pour laquelle de nombreux auteurs reconnaissent que l'assurance de protection de la vie privée est importante pour maintenir l'ordre social. Ainsi, pour Schwartz (1968) : « les garanties de respect de la vie privée, c'est-à-dire les règles stipulant qui peut ou ne peut pas observer ou divulguer des informations sur autrui, doivent être établies dans tout système social qui se veut stable ».

1.4 La distinction avec d'autres concepts proches et/ou liés

Le vocabulaire utilisé par les chercheurs dans le domaine de la protection de la vie privée est très hétéroclite car issu de nombreuses disciplines. A cela, on peut ajouter que beaucoup d'auteurs emploient un terme pour un autre et que certains concepts sont utilisés couramment

[36] Cette théorie sera présentée en détail au chapitre 7.
[37] Cités par Solove (2002)

dans des sens proches. De plus, de nombreuses notions sont imbriquées les unes dans les autres ou sont proches les unes des autres, ce qui accroît encore la difficulté de compréhension. Enfin, la traduction de certains mots (de l'anglais au français) ne permet pas toujours de rendre compte de la totalité du sens incombant au terme d'origine. Ainsi, le mot « *privacy* » - utilisé par tous les chercheurs anglo-saxons travaillant sur le sujet – est parfois traduit par « intimité » qui constitue une dimension particulière de la protection de la vie privée. Pour autant, ce dernier terme ne constitue pas forcément la traduction idéale. En français on peut ainsi lui préférer le terme de « confidentialité » lorsqu'on aborde la question des données personnelles.

Dans un souci de clarté, nous avons souhaité faire le point sur plusieurs concepts proches et/ou liés à celui de « protection de la vie privée ». Nous ferons ainsi la différence entre les notions d'intimité, de réserve, de secret et d'anonymat puis entre celles de protection de la vie privée et de sécurité, avant d'évoquer le concept d'invasion de la vie privée.

Westin (1970) a identifié quatre états de préservation de la vie privée : 1) la solitude (état d'être seul), l'intimité (quelques personnes qui veulent rester ensemble), la réserve (besoin de restreindre les interactions avec les autres et toute communication à propos de sa personne) et l'anonymat (fait de ne pas s'identifier). Dans une étude complémentaire, Pedersen (1979) ajoute la dimension d'isolement (séparation physique de soi vis-à-vis des autres en tant que choix de vie) - proche du concept de solitude - et distingue 2 cas d'intimité (l'intimité avec la famille et l'intimité avec les amis). Ces 6 états correspondent aux moyens physiques, informationnels et sociaux de préserver sa vie privée.
Le concept d'intimité est proche de celui de secret, lui-même peu éloigné de celui de réserve. Pour Bok (1984) les termes de secret et d'intimité se superposent puisque garder un secret est un moyen de maintenir une certaine intimité. Cependant, l'intimité ne signifie pas forcément un besoin de cacher et le secret peut porter sur des aspects non privés. En fait, intimité et secret sont tous deux des moyens de contrôler l'accès à ses informations personnelles, à ses activités, à son espace et à ses possessions.

Il y a aussi souvent confusion entre respect de la vie privée et sécurité. La question concerne plus particulièrement la collecte et l'utilisation de données personnelles par une entreprise (cas qui nous intéresse dans le cadre de cette thèse) et se pose avec plus d'acuité lorsque ces pratiques ont lieu sur Internet. Qu'en est-il précisément de la différence entre ces deux notions ? Tout d'abord, on peut dire qu'il y a respect de la vie privée si le propriétaire de

l'information - le consommateur - peut contrôler sa collecte, sa distribution et son accès. Il y a sécurité si le dépositaire de l'information – généralement l'entreprise - peut contrôler son stockage et sa diffusion. Le respect de la vie privée requiert donc un minimum de sécurité car, sans la possibilité de restreindre l'accès aux bases de données, elle ne peut pas être respectée. Pour résumer, il ne peut y avoir préservation de la vie privée sans sécurité, mais, l'assurance de sécurité ne garantit pas forcément le respect de la vie privée.

La notion de violation de la vie privée cache la complexité des questions liées à l'utilisation d'un mot si émotionnellement chargé que celui de violation. On lui préfère alors le terme d'invasion, souvent défini comme l'intrusion dans l'intimité de quelqu'un. Comme l'explique Boatright (2000) « un degré d'intimité est nécessaire pour pouvoir profiter pleinement de certaines activités, si bien qu'une invasion de la vie privée change le caractère de ces expériences et nous prive de la possibilité d'en tirer plaisir ». Le mal résultant de l'incapacité à prendre du plaisir est alors présumé supérieur à tout bénéfice (tel que le gain de confort), si bien que l'invasion est toujours vécue négativement.
La vie privée est préservée quand : 1) l'individu contrôle les interactions sociales et/ou les stimuli externes non désirés, 2) peut prendre des décisions autonomes sans interférence de l'extérieur, et/ou 3) peut contrôler la circulation de ses données personnelles (Stone et Stone 1990 ; Westin 1967). On peut donc dire qu'il y a invasion quand l'individu est incapable de contrôler ses interactions avec son environnement social et/ou physique (Laufer et al. 1976) ou quand ses actions sont structurées de manière inconnue (Simitis 1987). De plus, le fait qu'un comportement puisse être qualifié d'invasion de la vie privée dépend, notamment, de l'objet affecté. Plus l'« envahisseur » se rapproche du soi profond, plus le comportement est vécu comme une invasion (Jourard 1959).

Nous allons à présent insister sur l'aspect polymorphe du besoin de protection de la vie privée en montrant notamment en quoi il est à la fois culturellement (2.1) et individuellement (2.2) spécifique. Nous terminerons en évoquant l'aspect processuel du phénomène (2.3).

2 La protection de la vie privée : un besoin polymorphe

2.1 Un besoin différent selon les cultures et les périodes

Altman (1977) indique, qu'en tant que processus d'interaction sociale, le besoin de respect de

la vie privée est à la fois universel[38] et culturellement spécifique. Pour lui, les mécanismes par lesquels la préservation de la vie privée est rendue possible sont liés aux contextes physique, psychologique et social de l'individu, qui sont spécifiques à sa culture.

Le concept de protection de la vie privée peut donc avoir plusieurs sens selon les personnes, les cultures et/ou les périodes concernées. Il est ainsi associé à des valeurs différentes dans des pays aussi proches que les Etats-Unis et la Grande Bretagne. En effet, d'après les résultats trouvés par Petrison et Wang (1995), les Américains sont davantage préoccupés par l'aspect physique (intrusions par le biais de communications non sollicitées), tandis que les Anglais sont plus concernés par l'aspect informationnel (collecte et utilisation de données personnelles). Cette différence serait liée aux influences sociales et politiques présentes dans chaque pays, ainsi qu'à des normes différentes au sein de chaque culture.

Les formes de protection (ou d'invasion) de la vie privée évoluent aussi selon les périodes : elles ne sont pas immuables. Le fait qu'un comportement constitue (ou non) une forme d'intrusion est ainsi fonction des normes présentes à un moment donné. Au sein d'un même pays en effet, la manière d'aborder le sujet varie suite aux événements historiques et/ou aux changements au sein des forces sociales. Il y a donc des chances que les gens changent d'attitude, sur la base de ce qu'ils peuvent obtenir à ce moment là.

Finalement, chaque culture dispose de moyens propres à faire respecter le droit au RVP et/ou à s'en prévaloir (Altman 1977). Aujourd'hui encore, toutes les cultures n'offrent pas la possibilité de s'isoler physiquement. Dans la mesure où les mécanismes permettant de contrôler l'accès aux données personnelles varient selon les cultures, il semble raisonnable de dire que les individus peuvent eux-mêmes modifier leur manière de considérer le phénomène. On peut en effet penser que quelques aspects seront considérés comme plus importants par certaines personnes (et moins par d'autres), ces préférences évoluant elles-mêmes au sein des cultures et au cours du temps.

[38] Il convient toutefois de nuancer le côté universel du besoin de respect de la vie privée. Nous avons en effet indiqué dans l'introduction de ce chapitre que cette notion n'existait pas dans certaines cultures. Ceci est confirmé par Raymond Polin (2000) qui souligne que la manière dont elle est intégrée dans la culture est propre à chaque société et qu'il est difficile, de ce fait, de parler de valeurs et de droits de portée universelle.

2.2 Un besoin individuel

On estime que certaines personnes ont des besoins importants en matière de respect de leur vie privée, tandis que d'autres y sont peu sensibles (Bersheid 1977[39]). De même, Westin (1967) indique que les individus auraient des moyens préférés d'atteindre et/ou de maintenir leur intimité. Si certains préfèrent s'isoler physiquement (solitude ou isolement), d'autres choisissent de peu parler d'eux-mêmes (réserve). Ces différents constats amènent à supposer l'influence de variables individuelles sur le besoin d'intimité.

Des travaux visant à étudier cette hypothèse[40] montrent une différence significative en termes de préférence pour le respect de la vie privée selon le sexe et l'âge de la personne. Les adultes auraient ainsi une attirance pour la réserve, tandis que les jeunes préfèrent la solitude et l'intimité avec des proches. Les femmes auraient, quant à elles, une prédilection pour la réserve, la solitude, l'intimité et l'anonymat (Marshall 1974).
De son côté, Pedersen (1988) établit des corrélations entre certaines variables de personnalité et le besoin de la protection. Il déduit que l'échelle d'extraversion-intravertion est liée au score global de préférence pour le respect de la vie privée et au score de préférence pour la réserve. Les introvertis seraient moins nombreux (que les extravertis) à parler d'eux-mêmes. Les résultats d'études, visant à démontrer l'effet des variables de personnalité sur les choix individuels de protection de la vie privée, sont cependant largement contradictoires et parfois contraires à la théorie et à l'intuition. Ainsi, alors que dans une première étude, Pedersen (1982) montre que les sujets à faible estime de soi sont plus enclins à être réservés et à rechercher la solitude et l'anonymat, dans une seconde étude, ce même auteur (Pedersen 1988) montre qu'aucun trait de personnalité n'influence l'atteinte du niveau désiré de réserve (fait de ne pas dévoiler d'informations personnelles).

Finalement, la protection de la vie privée est un aspect multi-facettes du comportement humain et non une caractéristique unique et identique dans le temps et dans l'espace. Il s'agit d'un phénomène mouvant, qui amène les individus à opérer un processus d'ajustement permanent entre leur désir d'être accessibles aux autres et celui d'être seul.

[39] Cité par Newell (1995)
[40] Nous reviendrons sur l'étude de l'influence de variables individuelles dans le chapitre 4.

2.3 La protection de la vie privée vue comme un processus d'ajustement

Le besoin d'intimité est fortement dynamique. Westin (1967) explique que la protection de la vie privée implique un processus personnel d'ajustement, pour lequel l'individu doit trouver un équilibre entre son désir de protection et son désir de communication avec les autres. Cet ajustement a lieu dans un contexte de forces situationnelles variées telles que la pression des autres, les normes sociales, ainsi que les processus utilisés pour les combattre (Kimmel 1996).

La recherche d'intimité n'implique pas forcément d'établir une distance mais surtout de contrôler le niveau et le type de contacts avec les autres. Selon Altman (1975), c'est un phénomène de « contrôle des frontières » par lequel l'individu régule le contact qu'il établit avec les autres, en particulier avec qui il s'opère et de quelle manière. La protection de la vie privée résulte donc d'un processus mouvant, qui reflète le niveau de contact souhaité par un individu à un moment donné, pouvant aller du désir d'être accessible aux autres à celui d'être seul. Le niveau d'intimité idéal est une position intermédiaire sur un continuum mesurant l'interaction désirée, chaque extrémité représentant un niveau non satisfaisant. En effet, si une personne veut moins d'intimité qu'elle n'en a, il est probable qu'elle souffre de solitude. Si elle en désire plus, c'est qu'elle estime sa vie privée « envahie » par les autres.
Les conséquences du manque ou de l'absence d'intimité incluent l'apprentissage des limites de l'autonomie. Les individus expérimentent ainsi des feedbacks psychologiques et physiologiques négatifs, liés à l'absence de contrôle. Ceux-ci concernent plus particulièrement l'anxiété et le stress (Johnson 1974[41]), ce dernier se manifestant quand les individus souhaitent exercer un contrôle et ne peuvent pas le faire. Il a ainsi été montré que plus les données sont sensibles, plus le stress est élevé (Stone et Stone 1990).
La régulation vise donc à trouver un équilibre entre ces deux états caractérisés par un surplus ou une insuffisance d'intimité. Celle-ci dépend à la fois de facteurs personnels - besoin d'intimité, compétences interpersonnelles, variables de personnalité, capacité à recourir à des mécanismes de contrôle (Johnson 1974) - et situationnels, qui peuvent être sociaux (présence et caractéristiques personnelles des autres) ou physiques (barrières, distance interpersonnelle).

Après avoir défini la protection de la vie privée et avoir souligné l'aspect polymorphe de ce concept, nous allons nous pencher plus précisément sur la notion de protection des données à caractère personnel, une de ses principales dimensions, la plus concernée lors de la collecte

[41] Cité par Tolchinsky et al. (1981)

d'informations clients. Nous étudierons cet aspect, non dans sa dimension interpersonnelle, mais dans sa dimension « commerciale », à savoir dans le cadre de relations entre un consommateur et une entreprise, dans la mesure où c'est le cas qui nous intéresse ici.

3 La vie privée en tant que protection de données à caractère personnel

Nous commençons par définir le droit à la protection des données personnelles (3.1), avant d'identifier les différentes dimensions inhérentes à ce droit (3.2). Nous terminons en évoquant l'aspect historique et juridique du droit à la protection des données (3.3).

3.1 La définition du droit à la protection des données personnelles

Les avancées technologiques et les recherches menées sur le sujet ont montré que les atteintes en matière de respect de la vie privée touchent de plus en plus l'accès aux données personnelles, réduisant par là même l'autonomie à laquelle aspire tout individu eu égard aux informations le concernant. Cette situation est particulièrement prégnante dans le cadre des relations entre un individu et une entreprise, car toute transaction entre ces deux parties – y compris non financière – implique en général la fourniture d'informations, et notamment de données personnelles, de la part du client/usager.

Au-delà de l'aspect réglementaire, on trouve, au sein de la littérature, plusieurs définitions du droit à la protection des données personnelles. Turn (1985) définit ainsi ce concept comme les droits des personnes concernant la collecte, le stockage, la dissémination et l'utilisation de leurs données, sans pour autant expliquer ce que ces différents droits recouvrent.

La définition la plus courante du concept de protection des données à caractère personnel – que nous retiendrons dans le cadre de cette thèse - est celle donnée par Westin (1967) : « *the claim for individuals, groups or institutions to determine when, how and to what extent information about themselves is communicated to others* », que l'on peut traduire par : « le droit (ou la revendication) pour les individus, les groupes ou les institutions de déterminer quand, comment et dans quelle mesure des informations les concernant peuvent être communiquées à d'autres ». Il s'agit donc du droit pour tout individu – nous ne nous intéressons pas au cas des groupes et des institutions ici - de décider quelles informations le concernant peuvent être connues des autres. Cela implique de pouvoir déterminer à la fois quelles données peuvent être obtenues et quels usages les autres peuvent en faire. Au-delà de

l'utilisation des données, c'est donc aussi - et surtout - l'aspect « collecte » qui est concerné, puisque l'individu doit pouvoir décider des informations qu'il souhaite transmettre. Stone et al. (1983) reprennent cette définition en insistant sur la notion de contrôle. Pour eux, il s'agit de « la capacité d'un individu à contrôler personnellement les informations le concernant ».

Une grande part de la littérature se focalise sur cette idée de contrôle. Goodwin (1991) propose ainsi de définir la protection des données consommateurs comme la capacité de ces derniers à contrôler : 1) la présence d'autres personnes dans leur environnement au moment de transaction et 2) la dissémination des informations fournies au cours de la transaction à des personnes non présentes. Cet auteur caractérise donc la protection en termes de contrôle des informations, mais aussi de l'environnement, c'est-à-dire le contexte dans lequel la transaction a lieu. Cette deuxième catégorie de contrôle a été critiquée par plusieurs chercheurs dont Foxman et Kilcoyne (1993). Nous considérons, pour notre part, l'aspect environnemental comme particulièrement important dans le cadre d'une sollicitation de données sur Internet, dans la mesure où il s'agit d'un espace « ouvert » sur le monde. Il nous apparaît cependant lié davantage à la notion de sécurité, qu'à celle de protection des données personnelles. Pour un individu naviguant sur Internet en effet, contrôler la présence des autres dans son environnement revient à s'assurer que personne ne peut « capter » les informations qu'il est susceptible de donner à différents sites, ce qui dépend davantage de la politique de sécurisation des entreprises que de l'utilisation qu'elles peuvent faire des données. Bien qu'important, le contrôle environnemental ne nous semble donc pas constituer une partie substantielle de la protection des données consommateurs, contrairement à la dimension « dissémination des informations fournies ». Cette dernière demande toutefois à être clarifiée et complétée par les autres dimensions trouvées dans la littérature.

3.2 Les dimensions du droit à la protection des données personnelles

Il a été suggéré que la vie privée des consommateurs était respectée dès lors que ceux-ci étaient capables de restreindre l'utilisation faite par les entreprises de leurs données personnelles (Culnan 1995 ; Nowak et Phelps 1995). Les pratiques qui offrent un faible niveau de contrôle sur l'utilisation des données, comme la location ou la vente de fichiers clients, sont en effet souvent perçues comme préoccupantes par les consommateurs (Nowak et Phelps 1992). Ceci a conduit certains chercheurs à opérationnaliser la protection des données en termes de contrôle sur leur utilisation ultérieure. On distingue, à cet égard, plusieurs types d'utilisations possibles. A l'utilisation primaire – qui consiste à réaliser l'objectif pour lequel

les données ont été collectées – (par exemple : livrer un produit au domicile du consommateur dans le cas d'une commande) on oppose l'utilisation secondaire des données, c'est-à-dire dans un but autre que ce pour quoi elles ont été collectées à l'origine (par exemple : envoyer des offres commerciales à un consommateur ayant donné son adresse dans le cadre d'un jeu-concours). On oppose aussi l'utilisation dite interne (envoi de sollicitations par l'entreprise ayant collecté les données) et l'utilisation externe (transfert des données à des tiers).
A cette dimension de contrôle de l'utilisation des données, il convient d'ajouter celui opéré sur la collecte elle-même. Dans la définition du concept de protection des données à caractère personnel, Westin (1967) considère en effet la capacité à contrôler l'acquisition et l'utilisation d'informations personnelles, distinguant ainsi la décision de l'individu au moment de la collecte, de l'utilisation ultérieure des données, par celui les ayant reçues. Il s'agit donc pour le consommateur de pouvoir décider de donner (ou non) ses informations personnelles, et, le cas échéant, de choisir celles qu'il accepte de transmettre.

Par la suite, Foxman et Kilcoyne (1993) proposent d'ajouter, à la notion de contrôle, celle d'information. Pour eux, au delà du contrôle, une des questions importantes est de savoir si les personnes sont informées de la collecte et de l'utilisation de leurs données. De plus, pour que le consommateur ait un contrôle, il doit pouvoir donner son avis sur le type d'informations collectées et sur la manière de les utiliser par la suite. Or, pour donner un avis « éclairé » sur ces questions, il faut être préalablement informé. L'addition de cette dimension « information » permet donc d'accentuer les prérogatives laissées aux consommateurs et offre alors la possibilité de mieux évaluer leurs perceptions.

De l'avis même de la majorité des chercheurs ayant travaillé sur ce sujet, le concept de « protection des données personnelles » apparaît donc comme multidimensionnel, même si tous ne sont pas d'accord sur le nombre et l'intitulé des dimensions.
L'analyse précédente permet de distinguer deux dimensions majeures :

1) le droit à l'information : relatif à la collecte (a) et à l'utilisation des données (b) ;
2) le droit de contrôler : la collecte (a) et l'utilisation ultérieure des données (b).

Il s'agit, dans ce second cas, de la possibilité de donner son avis sur la dissémination des données (par exemple : en cas de transfert des données à des tiers, possibilité de choisir les tiers en question ou de définir la fréquence d'envoi des offres commerciales).

Afin de rendre cette classification plus opérationnelle, notamment managérialement parlant, il peut être utile de traduire ces dimensions en questions que les praticiens peuvent et/ou doivent se poser lors de la sollicitation de données personnelles :

- les consommateurs sont-ils conscients que des informations les concernant sont collectées ? Sont-ils informés de l'utilisation qui en sera faite ?
- Ont-ils la possibilité de s'opposer à la collecte de leurs données ?
- Ont-ils la possibilité de donner leur avis sur l'utilisation des données collectées ?

Dans l'ensemble, nombreux sont les chercheurs qui considèrent la protection des données consommateurs comme la conjonction de la connaissance et du contrôle de la collecte et de l'utilisation des données à caractère personnel (Culnan 1995 ; Foxman et Kilcoyne 1993 ; Goodwin 1991 ; Nowak et Phelps 1995). Il faut cependant souligner que quelques uns se focalisent davantage sur l'une de ces deux dimensions tandis que d'autres (comme Sheehan et Hoy 2000) proposent d'ajouter des dimensions supplémentaires (cf. tableau 2.2).

Tableau 2.2 Les principales dimensions du droit à la protection des données dans la littérature

Etudes	Nb de di-mensions	Connaissan-ce collecte	Connaissance utilisation	Contrôle collecte	Contrôle utilisation	Autres dimensions
Goodwin (1991)	2			X	X	
Culnan (1993)	3			X	X	X
Foxman et Kilcoyne (1993)	2	X			X	
Culnan (1995)	2 (a)	X	X		X	
Novak et Phelps (1995)	2 (a)	X	X		X	
Wang et al. (1998)	7 (b)	X	X	X	X	
Culnan et Armstrong (1999)	2			X	X	
Hoffman et al. (1999)	2			X	X	
Caudill et Murphy (2000)	2 (a)	X		X	X	
Milne (2000)	2 (a)	X	X	X	X	
Milne et Rohm (2000)	2	X	X			
Phelps et al. (2000)	2 (b)				X	
Sheehan et Hoy (2000)	5 (b)	X	X		X	X
TOTAL	-	8	6	7	12	2

(a) Une étude qui ne considère que 2 dimensions peut faire apparaître plus de deux catégories dans le tableau car une dimension peut couvrir plusieurs aspects.
(b) Bien qu'une étude puisse inclure plus de dimensions que celles présentées dans le tableau, il est possible qu'elles se classent toutes dans les dimensions identifiées.

Ainsi, pour opérationnaliser la préoccupation pour la protection des données personnelles, certains auteurs considèrent aussi les principes de base à respecter par les entreprises américaines en matière de traitement des données clients, et notamment l'accès non autorisé aux données ou les erreurs (Smith et al. 1996, Stewart et Segars 2002). Ces principes, connus sous le nom de *Fair Information Principles*[42] constituent le fondement de la politique de régulation américaine, là où l'Europe choisit une voie plus réglementaire.

3.3 Approche historique et juridique du droit à la protection des données

Les efforts pour protéger la vie privée du consommateur sont principalement de 2 ordres : légaux ou comportementaux (Goodwin 1991)[43]. Les aspects légaux visent à limiter les intrusions des entreprises – et à renseigner les consommateurs sur les règles à respecter en la matière - tandis que les aspects comportementaux suggèrent d'inclure des mécanismes de préférence permettant aux individus de choisir leur propre niveau de protection. Cette seconde approche offre l'avantage de reconnaître la diversité des préférences individuelles mais est plus difficile à mettre en place[44].

C'est donc essentiellement sur l'aspect légal que se concentre la protection du consommateur.

Avant d'aborder l'aspect juridique proprement dit, nous revenons brièvement sur l'évolution historique du droit à la protection des données.

3.3.1 Approche historique du droit à la protection des données

En tant que droit, le respect de la vie privée et donc des données à caractère personnel a ses racines dans l'histoire. On trouve ainsi une trace de protection de la vie privée (avec une focalisation sur le droit à la solitude) chez les Hébreux et dans la Chine ancienne. D'autres nations reconnaissent ce droit depuis des siècles. C'est notamment le cas en France à travers le droit de la personnalité. Dans de nombreux pays cependant, ce droit n'est réellement protégé que depuis une vingtaine ou une trentaine d'années, tandis que dans d'autres, on ne lui accorde encore que peu de place – voire aucune - au sein de la Constitution.

---

[42] Présentés plus en détail au paragraphe suivant

[43] Nous ajouterions pour notre part un troisième volet de solutions, lié à l'aspect « technique » du problème. Il s'agit pour le consommateur de recourir à des outils permettant de limiter les intrusions et/ou de garantir la protection de ses données : logiciels de cryptage ou anti-spam, systèmes de filtrage ou d'anonymat, etc…

[44] Les projets de P3P sont, notamment, encore balbutiants.

Les préoccupations pour le respect de la vie privée ne sont pas nouvelles. Elles émergent et/ou s'accentuent quand le public perçoit une menace due à l'existence de nouvelles technologies, qui accroissent les capacités de surveillance, stockage, mémorisation et communication des informations personnelles (Westin 1967, Miller 1971, Mason 1986). Dans la majorité des pays, le législateur tente alors de pallier les dérives les plus inquiétantes, par de nouvelles lois et/ou l'aménagement de lois plus anciennes. On distingue ainsi plusieurs périodes, au cours desquelles la régulation du phénomène de collecte et d'utilisation des données personnelles des individus a évolué, que nous présentons ci-après.

Avant la fin des années 60, l'intérêt porté aux questions de respect de la vie privée était quasi inexistant dans les sciences sociales (Margulis 1977) et dans la réglementation. Dans l'ensemble, les citoyens faisaient confiance au gouvernement et aux entreprises. On note donc peu de limites visant les pratiques de collecte et d'utilisation de données personnelles.

La perception des menaces posées par les systèmes informatiques a suscité l'attention du public à la fin des années 60 et/ou au début des années 70. Deux livres datant de cette période constituent d'ailleurs la base des recherches sur le sujet : « *Privacy and Freedom* » (Westin 1967) et « *The assault of privacy* » (Miller 1971). A cette époque, l'arrivée de nouvelles technologies – et l'accroissement de la surveillance et des capacités de stockage qui en découle - (ré)introduit la question du droit au RVP[45] dans de nombreux pays. De nouvelles lois sur la collecte et la dissémination des données deviennent alors nécessaires. L'une des premières lois sur la protection des données a ainsi été élaborée en Allemagne en 1970. Aux Etats-Unis, on assiste également au début d'une campagne visant à demander une législation plus protectrice dans ce domaine. De plus, dès cette période, et face à la montée des préoccupations des citoyens, plusieurs auteurs soulignent le besoin de recherches sur le sujet. Des études empiriques visant à connaître l'impact des nouvelles technologies sur la vie des individus sont alors menées. Les travaux américains montrent ainsi à quel point et dans quelle mesure les individus accordent de la valeur au respect de la vie privée, réagissent à une invasion de leur vie privée et attendent que la loi les protège davantage.

Dans les années 80, le sujet est de nouveau à l'ordre du jour, à cause du développement du marketing des bases de données et du marketing direct. On observe alors une augmentation du tissu législatif visant à protéger les individus des différents abus. Beaucoup de pays, parmi ceux qui ne l'ont pas encore fait, choisissent alors une voie réglementaire. C'est, notamment, la période de création de lois protectrices dans les pays européens (la Loi Française sur l'informatique et les libertés date de 1978).

[45] Abréviation utilisée pour le terme de « Respect de la Vie Privée ».

Dans les années 90, ce sujet devient une question politique et sociale de première ampleur aux Etats-Unis et ailleurs dans le monde. L'impact des nouvelles technologies devient en effet très important (Internet, communication sans fil, solutions de cryptage, …). Parallèlement, on observe une globalisation du problème, les données collectées étant amenées à circuler dans le monde entier. Ces deux phénomènes accroissent les préoccupations des citoyens à ce sujet, que l'approche réglementaire cherche à juguler.

3.3.2 Approche juridique du droit à la protection des données

Le respect de la vie privée est un droit fondamental reconnu comme tel par la Déclaration Universelle des Droits de l'Homme, la Convention Européenne pour la protection des individus au regard de la collecte automatique des données personnelles, et d'autres traités régionaux et internationaux. Presque chaque pays dans le monde reconnaît ce droit dans sa constitution et/ou ses lois.

Le traitement de la protection de la Vie Privée repose en réalité sur 2 types de sources : un aspect législatif et réglementaire et un aspect lié à la dynamique des marchés (auto régulation, contrats, surveillance privée et systèmes de garanties). Le choix entre les 2 approches est largement culturel. Ainsi, l'approche américaine est fondée sur l'intervention minimale de l'état, tandis qu'en Europe et en France, la présence du législateur est prégnante.

L'approche américaine

Dans les pays où la protection face aux abus en matière de collecte et d'utilisation de données personnelles n'est pas explicitement reconnue dans la constitution, la jurisprudence ou des lois ont généralement permis d'identifier un « droit au respect de la vie privée ». Ainsi, si les Etats-Unis n'ont toujours pas de loi fédérale dans ce domaine (Jarvis 2001), on compte de nombreuses lois au niveau des états. Il existe également un certain nombre d'amendements qui tombent sous la rubrique générale de « protection de la vie privée ». Enfin, loin d'être d'une portée générale, la réglementation américaine s'attache plutôt à protéger les données les plus sensibles (notamment médicales), les personnes les plus fragiles (enfants), ou les secteurs d'activités les plus exposés (comme celui du crédit, de la banque et de l'assurance).
L'absence de loi fédérale aux Etats-Unis s'explique principalement par le choix d'une approche fondée sur l'autorégulation. Celle-ci fut fortement encouragée dans les années 90 par le FTC, organisme chargé par le gouvernement de l'époque de réfléchir à la protection de

la vie privée des citoyens américains. Selon cette approche, les entreprises américaines sont incitées à adopter elles-mêmes des pratiques loyales. Cependant, afin d'encadrer ce régime et éviter ainsi les écarts de conduite, le FTC définit plusieurs principes à respecter en matière de collecte et de traitement d'informations personnelles. Appelés « *Fair Information Principles* », ils visent surtout à trouver un équilibre entre les intérêts privés des individus et les besoins légitimes des entreprises en informations. Ces principes sont au nombre de 5 :

- Information/Notification (*Notice*) : informer le consommateur des pratiques utilisées par l'entreprise en matière de collecte et de traitement des données personnelles ;
- Choix (*Choice*) : laisser au consommateur le choix de l'utilisation de ses données (en particulier si elle est différente de l'objectif d'origine) ;
- Accès (*Access*) : donner la possibilité au consommateur d'accéder aux données le concernant et d'en contester l'exactitude ;
- Sécurité et Intégrité (*Security & Integrity*) : obligation faite à l'entreprise de prendre des mesures « raisonnables » visant à assurer la sécurisation des données durant la transaction et le stockage ultérieur ;
- Respect des règles et Réparation (*Enforcement & Redress*) : mécanismes de surveillance visant à s'assurer que les entreprises respectent ces principes et prévoyant des sanctions en cas de manquement aux règles.

L'objectif majeur de ces principes est de signaler aux consommateurs américains que les entreprises suivront un ensemble de règles considérées comme justes et qu'elles ne se comporteront pas de façon opportuniste (Shapiro 1987). En minimisant les risques, ils aident donc à bâtir la confiance des individus et les encouragent à fournir leurs données personnelles (Culnan et Armstrong 1999 ; Milne et Boza 1999).

Les Etats-Unis offrent ainsi un exemple d'approche décentralisée et « fragmentée ». En effet, que ce soit au niveau des états ou au niveau fédéral, les mesures statutaires ou réglementaires se focalisent sur les pratiques des industries, dans certains secteurs spécifiques. Celles-ci ont créé des droits légaux défendables devant les cours, à défaut d'être assurés par une administration centrale, comme c'est le cas en Europe. De plus, la protection ne se limite pas aux mesures réglementaires puisque entreprises et organismes professionnels ont, de leur côté, mis en place des codes de conduite pour rassurer les consommateurs quant à la manipulation de leurs données. Il existe donc une grande variété de mesures et de pratiques qui créent toute une mosaïque de droits très particuliers et très spécifiques.

L'approche européenne

En Europe, la protection des données personnelles est principalement assurée par la Directive du 24 Octobre 1995 relative à « la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données ». Son adoption répondait alors à 2 objectifs majeurs : assurer la protection d'un droit fondamental et prévenir les restrictions d'échanges d'informations entre pays européens. Cette Directive repose sur des principes souvent différents de ceux préconisés par l'approche américaine. Elle s'applique ainsi à tous les secteurs de l'économie. Elle prévoit aussi l'existence, au sein de chaque pays, d'une autorité indépendante, disposant d'un pouvoir de surveillance et d'investigation[46]. Enfin, elle prévoit des sanctions en cas d'abus. Dans l'ensemble, elle vise donc à protéger les droits individuels de manière préventive, à travers une autorité de contrôle ayant un rôle actif.

Malgré ces différences, l'approche européenne est basée sur des principes fondateurs, qui ne sont pas sans rappeler ceux préconisés outre-Atlantique[47] (cf. tableau 2.3). En outre, il faut souligner que la nouvelle Directive Européenne, adoptée en 2002, apporte de nouvelles conditions à ces principes d'origine. Il y est notamment fait mention du consentement préalable du consommateur à recevoir des communications commerciales (principe d'opt-in), sauf en cas de relation déjà établie avec l'entreprise, cette dernière pouvant alors continuer à recourir au principe d'opt-out (consentement présumé sauf en cas d'opposition expresse)[48].

Tableau 2.3 Les principes de base de la Directive Européenne de 1995, comparés aux FIP

Principes de la Directive Européenne de 1995	FIP américains
Toute donnée doit être obtenue de manière éthique et légale	
Toute donnée doit être collectée dans un but légitime et explicite et utilisée de manière concordante	Information (Notice) et Choix (Choice)
Toute collecte doit être pertinente et non excessive par rapport à l'objectif d'origine	
Toute donnée est juste (non erronée), à jour, et doit être détruite une fois le but atteint	
Il existe un droit d'accès et de vérification des données pour le consommateur	Accès (Access)
Celui qui collecte et conserve ces données est responsable de leur sécurité	Sécurité (Security)

[46] Par exemple, la CNIL (Commission Nationale de l'Informatique et des Libertés) en France.
[47] Le rapprochement se fait plutôt en sens inverse : ce sont les FIP qui sont inspirés de la Directive Européenne.
[48] Nous reviendrons ultérieurement sur ces notions de consentement explicite (opt-in) et implicite (opt-out).

Dans l'ensemble, l'approche européenne est donc beaucoup plus restrictive que ne le sont les mesures volontaires prises par les entreprises américaines. En effet, si aux Etats-Unis, la loi protège surtout les citoyens des abus du gouvernement - pour ce qui est des industries, les protections sont largement sectorielles - les lois européennes donnent aux individus des droits vis-à-vis de l'ensemble des pratiques du secteur privé.
En tant que membre de l'Union Européenne, la France applique cette Directive Européenne, tout en ayant ses propres dispositifs réglementaires.

L'approche française

Notre pays possède dans ce domaine tout un arsenal juridique, avec des dispositions, parfois fort anciennes, qui protègent certains compartiments de la vie privée, comme l'aspect territorial (le domicile) ou relationnel (la correspondance).
Au-delà d'une protection de nature constitutionnelle, nous disposons d'une loi de droit commun qui assure une préservation générale de la vie privée. Ainsi, l'article 9 (premier alinéa) du Code Civil (1970), stipule que « chacun a droit au respect de sa vie privée ». On trouve aussi des textes visant des types particuliers d'atteinte à la vie privée. Il existe ainsi une loi qui concerne l'informatique, les fichiers et l'accès aux documents qui contiennent des informations nominatives - la loi relative à l'informatique, aux fichiers et aux libertés (1978)[49] - et, plus récemment, la Loi pour la Confiance dans l'Economie Numérique (2004) qui constitue un nouveau cadre juridique pour les relations commerciales sur Internet. Cette dernière dispose notamment d'un volet concernant l'envoi de propositions commerciales par email qui reprend les principes de la nouvelle Directive Européenne de 2002.

A l'analyse des dispositions en vigueur au sein de ces deux continents, il apparaît que la notion de vie privée joue un rôle différent dans chacune de ces sociétés. Si en Europe, elle est considérée comme un droit humain (dont on trouve l'origine dans la Convention Européenne des Droits de l'Homme de 1950), aux Etats-Unis, elle est vue avant tout comme le résultat d'une négociation contractuelle entre les parties. Les raisons de ces différences sont à la fois historiques et culturelles. Comme certaines recherches l'ont montré (Milberg et al. 1995), les facteurs culturels semblent d'ailleurs avoir une double influence dans ce domaine, jouant un rôle significatif à la fois sur le niveau de préoccupation de l'opinion publique et sur l'approche réglementaire adoptée.

[49] Modifiée par la loi relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel du 6 août 2004.

Conclusion du chapitre 2

L'aspect psychologique du concept de protection (ou de respect) de la vie privée, de même que les études portant sur son sens commun, insistent sur le contrôle, la régulation et la limitation de la surveillance des autres et de l'accès non désiré à soi. Un des aspects les plus importants de la protection concerne ainsi le contrôle de l'accès aux données personnelles, vu comme « le droit de déterminer de quelle manière des informations nous concernant peuvent être communiquées aux autres » (Westin 1967). Dans la suite de ce document, nous emploierons d'ailleurs souvent le terme de respect de la vie privée (RVP) pour désigner la protection des données à caractère personnel.

On reconnaît de multiples bénéfices à la recherche de protection de la vie privée. D'un point de vue sociopolitique, l'intimité donne la possibilité de s'exprimer, de choisir et de s'associer. Au niveau psychologique, elle est le fondement de l'interaction sociale, qui affecte le développement de soi et de son individualité. De manière générale, elle est donc à l'origine d'un fonctionnement psychologique normal et du développement de relations interpersonnelles stables.

La régulation de la vie privée implique un système de *feedback* complexe, à travers lequel des ressources sont mobilisées pour trouver un compromis entre les désirs - parfois antinomiques - de protection et de communication. Il existe alors un ajustement continuel entre les deux, et ce, dès qu'une nouvelle situation émerge. Ces ajustements peuvent être étudiés en termes de coûts occasionnés par la régulation et/ou par la non atteinte de l'équilibre voulu. Ces coûts peuvent être de natures différentes : effort physique, temps passé, dépenses psychologiques telles que le stress et l'anxiété.

Selon Rachels (1975), il existe un lien étroit entre notre capacité à contrôler l'accès à nos informations et la possibilité de créer et de maintenir des relations avec les autres. Le Respect de la Vie Privée est donc important puisqu'il permet de choisir avec qui nous souhaitons être en contact. De manière générale, et dans la mesure où nous sommes libres d'établir des liens avec les autres, si nous souhaitons garder une certaine réserve, nous devons alors limiter ces relations. La recherche d'intimité est donc intimement liée au phénomène de dévoilement de soi, notion que nous allons présenter dans le chapitre suivant.

Partie I - COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

CHAPITRE 3
DÉVOILEMENT DE SOI
ET DIVULGATION DE DONNÉES PERSONNELLES

« People tend to become attracted to others
as a result of having disclosed personal information to them ».
Collins and Miller (1994)

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

Introduction du chapitre 3

Le dévoilement de soi (*self disclosure* en anglais) correspond au fait de révéler des informations sur sa personne à d'autres individus. Il est donc un moyen - parmi d'autres - de contrôler l'accès à ses données personnelles, et plus généralement de protéger sa vie privée. En ce sens, il est aussi concerné lorsqu'un consommateur est sollicité pour remplir un formulaire, comprenant des données à caractère personnel. Dans une telle situation en effet, l'entreprise, par le biais d'une méthode de collecte « directe », convie celui-ci à fournir des informations, ce qui revient à lui demander de s'engager dans une forme de dévoilement.

Jusqu'à présent l'étude du concept de dévoilement de soi a majoritairement, si ce n'est exclusivement, concerné l'aspect interpersonnel c'est-à-dire les relations entre individus. L'essentiel des résultats issus de cette littérature sont présentés dans ce chapitre. Nous verrons ainsi, après avoir évoqué ses principales caractéristiques (1), comment s'opère le processus de dévoilement (2) et quels sont ses facteurs d'influence (3). Nous terminerons par l'étude du lien entre dévoilement de soi et relation à l'autre (4), qui s'avère particulièrement utile pour comprendre la manière dont une personne prend la décision de fournir des données, parfois de nature intime, à une autre, en fonction du lien qui les unit. Les principes d'attraction et de réciprocité, particulièrement étudiés dans la littérature, seront ainsi développés.

Si les travaux présentés concernent essentiellement les relations interpersonnelles, il n'en demeure pas moins vrai que la majorité des résultats peuvent se transposer à l'étude des relations entre un individu et une entreprise, y compris dans le cadre d'un échange sur Internet, même si quelques principes de base – comme celui de la réciprocité - y sont plus difficilement applicables. En atteste le recours fréquent à cette littérature par les chercheurs en marketing, qui étudient la réaction des consommateurs face à une sollicitation de leurs données personnelles, dans un but commercial. Les conclusions que l'on peut tirer de l'analyse du concept de dévoilement de soi semblent donc prometteuses.

1 Définition et attributs du dévoilement de soi

Nous donnons tout d'abord la définition du concept de dévoilement de soi (1.1) avant d'en présenter les différents attributs (1.2). Nous abordons ensuite les conséquences d'un tel phénomène (1.3)

1.1 Définition du dévoilement de soi

Le dévoilement de soi peut se définir comme le fait de se faire connaître des autres, de leur donner des informations sur sa personne (Jourard et Lasakow 1958). A l'origine, le phénomène concerne toute donnée personnelle communiquée verbalement aux autres. Par extension, on considère le processus par lequel on donne des informations de nature personnelle aux autres, quelle que soit la méthode employée (oral ou écrit), même si l'étude du dévoilement oral est encore majoritaire. Ce que l'individu est prêt à donner, sous quelle forme et à qui constituent alors les différentes dimensions de la distance sociale qu'il entretient avec les autres (Fitzgerald 1963).

1.2 Les attributs du dévoilement de soi

Les trois attributs de base du dévoilement de soi concernent : 1) le montant d'informations dévoilées (leur nombre) ; 2) le degré d'intimité de celles-ci (la profondeur ou sensibilité des données révélées) et 3) le temps passé à parler de soi. Un individu qui souhaite se dévoiler davantage à autrui peut donc le faire de plusieurs manières : en augmentant le nombre d'informations dévoilées, en augmentant le degré d'intimité (de sensibilité) des faits révélés ou en augmentant le temps passé à parler de lui. Si l'intimité et le temps semblent indépendants entre eux – une corrélation de .42 entre les deux a été trouvée par Cosby (1973) - il semble exister une relation négative entre le montant et le niveau d'intimité du dévoilement, telle que les individus se dévoileraient moins (en nombre) sur des sujets intimes (Jourard et Lasakow 1958 ; Taylor 1968 ; Taylor, Altman et Sorrentino 1969). Ce résultat semble stable, quelles que soient les cultures.

Les travaux sur le sujet, essentiellement de nature expérimentale, menés par des chercheurs en psychologie, visent à placer deux individus en face à face et à mesurer leur niveau de

dévoilement réciproque. Si le montant du dévoilement est évalué en nombre de sujets différents abordés au cours de la conversation et le temps, en nombre de minutes passées à parler de soi, la mesure du degré d'intimité pose davantage de problèmes, étant donné son caractère éminemment subjectif. Les chercheurs proposent généralement d'établir un codage *a posteriori* du niveau de sensibilité, sur la base de la retranscription écrite de la conversation, qui sera réalisé par au moins deux codeurs différents. Le calcul du pourcentage d'accord inter codeurs permet alors de garantir la validité des résultats obtenus.
D'autres travaux, basés plutôt sur des questionnaires, sont également proposés. Ceux-ci visent davantage à étudier le comportement habituel et/ou passé en matière de dévoilement. La fiabilité des résultats de telles études est toutefois sujette à controverse puisque les mesures de report du dévoilement passé sont probablement moins « justes » et/ou prédictives que des mesures réelles de dévoilement dans une situation donnée (Pedersen et Breglio 1968). De plus, le temps passé à se dévoiler est impossible à évaluer dans ce type de recherche, ainsi que dans celles basées sur l'analyse du dévoilement de soi par écrit. Ces études obligent alors à réduire le dévoilement à la quantité (*breadth*) et à la sensibilité et/ou profondeur (*depth*) des informations qu'un individu fournit à un autre (Jourard 1971), en omettant la dimension « temps passé », ce que de plus en plus de chercheurs choisissent de faire.

Les dimensions de quantité et de sensibilité des informations livrées sont donc majoritairement étudiées au sein de la littérature sur le dévoilement de soi dans le domaine interpersonnel. Ces deux aspects offrent en outre l'avantage d'être complémentaires et/ou interchangeables. En effet, une des méthodes employées pour prévenir la perte de contrôle sur les données personnelles, notamment en cas de risque de révélation de celles-ci à un tiers, est de remplacer la profondeur de la divulgation (*depth*) par son ampleur (*breadth*) (Jones et Archer 1976). Cette technique est souvent utilisée par les individus souhaitant protéger leur vie privée mais qui sont – ou se sentent - obligés de répondre au dévoilement préalable des autres : dans ce cas ils ne révèlent rien de ce qu'ils considèrent comme sensible. L'attente de réciprocité de la part de l'autre est alors en partie satisfaite, tout en évitant une révélation non désirée. Ce principe de réciprocité, fondement majeur de la littérature sur ce thème, sera présenté plus en détail dans la section 4.

1.3 Les conséquences du dévoilement de soi

Jourard (1959) explique que le montant et le niveau de sensibilité des informations qu'un individu est prêt à donner aux autres sont un indice du degré de proximité et d'ouverture de la

relation qu'il entretient avec eux. L'action de dévoiler des informations sur soi est donc centrale dans le développement des relations avec les autres (Altman et Taylor 1973).

La capacité de se dévoiler aux autres serait également une condition nécessaire au bien être psychologique (Jourard 1959). Plusieurs recherches ont montré le lien entre divulgation de soi et santé mentale. Pour Jourard (1959) et Cosby (1973), la relation entre les deux serait curvilinéaire : les gens se dévoilant trop ou pas assez auraient des problèmes mentaux. Le fait de peu se dévoiler serait en effet le signe d'une incapacité à devenir adulte et à établir des relations avec les autres, tandis qu'une personne qui se livrerait trop serait exagérément centrée sur elle-même. Dès lors, l'individu « normal » serait quelqu'un qui se dévoilerait beaucoup à des personnes proches mais peu aux gens moins proches (suffisamment pour établir une relation mais pas trop pour ne pas être envahissant).
Au-delà de la quantité d'informations dévoilées, il est aussi possible d'évaluer le moment ainsi que la personne auprès de laquelle on a choisi de le faire, pour juger du bien-être d'un individu. Ainsi, le fait de se dévoiler à un « mauvais » moment ou à une personne inappropriée pourrait refléter des problèmes de socialisation ou de « déréglage »[50], tout comme le fait de ne pas le faire alors que cela serait approprié (Chaikin et Derlega 1974).

Après avoir présenté la définition et les attributs du dévoilement de soi, voyons maintenant de quelle manière s'opère le processus. Nous verrons tout d'abord les liens entre dévoilement de soi et échange (2.1) avant d'expliquer en quoi ce processus est conduit par la motivation (2.2).

2 Le processus de dévoilement de soi

2.1 Dévoilement et échange

Le dévoilement de soi, dans le domaine interpersonnel, peut s'ancrer dans les théories de l'échange (Bagozzi 1975), de l'échange social (Homans 1961, Blau 1964) et de la pénétration sociale[51] (Altman et Taylor 1973). Selon cette dernière, la croissance d'une relation interpersonnelle – vue comme un échange réciproque de données personnelles - serait fonction de l'analyse des coûts et bénéfices de la relation, de caractéristiques personnelles et de déterminants situationnels (Cosby 1973). Ce principe d'une analyse coûts/bénéfices est aussi étudié dans la théorie de l'échange social, dont elle constitue le fondement majeur. Les chercheurs qui étudient le dévoilement de soi, dans le domaine interpersonnel, adoptent

[50] En anglais, *maladjustment*
[51] Nous reviendrons plus en détail sur les fondements et principes de cette théorie dans la section 4 de ce chapitre.

d'ailleurs majoritairement une perspective basée sur cette théorie de l'échange social (Thibaut et Kelley 1959, Ajzen 1977), suggérant que le dévoilement, comme d'autres comportements interpersonnels, sont interprétés en termes de coûts et bénéfices par les individus.

Les travaux menés dans le domaine interpersonnel accréditent cette idée d'équilibre : les gens se confient pour les bénéfices de la relation, ceux-ci étant comparés aux pertes potentielles (Derlega et al. 1993). Thibaut et Kelley (1959) confirment également l'idée selon laquelle les gens acceptent de se dévoiler si les bénéfices perçus sont supérieurs aux risques perçus. En évaluant les coûts et bénéfices du dévoilement, les individus réalisent ce que Laufer et Wolfe (1977) appellent un « *calculus of behavior* », calcul basé sur l'estimation des conséquences futures d'un tel comportement, ou plutôt sur leur capacité à gérer, dans le futur, les conséquences de leurs choix actuels (Laufer et Wolfe, 1977). Ce calcul dépendrait de leur *self ego*, de l'environnement dans lequel ils se trouvent et des relations qu'ils entretiennent avec les autres (Laufer et Wolfe 1977 ; Stone et Stone 1990). Plus les conséquences futures sont considérées comme bénéfiques, plus l'individu sera donc enclin à se dévoiler.
Cette dimension d'échange social pourrait aussi s'appliquer au contexte commercial, dans le cadre d'une sollicitation de données personnelles de la part d'une entreprise. Dans de telles circonstances, le consommateur ferait également une analyse coûts/bénéfices de la situation, qui se rapprocherait du « *calculus of behavior* » de Laufer et Wolfe (1977).

2.2 Un processus conduit par la motivation

Dans tout comportement de divulgation, il existe deux forces opposées, l'une poussant à se dévoiler, l'autre jouant un rôle inhibiteur. Pour Stanton et Stam (2003)[52], le fait pour un individu de contrôler le flux d'informations concernant sa personne est un processus conduit par la motivation. Celle-ci s'exprime dans un contexte de relations sociales qui guident « qui peut apprendre quoi, sur qui, quand et à quel prix ? ».

[52] Il est intéressant de souligner l'écart "générationnel" existant entre les premières recherches sur le sujet (datant des années 60-70, comme en attestent de nombreuses références présentées dans ce chapitre) et les recherches plus récentes (fin des années 90 et début 2000), comme celle de Stanton et Stam (2003) citée ci-dessus. Nous trouvons ici la manifestation du fait que la recherche s'inspire fortement de l'histoire et de l'actualité. En particulier, dans le domaine de la vie privée, elle suit l'évolution technologique qui conduit à de nouvelles menaces pour les citoyens, ce que nous avons déjà souligné au chapitre 2. Ceci explique, au moins en partie, le fait qu'il existe peu de références entre ces deux périodes, époque au cours de laquelle les préoccupations liées à ces questions étaient peu élevées et les besoins en travaux académiques moins importants. Nous soulignons, enfin, la différence de nature existant entre ces différents travaux, les plus anciens se focalisant sur l'étude des facteurs individuels tandis que les plus récents s'intéressent davantage à l'impact des nouvelles technologies et notamment de l'ordinateur et de l'Internet.

Stanton et Stam (2003) proposent un cadre d'analyse visant à déterminer les éléments de motivation qui conduisent les individus à dissimuler ou révéler des informations les concernant. Ce cadre s'inspire de différentes théories, dont la théorie du management de l'intimité (Petronio 1991) et la théorie des frontières de l'information[53]. Ainsi, selon Petronio (1991), lorsque les individus sont confrontés à une telle décision, ils se fient à un système de frontières basé sur des règles précises, pour décider du niveau d'accessibilité de leurs données personnelles. Dès lors, si le fait de révéler des informations les rend potentiellement vulnérable et s'ils estiment que les données leur appartiennent, ils préfèreront exercer un contrôle sur elles, en refusant de les dévoiler. La théorie des frontières de l'information s'avère également particulièrement adaptée à ce cadre d'analyse dans la mesure où elle permet de prévoir les motivations de l'individu concernant le nombre et la nature des informations qu'il souhaite révéler, à travers un médium particulier. Elle vise aussi à prédire les préférences et le comportement des individus sur la base de leurs croyances quant à la nature de la relation qu'ils entretiennent avec l'audience, l'utilisation attendue des informations divulguées et les bénéfices retirés du dévoilement. Les individus établissent alors un calcul des conditions pour lesquelles la divulgation est acceptable ou non.

Les recherches antérieures distinguent différentes sources de motivations, parmi lesquelles on trouve notamment l'éthique, la valeur instrumentale et la confiance. De son côté, Higgins (1997, 1998) pense que les gens adoptent un comportement pour autoréguler leur état émotionnel. Il décrit deux types d'autorégulations : le « *promotion focus* », qui se concentre sur l'atteinte d'idéaux et l'obtention de gains ; et le « *prevention focus* », qui vise à éviter les pertes en répondant à ses devoirs, obligations et responsabilités. Pour Stanton et Stam (2003), les individus choisiraient donc de donner ou de cacher des informations afin d'atteindre un état désiré, celui-ci pouvant être - selon la classification de Higgins - une focalisation sur l'obtention de gains ou la limitation de pertes. Ils obtiennent ainsi un cadre d'analyse constitué de 4 cadrans : révélation ou rétention d'informations pour obtenir des gains ou limiter les pertes. Leur objectif est alors de trouver les motivations qui interviennent dans chacun de ces cadrans. Ce faisant, ils s'intéressent plus particulièrement à la motivation « confiance ». Ils font l'hypothèse que la confiance est nécessaire pour la catégorie « *redemptive* » (révélation d'informations afin de limiter les pertes) alors que pour la catégorie « *exoteric* » (informations données pour obtenir des gains), il y a révélation quel que soit le niveau de confiance au sein de la relation. Le test de ce cadre d'analyse, à travers une étude qualitative, confirme ces hypothèses. Stanton et Stam (2003) montrent notamment que

[53] En anglais, « *Theory of Communication Privacy Management* » et « *Information Boundary Theory* »

la confiance peut faciliter la révélation volontaire d'informations, et qu'elle s'avère nécessaire quand on se dévoile pour limiter les pertes. A l'opposé, en cas de manque de confiance, les situations de révélations volontaires d'informations sont peu probables, du fait de l'existence de formes subtiles de résistance. Des motivations politiques ou protectrices peuvent alors conduire à cacher les informations.

Nous abordons, dans une troisième partie, les facteurs susceptibles d'avoir un impact sur le dévoilement de soi. Nous verrons, tour à tour, l'influence des facteurs culturels (3.1), individuels (3.2) et situationnels (3.3).

3 Les facteurs d'influence du dévoilement de soi

On trouve au sein de la littérature trois grands types de facteurs pouvant influencer le dévoilement de soi : les facteurs culturels et/ou environnementaux, les facteurs individuels (principalement de nature sociodémographique et psychologique) et les facteurs situationnels. Nous évoquerons chacune de ces catégories, en insistant davantage sur la seconde (variables individuelles) puisqu'elle est la plus développée au sein de la littérature. Nous présenterons surtout les résultats clairement établis, l'effet de certains facteurs ayant été soumis à polémique. C'est par exemple le cas des variables de personnalité, dont l'effet démontré dans quelques études est parfois infirmé dans les travaux ultérieurs.

3.1 L'influence des facteurs culturels

Nous n'insisterons pas sur l'aspect culturellement spécifique du dévoilement de soi. D'une part, ce résultat a déjà été indirectement évoqué à travers l'étude de la protection des données à caractère personnel (cf. chapitre 2). D'autre part, il a été largement démontré dans la littérature sur le dévoilement de soi, tous les chercheurs se montrant unanimes sur ce sujet. Ainsi, par exemple, Lewin (1948)[54] trouve qu'Allemands et Américains diffèrent quant à leur capacité à confier des informations. Ce résultat est confirmé par Plog (1965) qui montre que les Américains se dévoilent plus que les Allemands.

Certains chercheurs ont choisi d'étudier l'impact de l'origine ethnique plutôt que du pays dans lequel on vit. Ainsi, dans une étude de la société américaine, Jourard et Lasakow (1958) montrent que les personnes de peau blanche se dévoilent plus que celles à la peau noire.

[54] Cité par Jourard et Lasakow (1958)

3.2 L'influence des facteurs individuels

Nous étudierons dans un premier temps l'effet du sexe, dans la mesure où ce résultat est clairement établi au sein de la littérature, avant d'aborder l'influence des variables de personnalité, pour laquelle les résultats sont parfois contradictoires.

3.2.1 Le dévoilement de soi selon le sexe

S'il semble évident qu'il existe des différences individuelles en termes de dévoilement de soi, il reste à savoir dans quel sens vont ces différences. L'étude de la littérature permet de répondre à cette question concernant le sexe de la personne qui se dévoile.

Dès 1958, Jourard et Lasakow montrent que les femmes parlent davantage d'elles-mêmes que les hommes, ce qui a été confirmé ensuite par de nombreuses études (Cosby 1973 ; Jourard 1971), rendant ainsi ce résultat relativement solide. D'autres travaux, cités par Cosby (1973) trouvent cependant une absence de différence selon le sexe[55]. Jourard (1964) souligne que ces résultats contradictoires pourraient être liés à l'utilisation d'échantillons d'origine géographique différente et ayant des normes différentes quant aux rôles des personnes de chaque sexe, cette explication n'ayant toutefois pas été démontrée. Pour Pedersen et Breglio (1968), ce n'est pas tant le montant du dévoilement de soi qui différencie les hommes et les femmes, que la profondeur des informations dévoilées. Ainsi, dans une étude portant sur le dévoilement écrit, ils constatent que les femmes n'utilisent pas plus de mots pour se décrire mais qu'elles donnent des informations plus intimes que les hommes.

Au-delà de cette tendance à se confier davantage, il apparaît également que les femmes valorisent davantage le dévoilement de soi que les hommes. Cette différence peut être trouvée dès l'enfance, dans la façon d'être éduquée selon que l'on soit un garçon ou une fille. Les comportements appropriés au sexe de l'enfant sont ainsi encouragés alors que les comportements inappropriés sont, soit découragés, soit ignorés. Jourard (1971) explique à cet égard que les garçons se croient obligés de cacher une grande partie d'eux-mêmes, dès le plus jeune âge. En plus des expériences de l'enfance, les normes culturelles prennent le relais à l'âge adulte pour maintenir cette différence entre les deux sexes. Ainsi, les hommes qui s'identifient au rôle de mâle semblent craindre d'être rejetés ou d'être ridicules s'ils ne se conforment pas au comportement que l'on attend d'eux.

[55] On peut citer, entre autres, ceux de Plog (1965), Didmond et Hellkamp (1969), Doster et Strickland (1969).

3.2.2 Dévoilement de soi et personnalité

Les études visant à établir un lien entre le niveau de dévoilement et des variables de personnalité donnent des résultats contrastés. Il est en effet difficile de trouver, au sein de la littérature, des conclusions confirmées par des travaux ultérieurs. Ainsi, dans la mesure où la relation entre sociabilité et extraversion est établie, on supposait l'existence d'une relation négative entre le dévoilement et l'extraversion. Ce résultat était d'ailleurs confirmé par Mullaney (1964)[56] qui, cherchant à montrer le lien entre différentes variables de personnalité et le dévoilement de soi, trouve que seule la dimension « introversion » lui est liée : moins une personne se dévoile, plus elle est introvertie. Toutefois les études de Pedersen et Breglio (1968) et Pedersen et Higbee (1969) sont contradictoires et ne permettent pas d'affirmer l'existence d'un lien « solide » entre ces deux variables.

L'existence de relations entre certaines variables de personnalité et le dévoilement pourrait notamment dépendre du sexe de la personne étudiée. Ainsi, Pedersen et Breglio (1968) ne trouvent aucune relation significative entre personnalité et dévoilement lorsqu'il s'agit de femmes. Ils montrent par contre que les hommes émotionnellement instables dévoileraient plus d'informations sur leur personnalité, leur santé et leur apparence physique que les hommes non instables. Toutefois, ils ne trouvent aucune corrélation significative entre le trait de masculinité-féminité et les mesures de divulgation, ce qui tend à infirmer ce résultat.

On pensait également que l'estime de soi pourrait influencer la facilité avec laquelle un individu accepte de se dévoiler aux autres. Ainsi, dans la mesure où les gens ayant une bonne estime d'eux-mêmes sont sûrs de leurs opinions et ressentent peu le besoin d'approbation ou de support des autres, ils trouveraient moins d'utilité à se dévoiler. Au contraire, les gens ayant une faible estime d'eux-mêmes pourraient ressentir le besoin de se dévoiler aux autres, de manière à gagner leur attention. Les gens à l'estime de soi élevée devraient donc moins se dévoiler que ceux ayant une faible estime de soi. Cette hypothèse n'est cependant pas confirmée par la littérature. Ainsi, les résultats de l'étude menée par Fitzgerald (1963) semblent montrer que le niveau d'estime de soi n'affecte pas de manière significative le montant d'informations données. De plus, pour cet auteur, deux personnes au niveau d'estime de soi dissemblable pourraient révéler le même montant d'informations mais pour des raisons différentes. Un individu à l'estime de soi élevée peut révéler des informations personnelles parce qu'il les considère importantes, tandis que ceux à faible estime de soi peuvent se

[56] Cité par Pederson et Higbee (1969)

dévoiler pour se rassurer ou tester leur utilité, à travers la réaction des autres. Et Altman et Taylor (1973) de conclure qu'il est irréaliste d'espérer trouver des liens entre des traits de personnalité et le dévoilement de soi.

Cette étude sur l'impact des variables individuelles montre que les résultats sont peu concluants concernant les variables de personnalité, la plupart des études ayant du mal à confirmer les résultats trouvés dans les travaux précédents. Les conclusions sur le sexe sont plus encourageantes car la majorité des recherches montre que les femmes se dévoilent plus que les hommes.

3.3 L'influence des facteurs situationnels

De nombreuses études montrent que les facteurs situationnels peuvent jouer un rôle important dans la détermination du niveau de dévoilement choisi par l'individu. Pour West (1971) l'individu est flexible dans son dévoilement puisqu'il adapte son discours aux caractéristiques de la situation. La capacité à s'adapter ainsi est d'ailleurs liée à un fonctionnement interpersonnel réussi.

Parmi les facteurs situationnels influençant le dévoilement, on trouve le médium, c'est-à-dire le fait que le dévoilement soit oral ou écrit, sur papier ou par l'intermédiaire d'un ordinateur. Concernant le cas particulier du dévoilement par le biais d'un ordinateur, la théorie de la réponse sociale prédit que les gens traitent les ordinateurs comme des acteurs sociaux, même s'ils savent que les machines ne possèdent pas de sentiments ou de motivations (Nass et al. 1994). Par conséquent, la plupart des conventions qui guident les comportements interpersonnels sont aussi valables dans les interactions homme/ordinateur, même si ces règles n'ont pas de sens rationnel dans ce contexte (Nass et al. 1994).

Benner (1968)[57] note que le dévoilement est fonction des caractéristiques sociodémographiques de celui qui se dévoile, mais aussi de la situation sociale, du sujet de la discussion, de la cible (personne à qui on se confie) et de la relation entre les deux individus. Jourard et Lasakow (1958) confirment que les gens se dévoilent différemment selon la cible (ils se confient ainsi plus facilement à leur mère qu'à leur père ou qu'à un étranger). Ils montrent aussi que les gens se dévoilent différemment selon le sujet de la discussion. Enfin, ils trouvent un effet d'interaction significatif entre la cible et le sujet.

[57] Cité par Chelune (1975)

Le fait que le comportement soit différent selon le sujet abordé tend ainsi à prouver que le dévoilement dépend de la pertinence de celui-ci, d'une part et du degré d'intimité ou de sensibilité des informations qui s'y réfèrent, d'autre part.

La volonté de se dévoiler dépend naturellement de la nature de l'information à transmettre. Les données ayant un caractère « public » telles que nos centres d'intérêt, ce qui concerne le travail ou les études seraient plus facilement confiées que les données « privées », liées à l'ego (Jourard et Lasakow 1958). Taylor et Altman (1975) montrent aussi que le dévoilement est plus important pour les sujets non intimes (vs intimes), qu'il varie en fonction d'un ratio coûts/bénéfices et que l'impact du ratio est plus important pour les sujets intimes (effet d'interaction). Les divulgations « intimes » sont celles qui contiennent des informations à risque élevé et qui conduisent l'individu à se sentir vulnérable (Derlega et al. 1993). Dans le domaine interpersonnel, la divulgation d'informations sur ses émotions et/ou ses sentiments mènerait ainsi à une perception de vulnérabilité extrême (Derlega 1988), celle-ci pouvant être émotionnelle ou liée à une atteinte physique. La littérature suggère que les coûts associés à la divulgation d'informations intimes sont généralement élevés (Altman et Taylor 1973 ; Laurenceau et Barrett 1998), même si cela n'est pas toujours le cas. De plus, certains auteurs indiquent, qu'à un niveau extrême d'intimité, le lien entre dévoilement et appréciation de l'autre pourrait se rompre. Le fait de dévoiler des données intimes peut en effet être considéré comme inapproprié (violation des normes sociales), surtout en cas d'échange avec un étranger. Cela peut mettre celui qui les reçoit dans une situation d'embarras.

Nous résumons à travers le tableau suivant (i.e. tableau 3.1) l'ensemble des conclusions issues de la littérature concernant les facteurs d'influence du dévoilement de soi, qu'ils soient de nature culturelle (et/ou environnementale), individuelle ou situationnelle.

Tableau 3.1 Facteurs d'influence du dévoilement de soi identifiés dans la littérature

Facteurs d'influence	Auteurs (année)	Résultats et commentaires
Facteurs culturels et/ou environnementaux		
Nationalité	Lewin (1948) ; Plog (1965)	Dans certains pays, on se dévoilerait plus que dans d'autres (exemple : Américains vs. Allemands)
Origine ethnique	Jourard et Lasakow (1958)	Certaines « ethnies » se dévoileraient plus que d'autres (exemple : « blancs » vs. « noirs »)
Conclusion : l'influence des facteurs culturels et/ou environnementaux est clairement établie		

Facteurs d'influence	Auteurs (année)	Résultats et commentaires (SUITE)
Facteurs individuels		
Sexe	Jourard et Lasakow (1958) ; Cosby (1973)	les femmes se dévoileraient plus que les hommes
	Pedersen et Breglio (1968)	Les femmes donneraient des informations plus intimes que les hommes
	Jourard (1971)	Les femmes valoriseraient davantage le dévoilement que les hommes
Conclusion : l'influence du sexe est clairement établie		
Degré d'introversion	Mullaney (1964) ; Pedersen et Breglio (1968), Pedersen et Higbee (1969)	L'hypothèse selon laquelle « moins une personne se dévoile, plus elle est introvertie » ne se confirme pas dans toutes les études
Estime de soi	Fitzgerald (1963)	L'hypothèse selon laquelle « les gens à l'estime de soi élevée (vs. faible) devraient moins se dévoiler » n'est pas confirmée par la littérature
Conclusion : les liens entre les traits de personnalité et le dévoilement de soi sont difficiles à établir. Cela pourrait s'expliquer par l'existence d'effets d'interaction avec d'autres variables et notamment le sexe (Pedersen et Breglio 1968)		
Facteurs situationnels		
Sujet de la discussion	Benner (1968)	On parlerait plus facilement d'un sujet qui nous intéresse
Degré d'intimité des informations	Jourard et Lasakow (1958) ; Taylor et Altman (1975)	On dévoilerait plus facilement des informations non intimes (ayant un caractère « public »)
Cible	Jourard et Lasakow (1958)	On se confierait plus facilement à un ami qu'à un étranger
Conclusion : l'influence du sujet de la discussion, de la cible et du degré d'intimité des informations est clairement établie au sein de la littérature		

Le dévoilement de soi serait donc clairement lié à la relation que l'on entretient avec son interlocuteur (cible). Il dépendrait ainsi de l'avancement de la relation (4.1), de l'appréciation de l'autre (4.2) et ferait intervenir des conditions de réciprocité (4.3).

4 Dévoilement de soi et relation à l'autre

Jourard (1966) considère que le déterminant majeur du dévoilement de soi est l'identité de la personne à laquelle on se confie ainsi que la nature et l'état de la relation entre ces deux personnes. Nous verrons que le dévoilement est une composante critique de la formation et de l'évolution des relations que nous établissons avec les autres (Altman et Taylor 1973) et qu'il répond à deux principes majeurs : l'attraction et la réciprocité.

4.1 Dévoilement selon l'avancement de la relation

Nous devons tout d'abord distinguer les caractéristiques du dévoilement selon la relation établie avec la personne à qui nous nous confions, et ce, à deux niveaux : 1) le niveau de connaissance, correspondant au lien que nous avons avec cette personne (ami vs. étranger) ; 2) le niveau d'avancement de la relation (ami de longue date vs. nouvel ami).
Concernant le premier niveau (connaissance), une absence de dépendance entre les partenaires, comme dans le cas d'un dévoilement à un étranger fait disparaître certaines contraintes (notamment en termes de réciprocité) et augmenterait alors le dévoilement (en quantité), mais seulement quand la période d'interaction entre les deux individus est brève (Cosby 1973). Le second niveau (avancement de la relation) fait, notamment, appel aux conclusions issues de la théorie de la pénétration sociale (Altman et Taylor 1973), que nous présentons ci-dessous.

On peut considérer le dévoilement de soi comme un échange social dans un contexte relationnel. Le travail théorique le plus abouti dans ce domaine est présenté par Altman et Taylor (1973). Leur théorie de la pénétration sociale suggère que les relations se développent, à travers une augmentation graduelle du montant et de la profondeur du dévoilement des partenaires. Le niveau de dévoilement de chaque individu est alors considéré comme une mesure de son degré d'ouverture (Taylor 1968). Selon cette approche, le dévoilement est alors une récompense pour celui qui la reçoit, parce qu'elle communique le désir du divulgateur d'initier une relation intime avec lui. De plus, la divulgation à un ami proche (vs. à une simple connaissance) est considérée comme plus appropriée car elle s'adresse à quelqu'un de confiance.

La théorie de la pénétration sociale (Altman et Taylor 1973) considère donc la formation de l'amitié comme un processus au cours duquel, plus la relation se développe, plus les individus sont censés révéler un grand nombre d'informations sur eux et à un niveau de sensibilité de plus en plus élevé. Frankfurt (1965)[58] confirme cette hypothèse en montrant une augmentation du montant et du degré d'intimité de la divulgation, au fur et à mesure de l'avancement de la relation. Notons que, comme le dévoilement est de plus en plus intime, les normes concernant la profondeur et l'ampleur de la divulgation se modifient en parallèle.

[58] Cité par Cosby (1973).

Altman et Taylor (1973) étudient le caractère approprié (ou non) du dévoilement selon le moment auquel il intervient dans l'évolution de la relation que l'on établit avec l'autre. Ils considèrent alors que, même si le fait de se dévoiler signale l'intérêt de l'individu de développer une relation plus intime, il y a des moments où un tel dévoilement peut être considéré comme inapproprié. Ainsi, le fait de vouloir se dévoiler dès le début de la relation (i.e. dès la première rencontre) peut être considéré comme trop rapide et peu appréciable. Il semble donc exister des règles sociales strictes qui indiquent à quel moment une information peut être révélée aux autres et dans quel contexte (Derlega et Grzelak 1979).

4.2 Dévoilement et appréciation de l'autre

Dès 1959, Jourard note une relation entre le fait d'apprécier une personne et de se dévoiler à elle. Depuis, de nombreux auteurs suggèrent qu'on se dévoile plus facilement aux gens qu'on apprécie (Fitzgerald 1963) et que l'on apprécie davantage les gens qui se dévoilent à nous (Archer, Berg et Runge 1980). Le lien entre les deux concepts, bien qu'évident de prime abord, est cependant loin d'être aussi simple dans la réalité. Tout d'abord, ce lien n'est pas forcément direct. Ainsi, les modèles d'attraction suggèrent que le fait d'apprécier l'autre et d'être attiré par lui est déterminé par les sentiments à son égard. Selon cette approche, le lien entre divulgation et appréciation passe donc par la formation de croyances positives. Ensuite, comme le montrent certaines études (citées ci-après), la relation entre le dévoilement de soi et le fait d'apprécier l'autre ne serait pas toujours positive (absence de lien ou relation négative). Enfin, la relation ne serait pas forcément linéaire. Cosby (1973) pense ainsi qu'il existe une relation curvilinéaire entre divulgation et appréciation : l'appréciation serait plus forte quand la divulgation est modérée et plus faible quand la divulgation est très faible ou très élevée.

Dans une méta-analyse sur la relation entre le dévoilement de soi et le fait d'apprécier l'autre, Collins et Miller (1994) identifient trois liens (effets) possibles entre ces deux variables : 1) les individus qui se dévoilent de façon intime sont préférés à ceux qui dévoilent des informations moins intimes ; 2) les individus se dévoilent plus facilement aux gens qu'ils apprécient ; 3) les individus apprécient les personnes auxquelles ils se dévoilent (résultat du dévoilement). La relation entre dévoilement de soi et appréciation de l'autre serait modérée par quelques variables comme le type de dévoilement et le sexe de celui qui se dévoile.
Nous avons déjà parlé de l'effet 1 à deux reprises. D'une part, quand nous avons évoqué l'influence des facteurs situationnels sur le dévoilement de soi, et notamment l'effet de la sensibilité des informations dévoilées (cf. paragraphe précédent). D'autre part, quand nous

avons discuté du dévoilement selon l'avancement de la relation. Nous ne reviendrons donc pas dessus, préférant nous attarder plus particulièrement sur les deux autres liens possibles.

L'effet 2 cherche à mesurer si nous nous confions plus aux personnes que l'on apprécie. Dans la mesure où la divulgation est un acte risqué, il semble raisonnable de penser que les gens seront sélectifs dans le partage des informations. Jourard (1964) suggère que nous nous dévoilons plus aux gens que nous apprécions et en qui nous avons confiance. La théorie de la pénétration sociale de Altman et Taylor (1973) suggère que l'on anticipe un bénéfice à faire en sorte que les gens nous connaissent mieux. Worthy et al. (1969) confirment cette idée en indiquant que la divulgation est une récompense pour celui qui la reçoit et que nous voulons davantage récompenser les gens que nous apprécions. Ainsi, tout déterminant de l'appréciation de l'autre pourrait servir d'antécédent à un degré élevé de dévoilement de soi (Lynn 1978). Il semble que ce soit également le cas lors de la première rencontre, même si l'appréciation ne fournit, dans ce cas, qu'un montant limité de la variance expliquée du dévoilement. Collins et Miller (1994) montrent aussi que les hommes et les femmes ne diffèrent pas dans leur tendance à se dévoiler aux gens qu'ils apprécient (pas d'effet modérateur du sexe). De son côté, Frankfurt (1965), déjà cité précédemment, montre qu'on se dévoile plus à des partenaires appréciés (vs. non appréciés) quand il s'agit de sujets intimes. Il y aurait donc un effet d'interaction entre le niveau d'intimité de la conversation et l'appréciation de la personne à laquelle on se confie.
Le fait qu'on se dévoile plus aux gens qu'on apprécie ne signifie cependant pas qu'on ne se dévoile qu'à eux ni qu'il soit nécessaire d'apprécier l'autre pour se dévoiler. Il semble d'ailleurs que la divulgation ait parfois lieu en l'absence d'appréciation (Archer 1979 ; Derlega, Harris et Chaikin 1973). Il peut donc y avoir de multiples motivations à la divulgation dont l'une d'elles est le fait d'apprécier l'autre.

L'effet 3 cherche à étudier si se dévoiler à quelqu'un fait qu'on l'apprécie davantage par la suite. Si nous nous dévoilons, cela signifie que nous avons confiance en l'autre, que nous valorisons ses opinions et ses réponses et que nous sommes intéressés par le fait de mieux le connaître et d'être connu de lui (Derlega et al. 1993). La divulgation sert donc une fonction symbolique importante dans les relations interpersonnelles. De même, d'après le modèle de l'échange social, les gens devraient être attirés par ceux qui leur permettent d'obtenir une récompense en recevant leur dévoilement. Il n'y a cependant pas de preuve montrant que la divulgation est bénéfique pour celui qui s'est dévoilé (Archer 1979). Il existe même de nombreux exemples de situations où cela est considéré comme désagréable ou stressant.

Comme le signale Taylor (1979), la divulgation est à la fois bénéfique et risquée. Elle peut ainsi entraîner des sentiments négatifs - amenant l'individu à comparer son état actuel avec celui qu'il souhaiterait avoir - qui peuvent conduire à moins apprécier celui à qui on s'est confié, surtout s'il s'agit d'expériences personnelles douloureuses. Les résultats de la méta-analyse menée par Collins et Miller (1994), montrent toutefois que, dans l'ensemble, l'effet 3 est bien vérifié. De plus, plus le dévoilement porte sur des choses intimes plus nous apprécions celui envers qui nous nous sommes révélés.

On peut ajouter, à ces trois liens identifiés par Collins et Miller (1994), un quatrième effet, également étudié au sein de la littérature, selon lequel on apprécie les individus qui acceptent de se dévoiler. Cet effet serait cependant lié au caractère approprié ou non du dévoilement en question. En effet, si la divulgation est considérée comme appropriée et socialement bénéfique, elle peut conduire à une plus forte attraction envers celui qui se dévoile (Jourard 1959, 1971 ; Worthy et al. 1969). A l'opposé, une divulgation vue comme inappropriée ou non bénéfique peut être non liée ou non à l'acceptation (*acceptance*) de se dévoiler en retour.

4.3 Dévoilement et réciprocité

Il existerait une attente de réciprocité, associée aux dévoilements de soi volontaires (Derlega, Wilson et Chaiken 1976). Le fait de se dévoiler est en effet associé au fait que la confiance témoignée par ces révélations sera rendue – ou récompensée - par une révélation de l'autre, à son tour. En ce sens, le dévoilement de soi peut être vu comme un moyen d'accéder à des informations concernant l'autre.

Mauss (1990) définit la réciprocité comme « le principe par lequel chacun accepte de donner et de recevoir ». Pour Blau (1964) « donner est un moyen d'obtenir du pouvoir ». S'il existe une relation entre A et B et si A donne à B, alors B est redevable à A. Il s'agit donc de l'idée selon laquelle, lorsqu'un individu fournit à un autre des informations sur sa personne, il l'oblige à donner quelque chose en retour. Ce principe serait lié à la notion de distance sociale : quand une des parties ne reçoit pas de retour approprié, la distance sociale avec l'autre partie augmente. Selon Gouldner (1960), cette norme de réciprocité serait universelle, tandis que pour d'autres auteurs elle serait culturellement spécifique.

La probabilité pour qu'il y ait réciprocité serait liée à trois facteurs principaux : 1) l'opinion concernant les raisons du dévoilement de l'autre (une attribution positive augmenterait les

chances de réciprocité), 2) le niveau d'intimité (ou de profondeur) de la révélation reçue ; 3) le type de relation souhaitée et/ou le stade de développement de la relation entre les partenaires. Nous insisterons plus particulièrement sur les deux premiers.

Levin et Gergen (1969) prédisent une relation curvilinéaire entre la divulgation de l'autre et sa propre divulgation. Un montant de dévoilement moyen, de la part du partenaire, soulignerait son désir de nouer une relation plus intime, tandis qu'en se dévoilant trop, il pourrait être considéré comme indiscret et ne mériterait donc pas qu'on lui fasse confiance. En réalité, ils trouvent une relation linéaire : plus l'autre se dévoile, plus on se dévoile en retour.

Pour Berg et Archer (1983), toute divulgation qui interviendrait en réponse à une divulgation de l'autre, se ferait à un niveau d'intimité supérieur. Sous cette norme de réciprocité, on trouve l'idée selon laquelle la personne ajuste la profondeur de son dévoilement à celui de la divulgation reçue, de manière à maintenir l'équité de l'échange - égalité des coûts et bénéfices pour chacune des parties - (Derlega et al. 1993). De plus, les individus qui croient que l'autre les apprécie, se dévoileraient de manière plus intime (Curtis et Miller 1986). Cosby (1973) indique que si une divulgation intime peut être considérée comme une récompense élevée pour celui qui la reçoit, elle engendre aussi des coûts non négligeables, notamment une forte anxiété, du fait de devoir (au sens de se sentir obligé de) révéler plus d'informations qu'on ne le souhaiterait. Il pense donc que l'effet de réciprocité serait dans ce cas réduit. Plusieurs études confirmeront cette hypothèse, en montrant que l'effet de réciprocité ne serait pas linéaire : il serait moins important - tout en restant significatif - en cas de dévoilement de l'autre à un niveau d'intimité élevé. Concernant l'impact du niveau de développement de la relation, et contrairement à ce qu'on pourrait penser, Derlega et al. (1976) montrent que les amis ont une moins forte réciprocité que les étrangers. Cela pourrait être dû au fait que les normes de réciprocité sont plus fortes au début d'une relation (Altman 1975).

Ainsi, dans l'ensemble, nous serions non seulement attirés davantage par les personnes qui se sont confiées à nous, mais aussi par celles à qui nous nous sommes confiés, ce qui engendrerait des dévoilements réciproques. Ces effets de réciprocité et d'attraction, très développés dans la littérature, semblent donc robustes et expliqueraient - au moins en partie – certains de nos comportements, à ce sujet.

Conclusion du chapitre 3

Le dévoilement de soi peut se définir comme le processus par lequel on donne des informations personnelles aux autres (Jourard et Lasakow 1958). Parmi les principaux attributs du dévoilement, on distingue l'ampleur (quantité d'informations dévoilées), la profondeur (niveau d'intimité des faits révélés) et le temps passé à parler de soi. Le fait de donner des informations sur soi est d'ailleurs considéré par de nombreux chercheurs, comme un signe d'ouverture aux autres et de bonne santé mentale (Altman et Taylor 1973).

Le dévoilement de soi est généralement considéré comme un processus d'échange, au cours duquel l'individu donne une partie de lui-même contre des bénéfices sociaux ou relationnels. La théorie de l'échange social fournit, à ce titre, un des principes majeurs d'un tel échange, selon lequel la décision de révéler des informations est liée au calcul des coûts et bénéfices du dévoilement. Ce calcul que Laufer et Wolfe (1977) appellent « *calculus of behavior* », serait basé sur l'estimation des conséquences futures d'une telle décision.

Le dévoilement de soi serait soumis à une triple influence : celle des facteurs culturels, individuels et situationnels. Il serait ainsi fonction du sexe de la personne qui se dévoile, mais aussi de la cible (personne à qui on se confie), du sujet de la discussion (et du degré d'intimité qui lui est lié) et de la relation entre les partenaires (Benner 1968)[59]. Deux principes majeurs guideraient alors l'échange : l'attraction – qui lie dévoilement et attrait pour le partenaire - et la réciprocité, qui prédit que se dévoiler oblige l'autre à se confier en retour.

Les recherches en psychologie ont montré que, d'une manière générale, le dévoilement de soi conduit à des rencontres sociales satisfaisantes car il accroît la confiance et l'attraction vis-à-vis du partenaire. Il semble par ailleurs que, lors d'une négociation commerciale, le type et la magnitude des révélations entre un client et un vendeur peuvent refléter la force de la relation entre les partenaires. Développés à l'origine dans le domaine interpersonnel, les principes qui s'appliquent au dévoilement de soi, pourraient ainsi être transférés au domaine commercial, dans le cadre de relations entre un consommateur et une entreprise.

Nous abordons, dans le chapitre suivant, le lien entre les préoccupations pour le respect de la vie privée et les réactions des consommateurs face à une sollicitation de données personnelles.

[59] Cité par Chelune (1975)

Partie I - COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

CHAPITRE 4
SOLLICITATION DE DONNÉES PERSONNELLES ET RÉACTIONS DES CONSOMMATEURS : SYNTHÈSE DE LA LITTÉRATURE

« Consumers worry that supposedly legitimate companies will take advantage of them by invading their privacy to capture information about them for marketing and other secondary purpose without their informed consent ».

US Public Interest Research Group (2000)

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 4

Nous avons vu, dans un premier chapitre, quels étaient les enjeux liés à la collecte de données consommateurs, notamment sur Internet. Nous avons poursuivi l'analyse de la littérature en étudiant le concept de vie privée, et avons insisté sur la notion de protection des données à caractère personnel, concept majoritairement mobilisé dans les recherches. Nous avons ensuite investigué les théories psychosociales du dévoilement de soi, en insistant, plus particulièrement, sur les facteurs à l'origine du dévoilement d'une part, et sur le lien existant entre le dévoilement et la relation à l'autre, d'autre part. Les différents principes issus de ces théories, développés à l'origine dans le domaine interpersonnel, peuvent néanmoins s'appliquer au dévoilement de données dans le cadre d'une sollicitation commerciale (en particulier sur Internet)[60], à condition toutefois de prendre les précautions de rigueur[61].

Nous terminons ici l'étude de la littérature en analysant les liens existants entre la sollicitation et les réactions des consommateurs. Nous verrons, tout d'abord, de quelle manière les chercheurs ayant déjà travaillé sur ce sujet évaluent les réactions de ces consommateurs, que ce soit au niveau attitudinal ou comportemental. Nous étudierons aussi les raisons pouvant expliquer l'écart entre les convictions profondes des individus (leurs préoccupations au sujet du respect de la vie privée) et leur réaction effective, quand ils se trouvent sollicités.

La deuxième section de ce chapitre sera consacrée à l'étude des facteurs pouvant influencer les réactions des consommateurs quand ils se trouvent sollicités pour fournir des données personnelles. Nous retrouvons ainsi un découpage traditionnel en facteurs environnementaux, individuels et situationnels. Nous détaillons les éléments spécifiques à chacun d'eux.

Nous terminons ce chapitre par l'étude d'un paradigme émergent dans les travaux portant sur la sollicitation de données clients, à savoir la théorie du contrat social. Après avoir présenté les principes de base de cette théorie, nous montrons en quoi elle s'applique utilement à la sollicitation de données consommateurs, en particulier sur Internet.

---

[60] En atteste le fait que la quasi-totalité des chercheurs travaillant sur le sujet (i.e. dévoiement du consommateur) procèdent comme tel (avec succès), démontrant ainsi la possibilité de le faire. Moon (2000) vient confirmer ce principe en s'appuyant sur la théorie de la réponse sociale (évoquée au chapitre 3, paragraphe 3.3) qui prédit que les gens traitent les ordinateurs comme des acteurs sociaux.

[61] Il s'agit notamment, quand on s'intéresse au domaine commercial (en particulier sur Internet), de prendre en compte les particularités de la cible et de la relation qu'on entretient avec elle (éloignement, absence de face à face, …). Certains principes issus de l'étude des relations interpersonnelles peuvent ainsi ne pas s'appliquer (ou être difficilement transposables), notamment celui qui concerne la réciprocité.

1 Sollicitation de données personnelles et réactions des consommateurs

La réaction du consommateur face à une sollicitation de ses données personnelles peut être étudiée à plusieurs niveaux. De nombreuses recherches se sont ainsi attachées à mesurer la préoccupation des personnes face aux possibles invasions de leur vie privée par les entreprises (aspect attitudinal). Ainsi, Cespedes et Smith (1993) ont, par exemple, montré que la préoccupation des consommateurs pour la protection de leur vie privée augmentait quand ils apprenaient qu'on avait collecté des données relatives à leur personne, sans les avoir préalablement mis au courant, ni avoir demandé leur accord. D'autres chercheurs ont plutôt raisonné au niveau du comportement des individus. Milne (2000) a ainsi proposé un cadre d'analyse visant à tenir compte des antécédents affectant à la fois les stratégies des entreprises en la matière (réglementation en vigueur, codes de conduite de l'industrie, éthique, actions des concurrents et technologie) et le comportement des consommateurs, en cas de sollicitation effective (données sociodémographiques, perceptions, facteurs situationnels et technologie).

On constate, cependant, une forte disparité des recherches dans ce domaine. En effet, si de nombreux auteurs s'intéressent à la préoccupation des consommateurs pour le respect de leur vie privée (donc pour la protection de leurs données personnelles)[62], peu de travaux étudient l'aspect comportemental. Pour Regan (1995), la littérature actuelle est d'ailleurs limitée du fait qu'elle se soit insuffisamment intéressée au comportement des individus.
Nous verrons, dans une première partie, la mesure des réactions comportementales des consommateurs en cas de sollicitation de leurs données personnelles (1.1), avant d'étudier le lien entre la préoccupation d'une part et le comportement d'autre part (1.2). Nous terminerons en évoquant les raisons pouvant expliquer les écarts entre ces deux niveaux d'analyse (1.3).

1.1 La mesure des réactions comportementales des consommateurs

Comme nous venons de le souligner, peu de recherches ont étudié le comportement[63] des consommateurs, en cas de sollicitation de leurs données personnelles. Une des raisons possibles au manque de recherches comportementales a été suggérée par Webb (1981).

[62] Par préoccupation pour le Respect de la Vie Privée (RVP), nous entendons le niveau de préoccupation ressenti par l'individu concernant les menaces qui pèsent sur sa vie privée lorsque des entreprises sollicitent et/ou utilisent ses données personnelles. Nous reviendrons sur la définition de ce terme dans les chapitres 7 et 9.
[63] Nous incluons, dans cette catégorie, les recherches mesurant des intentions comportementales, comme celle de Milne et Rohm (2000) qui étudie le désir de retirer son nom d'un fichier.

Celui-ci explique qu'il est difficile d'étudier les comportements adoptés par les individus face aux problèmes liés à la vie privée, parce que ceux-ci sont susceptibles de varier en fonction de la situation rencontrée. Il estime qu'il existe en effet un continuum de situations, pour lesquelles le comportement peut être observé. A une extrémité de ce continuum, on trouve les situations où il n'y a pas violation et à l'autre, celles où la violation est flagrante. Entre les 2, les frontières sont floues et le fait qu'un individu estime qu'il y a violation ou non, dépendra des caractéristiques de la situation et de son propre jugement de celle-ci. Cette nature contextuelle de la vie privée a d'ailleurs été reconnue, à la fois par les chercheurs (Schoeman 1984) et les autorités régulatrices.

Parmi les études ayant cherché à mesurer les réactions comportementales (intentions ou comportements réels) des consommateurs, une majorité d'entre elles se sont intéressées à l'achat en ligne (Jarvenpaa et Tractinsky 1999 ; Miyazaki et Fernandez 2000 et 2001) ou à l'utilisation ultérieure des données plutôt qu'à leur réaction en cas de sollicitation. Ces dernières concernent notamment le fait d'accepter de s'inscrire sur une liste de marketing direct ou d'autoriser le transfert de ses données à des tiers (Milne 1999), l'acceptation d'ouvrir un mailing (Fournier 1999) ou de voir ses données utilisées pour faire du « profiling »[64] (Culnan et Armstrong 1999, Farag et Krishnan 2003) ou bien encore de recourir à des services de personnalisation (Chellappa et Sin, 2005). Seules quelques études investiguent réellement la réponse du consommateur face à une sollicitation de ses données personnelles. La mesure est alors, le plus souvent, sous la forme d'une fréquence (de jamais à toujours). Des auteurs, comme Miyazaki et Krishnamurthy (2002) évaluent aussi la réponse en demandant aux répondants d'indiquer leur niveau d'accord (échelle de format Likert) pour fournir un ou plusieurs types de données (i.e. nom, adresse, etc, ..). L'ensemble des variables comportementales identifiées dans la littérature sont présentées au tableau 4.1.

Sheehan et Hoy (1999) ont été les premiers à évoquer les différentes solutions qui se présentent au consommateur lorsqu'il se trouve confronté à un formulaire à remplir. Ces auteurs analysent en effet les stratégies adoptées par les internautes face à deux catégories de pratiques : la collecte d'informations d'une part, et l'utilisation des données pour des contacts marketing, d'autre part (sous la forme d'emails non autorisés). Leurs résultats suggèrent d'ailleurs qu'il n'y a pas de réponse spécifique à chacune de ces pratiques, ce qui signifie que l'internaute adopte un comportement holistique, qui ne dépendrait pas du traitement en cause.

[64] Nous considérons ce terme sous l'angle utilisation de données pour établir un profil et proposer ensuite des offres adaptées et personnalisées.

Sheehan et Hoy (1999) identifient ainsi 6 types de comportements susceptibles d'être adoptés par les consommateurs (internautes) en réponse au traitement de leurs données (les deux premiers concernant plus particulièrement la réponse à une sollicitation) :

1) fournir des informations incomplètes et/ou erronées ;

Cette solution semble fournir une réponse idéale au paradoxe soulevé par Cespedes et Smith (1993) sur la manière de participer au e-commerce sans perdre le contrôle de sa vie privée.

2) ne pas s'enregistrer sur le site (ne pas divulguer d'informations personnelles) ;

Cette solution est vue comme le moyen optimal pour protéger sa vie privée.

3) demander à se faire rayer des listes (fichiers) ;
4) ne pas lire les emails non autorisés ;
5) se plaindre auprès des firmes responsables de l'envoi d'emails non sollicités ;

Cette solution est globalement peu utilisée, mais elle constitue un type de réponse possible, en cas de niveau d'insatisfaction élevé.

6) notifier au fournisseur d'accès la réception d'emails non sollicités

Ceci suggère que les individus cherchent une assistance vers qui se tourner pour se plaindre.

Pour leur part, s'inspirant des travaux de Stone et al. (1983), Stewart et Segars (2002) étudient plus particulièrement quatre comportements, qui reprennent certains de ceux identifiés précédemment : la tendance à retirer son nom d'un fichier, à refuser de divulguer des données, à se plaindre auprès de l'entreprise indélicate ou d'organismes gouvernementaux.

Tableau 4.1 Principales variables comportementales identifiées dans la littérature

Auteurs	Variables comportementales
Sheehan et Hoy (1999)	6 comportements de réponse : enregistrement sur le site, fourniture d'informations erronées ou incomplètes, …
Milne (1999)	Inscription sur une liste de marketing direct, autorisation de transfert des données à des tiers
Jarvenpaa et Tractinsky (1999)	Intention d'achat en ligne
Milne et Rohm (2000)	Désir de retrait du fichier
Phelps, Nowak et Ferrell (2000)	Intention d'achat à distance, désir de retrait du fichier
Miyazaki et Fernandez (2000)	Intention d'achat en ligne
Miyazaki et Fernandez (2001)	Achat en ligne
Stewart et Segars (2002)	4 intentions de comportement : retirer son nom du fichier, refuser de divulguer des données personnelles, se plaindre à l'entreprise ou aux organisations gouvernementales
Miyazaki et Krishnamurthy (2002)	Probabilité de divulguer 6 catégories de données, probabilité d'acheter en ligne
Chellappa et Sin (2005)	Probabilité de recourir à des services de personnalisation

Nous pouvons replacer les listes proposées par Sheehan et Hoy (1999) et Stewart et Segars (2002) dans des modèles plus généraux visant à élaborer une typologie de comportements. Hirschman (1970) a, par exemple, été le premier à suggérer une taxonomie des réponses à l'insatisfaction à travers son modèle EVL (*Exit Voice Loyalty*). Des recherches ultérieures montrent d'ailleurs que la dernière option (*Loyalty* = loyauté), caractérisée par le fait de souffrir en silence, est choisie par la majorité des répondants. Singh (1988) identifie, pour sa part, 3 catégories de réponses un peu différentes de celles du modèle EVL. Ce modèle, qu'on pourrait appeler « VPT », distingue les trois catégories de réponses suivantes : *Voice* (réponse à l'entreprise), *Private* (en parler aux connaissances) et *Third Party* (en parler à des tiers non directement impliqués dans l'échange comme des organismes professionnels ou des associations de défense des consommateurs). Il n'est toutefois pas certain que les stratégies de réponse identifiées par ces deux auteurs correspondent parfaitement aux situations de divulgation d'informations personnelles. En effet, comme le soulignent les résultats de Sheehan et Hoy (1999), d'autres tactiques pourraient être utilisées par les consommateurs en réaction à la sollicitation de leurs données, comme le fait de mentir (fourniture d'informations erronées). Sheehan et Hoy (1999) suggèrent d'ailleurs d'élaborer une typologie de réponses plus spécifique aux questions de respect de la vie privée.

En 1995, Regan faisait remarquer que si des relations existaient entre les attitudes et les comportements en matière de vie privée, aucune étude ne les avait jusqu'à présent analysées. Depuis, plusieurs travaux ont cherché à mesurer les effets de la préoccupation pour le respect de la vie privée sur le comportement du consommateur, en cas de sollicitation de ses données. Nous indiquons à présent les principaux résultats établis dans ce domaine.

1.2 L'impact de la préoccupation sur le comportement des consommateurs

La majorité des études menées à ce sujet confirment l'impact de la préoccupation pour le respect de la vie privée sur le comportement du consommateur face aux pratiques de traitement de ses données personnelles, et notamment concernant l'utilisation de celles-ci. Ainsi, Culnan et Armstrong (1999) montrent que les individus très préoccupés sont moins enclins à fournir des informations personnelles dans le but d'être profilés, résultat qui sera confirmé ultérieurement par Farag et Krishnan (2003). De même, Chellappa et Sin (2005) montrent que la probabilité de recourir à des services personnalisés est négativement influencée par la préoccupation pour le respect de la vie privée. Enfin, Phelps, Nowak et Ferrell (2000) trouvent une relation forte entre le niveau de préoccupation, les croyances

relatives aux pratiques des entreprises, d'une part, et le comportement en résultant, d'autre part. Les individus les plus préoccupés désapprouveraient ainsi plus volontiers le fait qu'une entreprise détienne des informations les concernant et seraient alors plus nombreux à demander à retirer leur nom du fichier.

Il a également été démontré que la préoccupation influençait le comportement de réponse face à la sollicitation d'informations. Ainsi, Smith, Milberg et Burke (1996) montrent que les individus préoccupés auraient des comportements spécifiques, comme le fait de se plaindre ou de ne pas divulguer leurs données. De même, Stewart et Segars (2002) trouvent que les individus préoccupés ont une plus forte tendance à retirer leur nom des fichiers, à refuser de fournir des données personnelles ou à se plaindre. Enfin, Sheehan et Hoy (1999) démontrent l'impact de la préoccupation sur 4 des 6 comportements de réponse présentés précédemment, dans le sens suivant : plus la préoccupation est élevée, moins les individus acceptent de s'enregistrer, plus ils donnent des informations incomplètes, plus ils notifient les problèmes à leur fournisseur d'accès et plus ils se plaignent auprès des entreprises responsables. Ils ne trouvent cependant aucune corrélation significative entre le niveau de préoccupation, le fait de lire des emails non sollicités, d'une part et de fournir des informations erronées, d'autre part. Ceci suggère que le lien entre préoccupation et comportement, bien que largement démontré, ne serait pas forcément validé dans tous les cas de figure. Sheehan et Hoy (1999) reconnaissent d'ailleurs que leurs résultats pourraient ne pas correspondre à la réalité, dans la mesure où peu de personnes adoptent un comportement de protection dans le cadre de leur étude alors que les consommateurs y ont largement recours en pratique.

Au-delà d'une absence de lien entre la préoccupation et le comportement (loin d'être remis en cause), ce qui est avancé ici, c'est le décalage qui pourrait exister entre un niveau général de préoccupation pour le respect de la vie privée et la manière de réagir concrètement, une fois confronté au formulaire à remplir. L'écart concernerait donc la différence entre les convictions personnelles des individus et leur manière de réagir en situation réelle. Les raisons possibles d'un tel écart seront exposées brièvement au paragraphe suivant. Bien que difficiles à identifier et à mesurer, elles conduisent à remettre en cause la majorité des études déjà réalisées sur le sujet qui adoptent une logique de sondage. Dans ce cas en effet, les répondants doivent indiquer de quelle manière ils pensent qu'ils réagiraient, ce qui peut être différent de la façon dont ils se comporteraient s'ils étaient réellement face au formulaire à compléter.

1.3 L'écart entre les convictions personnelles et le comportement réel

Plusieurs sondages relatifs au respect de la vie privée, menés notamment aux Etats-Unis [65], montrent que la majorité des individus se disent préoccupés par la question et cherchent à se protéger. L'atteinte aux données personnelles serait ainsi considérée comme la préoccupation la plus importante des internautes (Kekoe et al. 1998), ce qui aurait une incidence non négligeable sur l'acceptation de faire des achats ou de se dévoiler en ligne. En revanche, d'autres recherches prouvent que peu de gens se protègent réellement, même quand le coût de la protection s'avère limité, voire totalement gratuit. Ainsi, les consommateurs seraient de plus en plus confiants quant au respect de leur intimité sur Internet, en atteste la progression fulgurante de l'achat en ligne, notamment en France. Pour Gandy (1993), ce phénomène est lié au fait que de plus en plus d'individus réalisent qu'ils doivent abandonner un peu de leur intimité pour participer au e-commerce. Le confort et la valeur que fournit l'achat en ligne, seraient alors en mesure, pour certains d'entre eux au moins, de dépasser leur niveau de préoccupation dans ce domaine. De même, Singer et Presser (1996) constatent que l'attitude des citoyens américains face au problème de confidentialité de leurs données personnelles n'est pas un facteur qui explique leur degré de participation aux sondages, soulignant ainsi l'absence de lien entre un niveau général de préoccupation et leur degré de coopération.

Plusieurs recherches récentes menées sur ce thème soulignent donc que certains individus, une fois en ligne, oublient leurs préoccupations et ont tendance à se dévoiler, même quand les questions sont personnelles ou en l'absence de raison objective d'agir ainsi (Spiekermann, Grossklags et Berendt 2001). Cet écart entre les croyances et la réponse effective est un phénomène connu en comportement du consommateur, lié aux aspects sociaux et cognitifs qui interviennent lors d'une prise de décision. Nous essayons, dans les paragraphes suivants, d'identifier les raisons susceptibles d'expliquer et de justifier un tel écart.

De tels écarts entre une attitude générale et un comportement effectif, pourtant conceptuellement liés, ont été observés dans de nombreux aspects de la vie humaine et ont été étudiés en psychologie sociale (LaPiere, 1934). Certains de ces écarts pourraient ainsi être attribués aux procédures même de la recherche, telles que le recours à un environnement contrôlé dans les expérimentations, à opposer à un fort biais de désirabilité sociale[66] lors d'un

[65] Par exemple : Jupiter (1999) et Price Waterhouse (2000), cités par Acquisti et Grossklags (2004).
[66] Nous reviendrons sur cette notion et la définirons plus en détail dans le chapitre 8.

sondage. Ainsi, dans ce dernier cas, les répondants se disent majoritairement concernés par la question du respect de la vie privée (alors qu'ils ne le sont pas forcément) parce qu'ils pensent que c'est la réponse que l'on attend d'eux.

Une autre explication du phénomène est apportée par Acquisti et Grossklags (2004). Ces auteurs indiquent que, comme tout autre bien, la valeur accordée au respect de sa vie privée, est affectée par des asymétries d'informations, des interprétations subjectives et des distorsions psychologiques, toutes susceptibles de biaiser les résultats. Nous détaillons l'ensemble de ces raisons dans les paragraphes qui suivent.

Le processus de décision, dans le cadre de la réponse à une sollicitation de données personnelles, pourrait être l'objet d'une information incomplète et de rationalité limitée.

Dans le domaine du respect de la vie privée en effet, les individus ne disposent pas toujours de l'ensemble des informations nécessaires pour faire leur choix, en particulier en ce qui concerne la politique de confidentialité et les pratiques réelles de l'entreprise qui les sollicite. Certains individus ne connaissent pas non plus les risques associés à la divulgation de leurs données, ni l'existence de certaines mesures de protection. Cette information incomplète peut alors affecter l'estimation des risques et bénéfices liés à la divulgation et donc conduire à une décision qui semble incohérente.

Concernant le principe de rationalité, Acquisti (2004) estime qu'il est irréaliste de croire que les individus sont systématiquement rationnels dans un tel contexte. En effet, si les modèles économiques traditionnels postulent que les individus cherchent à maximiser leur bien-être, les recherches récentes en comportement du consommateur montrent que plusieurs inconsistances psychologiques se heurtent à cette idée de rationalité. Dans une situation de sollicitation de données personnelles, les individus pourraient donc davantage faire preuve de rationalité limitée. Celle-ci fait référence à l'incapacité à calculer et à comparer toutes les conséquences associées aux différents choix possibles pour l'individu, ainsi qu'à l'incapacité à analyser toutes les informations disponibles (Simon, 1976). Dans un tel contexte en effet, l'individu se trouve souvent dans une situation d'asymétrie d'information qui l'oblige à baser ses choix sur des estimations difficiles à réaliser, d'autant que certaines des conséquences ne sont connues qu'*ex post*. Les coûts cognitifs associés à l'analyse de la situation sont alors généralement très élevés et obligent la personne à se baser sur de simples heuristiques (raisonnement global) plutôt que sur une analyse détaillée de l'ensemble des paramètres.

La seconde catégorie de raisons concerne la subjectivité du consommateur lorsqu'il évalue la situation de sollicitation de ses données personnelles. Du fait de l'incertitude face à l'avenir,

les individus pourraient en effet comparer les dommages potentiels liés à la perte de contrôle (évalués subjectivement) avec la probabilité pour qu'un tel résultat se produise. Ils percevraient alors les coûts et la probabilité de pertes comme étant relativement faibles. Comparés aux coûts de la protection (notamment celui de ne pas fournir les données demandées), ils pourraient alors décider, soit de ne pas se protéger (i.e. répondre), soit de renoncer, ce choix étant largement guidé par la situation proposée. Pour Acquisti et Grossklags (2003), une des explications à ce phénomène est liée au fait que les perceptions des individus relatives à la situation de sollicitation de leurs données peuvent varier selon qu'elles sont considérées d'un point de vue théorique (comme c'est généralement le cas dans les sondages) ou réellement expérimentées (cas réel ou simulé à travers une expérimentation).

La littérature en psychologie sociale met également en évidence l'existence de distorsions psychologiques susceptibles d'affecter le comportement, en l'occurrence la décision de se dévoiler ou non (Acquisti 2004). Cet auteur cite notamment les problèmes d'auto-contrôle et de gratification immédiate. Ce dernier est défini par O'Donogue et Rabin (1999) comme la préférence d'un individu pour un bien-être immédiat plutôt que différé, d'autant plus que l'échéance est proche. Le problème d'auto-contrôle est lié, quant à lui, à la tendance de l'individu à préférer une gratification immédiate plutôt que différée, même si celle-ci s'avère peu efficace (ou rentable) à long terme. Ces deux aspects inciteraient donc certains individus à déclarer qu'ils souhaitent se protéger des risques d'invasion de leur vie privée mais à ne pas le faire en réalité, s'ils identifient des bénéfices à se comporter ainsi. Dans le cadre d'une sollicitation de données personnelles, certaines personnes, préférant pourtant ne pas se dévoiler, opteraient alors pour une gratification immédiate, prenant la forme d'une récompense offerte par l'entreprise en échange de leurs données personnelles. D'autres individus souffriraient de biais d'optimisme les amenant à considérer les risques qu'ils encourent comme inférieurs à ceux auxquels les autres sont confrontés dans une situation identique. Enfin, la majorité des gens auraient des difficultés à prendre en compte des risques cumulatifs, qui sont pourtant courants dans le domaine de la vie privée. En effet, une fois les données personnelles divulguées à une entreprise, celles-ci deviennent utilisables par cette dernière et éventuellement par d'autres pendant de longues périodes d'où un risque global supérieur à la somme des risques associés à chaque donnée divulguée. Du fait de ces distorsions psychologiques, certains chercheurs considèrent ainsi les individus comme des personnalités multi-facettes, dont il est difficile d'évaluer le processus de décision.

Pour toutes ces raisons, de plus en plus de recherches soulignent l'existence d'un écart entre la manière dont les individus agissent dans la réalité et leurs préférences générales dans ce domaine. Plus précisément, dans le cadre d'une sollicitation de données personnelles, il apparaît que les consommateurs donnent généralement plus d'informations que leur attitude générale face au respect de la vie privée ne le laisse présager, parfois même, en contrepartie de récompenses minimes (Acquisti et Grossklags 2004). Cette réaction peut cependant être considérée comme une manière rationnelle de répondre aux incertitudes, ce que Lemley (2000) qualifie d'« ignorance rationnelle ».

Au final, la décision d'un individu de fournir (ou non) les données qui lui sont demandées ne serait donc pas uniquement liée à des caractéristiques individuelles (et notamment à ses préoccupations en termes de respect de la vie privée) mais aussi à l'effet d'autres facteurs (dits situationnels), dont l'impact ne serait d'ailleurs effectif qu'en cas de situation réelle. L'influence de ces deux types de facteurs (et des facteurs environnementaux) sur l'attitude et le comportement de l'individu en cas de sollicitation de ses données est à présent étudiée.

2 Les antécédents aux préoccupations et aux réactions des consommateurs

Nous verrons, dans un premier temps, qu'il existe trois grandes catégories d'antécédents (2.1), avant de détailler l'influence de chacun d'entre eux. Nous débutons par l'impact des facteurs environnementaux (2.2), avant d'aborder les antécédents individuels (2.3). Nous terminons par l'étude de l'influence des facteurs situationnels (2.4).

2.1 Un modèle à 3 entités

Dans le contexte du marketing direct et des bases de données, la préoccupation pour la protection des données à caractère personnel[67] est influencée par un mix multidimensionnel de facteurs (Turn 1985, Gandy 1993). Ces facteurs sont si nombreux et si divers qu'aucune liste exhaustive n'existe à ce jour. De plus, certains chercheurs ont analysé l'impact de ces facteurs sur le niveau de préoccupation tandis que d'autres étudient leur influence sur la réaction du consommateur face à la sollicitation de ses données, ce qui ajoute à la confusion. Il est néanmoins possible d'indiquer ceux qui sont souvent étudiés par les chercheurs qui travaillent sur ce sujet, ce à quoi nous nous attachons dans cette section.

[67] Pour rappel, nous emploierons parfois le terme de respect de la vie privée (RVP) pour désigner la protection des données personnelles.

Les variables ayant un effet sur la préoccupation du consommateur en matière de protection de ses données et/ou sur sa réaction en cas de sollicitation peuvent être classées dans 3 catégories distinctes, qui correspondent à 3 niveaux d'influence (Patterson, O'Malley et Evans 1997). Bien que certains d'entre eux aient déjà été évoqués à travers l'étude des caractéristiques du droit à la protection de la vie privée (chapitre 2) et du dévoilement de soi (chapitre 3), nous les citons à nouveau dans ce chapitre. En effet, nous nous intéressons ici, plus spécifiquement, à l'influence directe ou indirecte de ces facteurs sur le comportement de réponse du consommateur en cas de sollicitation de ses données.

Au premier niveau, on trouve les variables liées à l'interaction en cours (que nous appellerons l'environnement « spécifique »). De tels facteurs, dits situationnels, comme le type de données collectées et l'entreprise à l'origine de la sollicitation, conduisent à une évaluation majoritairement cognitive de la situation, et influencent la manière dont le consommateur perçoit la sollicitation de ses données personnelles.

Le niveau 2 correspond aux variables de nature individuelle, et est à rapprocher de ce que Patterson, O'Malley et Evans (1997) appellent l'« environnement idéologique ». De tels facteurs influencent également l'évaluation de la situation, notamment dans sa dimension affective. Parmi ceux-ci, on trouve : le niveau de connaissance de l'individu vis-à-vis des pratiques des entreprises en matière de collecte et d'utilisation de données personnelles et son expérience en la matière. Nous incluons également, à ce niveau, les variables d'ordre sociodémographique et psychologique, dans la mesure où elles sont propres à l'individu.

Le niveau 3 est constitué de variables « culturelles », liées au macro-environnement autour duquel la sollicitation a lieu. Celui-ci inclut, notamment, l'infrastructure et la technologie du pays en matière de collecte et d'utilisation des données ainsi que la législation en vigueur.

L'influence de ces trois niveaux peut être synthétisée à travers le modèle présenté figure 4.1.

La particularité de ce modèle (adapté de Patterson, O'Malley et Evans 1997) est d'être à la fois dynamique et intégrateur, dans le sens où chaque niveau de variables est influencé par les autres et/ou influence lui-même les autres. Ainsi, chaque consommateur évalue la situation sur la base d'éléments cognitifs, qui sont eux-mêmes influencés par sa disposition idéologique. Cette dernière dépend à son tour de l'environnement culturel dans lequel l'individu se trouve. Il faut toutefois souligner que l'influence entre ces 3 niveaux de variables se fait principalement du macro-environnement vers l'environnement spécifique. Ainsi, les variables spécifiques à l'interaction en cours n'influencent pas directement l'environnement idéologique de l'individu, si ce n'est à travers l'expérience que chaque situation lui permet

d'acquérir. Ce sont cependant davantage les conséquences de ces interactions plutôt que les interactions elles-mêmes, qui constituent et donc influencent cette expérience.

Ce modèle s'avère conforme à la majorité des recherches sur le sujet qui indiquent que la décision du consommateur de divulguer ses informations personnelles varie à la fois selon l'individu et le contexte (Cranor, Reagle et Ackerman 1999). La dimension « macro-environnement » est en effet occultée dans de nombreuses études, étant donné que les travaux sont généralement menés dans un pays unique.

Nous allons à présent évoquer, plus en détail, chacun de ces facteurs et indiquer leur effet sur les variables dépendantes majoritairement étudiées dans la littérature, à savoir la préoccupation générale pour la protection des données d'une part et la réaction du consommateur en cas de sollicitation effective d'autre part.

Figure 4.1 Catégories de variables influençant la réaction à une sollicitation de données

Adaptée de Patterson M., O'Malley L. et Evans M. (1997)

2.2 Les antécédents environnementaux

Les antécédents environnementaux concernent les aspects sociaux, culturels et légaux qui entourent les pratiques de collecte et d'utilisation des données consommateurs. Bien que leur effet soit largement admis au sein de la littérature, il n'est pas toujours clairement démontré, d'un point de vue empirique. En effet, si plusieurs chercheurs abordent la question (comme Zhang, Wang et Chen 2001 ou Hann et al. 2002), ils sont peu nombreux à mesurer véritablement ce phénomène, Milberg et al. (1995) ou encore Milne et al. (1996) mis à part.
De plus, il manque encore d'études transnationales de grande ampleur qui prennent véritablement en compte cet aspect culturel. Celles qui le font sont en effet souvent basées sur des échantillons d'étudiants et/ou de faible taille, ce qui réduit la portée des résultats obtenus. C'est par exemple le cas de l'étude menée par Cockcroft et Clutterbuck (2001), auprès de 170 étudiants de nationalités différentes, qui montre que les Australiens sont plus préoccupés que les Américains au sujet de la protection de leurs données personnelles. Ces travaux offrent cependant des résultats encourageants et permettent de conclure, malgré le manque de validité externe, à l'influence plus que probable des facteurs culturels et environnementaux sur la préoccupation des individus et leur réaction face à une sollicitation.

S'agissant tout d'abord des normes sociétales en matière de protection des données, il faut souligner que celles-ci diffèrent d'un Etat à l'autre, et d'une période à l'autre au sein d'un même pays (Kelvin 1973). Il semble néanmoins logique de penser que l'attitude des consommateurs concernant la protection de leurs données à caractère personnel est liée à la manière dont les citoyens de ce pays considèrent ce droit comme absolu ou relatif, à une période et/ou dans un contexte donnés. Derlega et chaikin (1977) notent ainsi que les normes culturelles américaines invitent les femmes à se dévoiler davantage que les hommes.

De nombreux auteurs identifient une dimension culturelle à la préoccupation pour la préservation des données personnelles (Goodwin 1991, Maynard et Taylor 1996, Milne et al. 1996, Petrison et Wang 1995). Milne et al. (1996) ont ainsi testé un modèle expliquant les effets de la culture sur les attitudes des consommateurs vis-à-vis du marketing direct et de la question du respect de la vie privée. Dans ce modèle, la culture est évaluée à travers des éléments tels que la technologie et l'infrastructure du pays dans le domaine de la collecte et de l'utilisation des données consommateurs (i.e. pénétration des différents médias, technologies employées en matière de marketing direct, etc ...).

Milberg et al. (1995) ont également étudié l'influence de l'environnement sur l'attitude des consommateurs en matière de préservation de leur vie privée d'une part, et sur le système de régulation en vigueur dans le pays étudié, d'autre part. Ils mettent ainsi en évidence le fait que l'attitude des individus face au respect de leur vie privée est différente selon les pays et la réglementation en place. Une relation en forme de U inversé apparaît entre le degré de régulation et le sentiment d'être concerné par les problèmes de protection des informations personnelles. Ils trouvent ainsi qu'un faible niveau de préoccupation est associé aux pays sans régulation ou avec le plus haut niveau d'implication de l'Etat dans la régulation. Cela suggère alors une relation complexe entre ces différentes variables, selon laquelle lorsque le niveau de connaissance des citoyens sur le sujet est faible, le niveau de préoccupation l'est aussi. Une augmentation de ce niveau de préoccupation conduit à un accroissement de la législation en vigueur dans le pays jusqu'au moment où les préoccupations diminuent à nouveau.

Sur Internet, cette dimension environnementale est d'autant plus importante que les risques encourus sont élevés. Elle serait donc à même d'influencer la capacité de l'individu à évaluer la situation d'une part et à utiliser les différentes options qui s'offrent à lui, d'autre part. En ligne, les internautes sont en effet dépendants à la fois de la réglementation en vigueur dans ce domaine et des codes de bonne conduite mis en place par les entreprises pour protéger leurs données clients. Ainsi, d'après les études menées à ce sujet, lorsque le consommateur estime que ces deux types de régulations sont insuffisants pour le protéger, il est davantage préoccupé, ce qui le conduit à des mesures défensives, telles que la fourniture d'informations incomplètes ou erronées (Sheehan et Hoy, 1999).

2.3 Les antécédents individuels

Parmi les antécédents individuels étudiés dans la littérature comme étant susceptibles d'influencer la préoccupation du consommateur en terme de protection de ses informations personnelles d'une part, et la réaction en résultant en cas de sollicitation de celles-ci, d'autre part, on trouve des variables appartenant à quatre catégories différentes (cf. figure 4.2) :

- d'ordre sociodémographique
- liées à l'expérience, à la connaissance et à l'attitude de l'individu relatives aux pratiques de collecte et d'utilisation de données personnelles
- d'ordre psychologique, liées à la personnalité de l'individu
- d'ordre idéologique, liées aux valeurs de l'individu

De nombreuses recherches visent à démontrer l'effet d'une ou de plusieurs de ces variables individuelles sur le niveau de préoccupation sans parvenir, dans la majorité des cas, à confirmer les résultats issus d'études précédentes. Il faut toutefois remarquer que les écarts entre ces travaux, notamment en termes de période (année de publication) et d'échantillon (origine géographique des répondants, méthode d'échantillonnage) peuvent expliquer certaines conclusions contradictoires, tout comme le manque de variance au niveau des caractéristiques des individus interrogés. Cela amène en tout cas à se poser la question de l'influence réelle ou seulement supposée de ces facteurs individuels. La distinction entre l'existence d'un effet et son sens (positif ou négatif) est, à ce titre, importante, dans la mesure où si des chercheurs trouvent un effet significatif pour certaines variables individuelles, il ne va pas toujours dans le sens supposé, ce qui ne permet pas de valider empiriquement ces hypothèses. Nous allons à présent détailler l'impact de ces quatre catégories d'antécédents individuels, en distinguant l'existence d'un effet et, le cas échéant, son sens.

Figure 4.2 Les antécédents individuels étudiés dans la littérature

2.3.1 Les variables d'ordre sociodémographique

Concernant les facteurs sociodémographiques, on peut distinguer leur effet sur le niveau de préoccupation, et sur la réaction attitudinale et surtout comportementale de l'individu en cas de sollicitation de ses données[68]. Nous insisterons plus particulièrement sur les variables les plus étudiées dans la littérature, à savoir le sexe, l'âge, le niveau d'études et la CSP et aborderons plus brièvement les autres, dans la mesure où le plus petit nombre de recherches à leur sujet empêche généralement de conclure de manière formelle. Nous avons regroupé dans le tableau 4.2 des exemples de recherches ayant abordé cette question ainsi que les conclusions auxquelles elles ont abouti. Ce tableau, bien que non exhaustif, permet de visualiser la contradiction apparente qui transparaît de tous ces résultats. Si l'absence de conclusion uniforme n'invite pas à conclure à l'influence de ces facteurs individuels sur les variables dépendantes étudiées, cela n'autorise toutefois pas à en déduire qu'ils n'ont aucun effet, celui-ci pouvant être indirect ou soumis à des conditions particulières. Il conviendra donc, dans toute recherche menée sur ce sujet, de prendre en compte l'existence d'un effet possible de ces variables, bien que celui-ci n'ait jamais été formellement démontré.

D'un point de vue démographique, la variable dont l'influence sur le niveau de préoccupation envers la protection des données personnelles est la plus étudiée dans la littérature est le sexe. A ce sujet, deux types de théories s'opposent. D'un côté, les tenants des théories sur le dévoilement de soi qui prédisent que, dans la mesure où les femmes parlent plus facilement d'elles-mêmes (Archer 1979, Cosby 1973), elles seraient moins préoccupées par la protection de leurs données et accepteraient plus volontiers de donner des informations personnelles à une entreprise. De l'autre, les théoriciens de la vie privée qui expliquent que les femmes ont une plus forte préférence pour le respect de leur intimité et qu'elles seraient donc plus préoccupées par la protection de leurs données personnelles que les hommes.
Le fait que la sollicitation s'opère sur Internet amène à considérer de nouveaux facteurs, qui viennent compliquer davantage la situation. En effet, plusieurs études montrent des différences entre les sexes concernant l'orientation à la technologie et son usage (Gilroy et Desai 1986, Frankel 1990), ces écarts pouvant influencer à leur tour le comportement sur Internet, et notamment le dévoilement de données personnelles en ligne. Dans une étude menée auprès d'internautes américains, Sheehan (2002) ne trouve cependant aucune

[68] Certains des travaux menés concernent l'impact de ces facteurs sur la réaction des consommateurs dans le cadre de la participation à un sondage. Même si celle-ci est différente de la réaction face à une sollicitation de données personnelles, nous présenterons les résultats issus de ces recherches, dans une logique d'accumulation.

différence de préoccupation entre les deux sexes. En réalité, comme le montre le tableau 4.2, l'absence de résultats concordants concernant l'effet du sexe amène à se poser la question de l'existence même d'un tel effet, au moins s'agissant d'un effet direct. Il se pourrait en effet que l'influence de cette variable soit indirecte, par exemple en interaction avec d'autres variables. C'est en tout cas ce que suggèrent les résultats trouvés par Sheehan (1999) qui étudie l'impact des données sociodémographiques, à la fois sur la préoccupation envers la protection des données et sur certains comportements sous-jacents comme la demande de retrait d'un fichier ou le fait de fournir des données erronées ou incomplètes. Alors que cet auteur ne trouve aucune différence entre les sexes en termes de préoccupation, il montre que les femmes demandent à supprimer leur nom du fichier quel que soit leur niveau de préoccupation, alors que les hommes le font surtout quand leur préoccupation s'accroît.

Concernant l'âge, là encore, les thèses s'affrontent. D'un côté, certains chercheurs comme Dommeyer et Gross (2003), qui expliquent que les gens âgés ont plus d'expérience au sujet des sollicitations de marketing direct, qu'ils connaissent davantage les moyens de se protéger et qu'ils sont donc dans l'ensemble moins préoccupés que les jeunes. De l'autre, ceux qui disent qu'étant davantage sollicités, ils sont plus susceptibles de demander leur retrait des fichiers (Milne et Gordon 1993). Cette dernière hypothèse semble d'ailleurs souvent se confirmer dans les faits, puisqu'un grand nombre de recherches concluent à une plus grande préoccupation chez les personnes âgées. C'est cependant loin d'être systématiquement le cas, Gervey et Lin (2000) trouvant au contraire que les jeunes ont une attitude plus positive que les gens âgés concernant la collecte de leurs données personnelles, notamment si on leur promet une compensation en échange. Ce résultat laisse, là encore, supposer l'existence possible d'un effet d'interaction entre ces deux catégories de variables (l'âge et le niveau de compensation). S'agissant d'une collecte de données sur Internet, on estime généralement que les plus jeunes étant davantage familiers avec cet environnement seront moins préoccupés que leurs aînés. Malgré l'existence de travaux allant dans ce sens (Harris 1995, Liebermann et Stashevsky 2002), l'absence de généralisation possible de ces résultats ne permet pas de conclure de manière certaine à l'existence d'un tel effet.

S'agissant du niveau d'études, Stone et Stone (1990) estiment que plus il est élevé, plus la capacité cognitive de la personne l'est aussi et donc plus elle est capable de structurer son environnement de manière à protéger sa vie privée, réduisant dans le même temps le niveau de préoccupation qui lui est associé. A l'opposé, plus la capacité cognitive de l'individu est faible, moins il pourra développer de cognitions cohérentes et justes concernant les liens entre

les informations dévoilées et les conséquences à venir, ce qui le conduit à être davantage préoccupé et donc plus réticent à répondre. Dans les faits, cette hypothèse n'a jamais pu se vérifier de manière certaine. Ainsi, alors que Rosenbaum (1973) montre une relation entre le niveau d'études de recrutés et leur niveau de préoccupation par rapport au fait de devoir donner des informations aux entreprises, une étude menée par le *Georgia Institute of Technology* en 1998 ne trouve aucune relation significative entre ces deux variables.

Enfin, concernant la Catégorie Socioprofessionnelle (CSP) – certains auteurs étudient plutôt la profession ou le niveau socio-économique de l'individu – les résultats sont tout aussi déroutants, même si un grand nombre d'études tendent à montrer l'influence de ce type de variables. Ainsi, plusieurs chercheurs (dont Kelvin 1973, Laufer et Wolfe 1977) trouvent que les individus ayant un faible pouvoir social au sein de la société (par exemple ceux qui sont sans emploi) tendent à sacrifier leur intimité pour obtenir des résultats positifs de la part du collecteur d'informations (par exemple un travail). Au contraire, Sheehan et Hoy (2000) ne trouvent aucune relation significative entre la CSP et le niveau de préoccupation de l'individu en matière de protection de ses données. De plus, même lorsque des travaux tendent à montrer l'existence d'un lien entre cette variable et certaines croyances et/ou comportements en matière de dévoilement et/ou d'utilisation de données personnelles, le sens d'un tel effet est loin d'être évident, puisqu'il est parfois inversé d'une étude à l'autre. Toutefois, l'absence d'unité dans la manière d'opérationnaliser ces variables est peut-être pour beaucoup dans l'incapacité à valider ce résultat.

D'autres variables individuelles de nature sociodémographique – au sens large – sont également étudiées dans les recherches menées sur ce thème. Il s'agit notamment de l'origine ethnique et/ou géographique de l'individu ou du niveau de revenus du foyer. Etant moins étudiées que les précédentes, il est plus difficile de conclure à l'impact de ces variables, puisque l'on manque souvent de matière pour pouvoir valider, le cas échéant, des résultats qui iraient dans ce sens. Nous n'aborderons donc pas les résultats déjà trouvés en la matière, nous contentant de souligner l'effet possible, mais non prouvé à ce jour, de telles variables.

Nous reprenons dans le tableau ci-après les principaux résultats trouvés dans la littérature concernant l'impact des variables sociodémographiques. Nous les avons classés par facteur analysé (i.e. sexe, âge, …), par type de variable dépendante étudiée (i.e. préoccupation, attitude, comportement), et, plus accessoirement, par ordre chronologique.

Tableau 4.2 Impact des variables sociodémographiques dans le cadre d'une sollicitation de données personnelles

Variable socio-démographique(*)	Auteurs (année)	Variable dépendante étudiée	Résultats
Sexe	Singer, Mathiowetz et Couper (1993)	Niveau de préoccupation	Pas de relation significative
	Westin (2001)		Significatif (femmes plus préoccupées que les hommes)
	Kekoe et al. (1997)	- Niveau de préoccupation - Fourniture d'informations erronées	Significatif (femmes plus préoccupées que les hommes) Significatif (plus de mensonge chez les femmes)
	Sheehan (1999)	- Niveau de préoccupation - Fourniture d'informations incomplètes	Pas de relation significative Significatif (les femmes le font plus que les hommes)
	Milne et rohm (2000)	Désir de retirer son nom du fichier	Pas de relation significative
	Nowak et Phelps (1992)	Demande de retrait du fichier	Pas de relation significative
Age	Campbell (1997)	Niveau de préoccupation	Pas de relation significative
	Phelps Nowak et Ferrell (2000)		
	Graeff et Harmon (2002)		Significatif (les plus âgés sont plus préoccupés)
	Gervey et Lin (2000)	Attitude face à la collecte	Significatif (les jeunes ont une attitude plus positive)
	Milne et rohm (2000)	Désir de retirer son nom du fichier	Significatif (les plus âgés désirent davantage un retrait)
	Novak et Phelps (1992)	Demande de retrait du fichier	Significatif (les jeunes ont plus demandé le retrait)
Niveau d' études	Rosembaum (1973)	Niveau de préoccupation	Significatif (les gens éduqués sont moins préoccupés)
	Nowak et Phelps (1992)		Pas de relation significative
	Etude Georgia Institute of Technology (1998)		
	Milne et Rohm (2000)	Désir de retrait du fichier	Pas de relation significative
	Sondage Harris (1995)	Acceptation de fournir des données	Significatif (les gens éduqués sont plus enclins à refuser)
CSP et/ou Statut socio-économique (**)	Wang et Petrison (1993) Milne et al (1996)	Niveau de préoccupation	Significatif (les gens de bas statut sont plus préoccupés)
	Sheehan et Hoy (2000)		Pas de relation significative
	Robins (1963)	Participation à un sondage	Significatif (les «cols blancs» sont plus nombreux à refuser)
	Zhang, Wang et Shen (2001)	Acceptation de fournir des données	Significatif (les étudiants sont plus enclins à refuser)

(*) Seules les quatre variables sociodémographiques les plus étudiées dans la littérature (sexe, âge, niveau d'études et CSP) sont reprises dans ce tableau.

(**) Mesuré par le niveau d'études, l'origine ethnique et la profession

2.3.2 La connaissance, l'expérience et l'attitude des consommateurs

Outre les données sociodémographiques, d'autres facteurs individuels sont également susceptibles d'influencer les croyances et les comportements des individus en matière de dévoilement d'informations personnelles. Il en va ainsi des variables liées à la connaissance, à l'expérience et à l'attitude des consommateurs dans le domaine de la protection de la vie privée mais aussi concernant le marketing direct, l'achat à distance et les nouvelles technologies, tous domaines qui sont liés de près ou de loin aux pratiques des entreprises en matière de collecte et/ou d'utilisation de données clients.

En matière de connaissance de l'individu, il faut distinguer celle relative aux pratiques des entreprises de celles liées à la réglementation en vigueur et aux moyens de protéger ses données personnelles. Nous insisterons plus volontiers sur la première, dans la mesure où elle est la plus étudiée au sein de la littérature. La connaissance des consommateurs concernant les pratiques des entreprises en matière de collecte et d'utilisation de données clients a été identifiée par certains chercheurs comme une variable clé dans la détermination de leurs préoccupations à ce sujet (Culnan 1995, Foxman et Kilcoyne 1993). Pour ces auteurs, plus les individus sont informés (connaissance élevée), plus ils se sentent concernés par la collecte et l'utilisation possibles de leurs données. Walczuch, Seelen et Lundgren (2001) trouvent au contraire que plus le consommateur a connaissance des pratiques des entreprises à ce sujet, plus il leur fait confiance. D'autres travaux ne montrent, quant à eux, aucun lien entre la connaissance des pratiques et le niveau de préoccupation (Campbell, 1997) ou le compromis coûts/bénéfices lors du dévoilement de données personnelles (Hann et al. 2002). Dans certains cas en effet, si les individus comprennent que la collecte de leurs informations personnelles a pour but de bâtir une relation à laquelle ils pourront participer, leur préoccupation pourrait diminuer ou être supplantée par leur désir de participer à cette relation (Glazer 1991). Enfin, les résultats trouvés par Nowak et Phelps (1992) suggèrent l'existence d'une relation non linéaire entre la connaissance et le niveau de préoccupation. Ils montrent, en effet, que le niveau de préoccupation est plus élevé quand il y a connaissance des pratiques mais que, plus cette connaissance augmente, moins la préoccupation est importante.

La connaissance des pratiques se situe en réalité à 2 niveaux : tout d'abord, la connaissance générale que les consommateurs ont des procédés utilisés par les entreprises dans ce domaine (notamment par le biais de reportages diffusés à travers les différents médias ou de bouche à oreille issu de l'entourage), et ensuite, le résultat de l'exposition et de l'expérience

personnelle à ce sujet (expérience de collecte illicite ou d'envoi de propositions commerciales non désirées, par exemple). Ces notions de connaissance et d'expérience semblent donc étroitement liées, dans la mesure où elles s'influencent mutuellement au cours du temps. Ainsi, il ne peut y avoir connaissance sans un minimum d'expérience dans ce domaine, qu'elle soit directe (personnelle) ou indirecte (par le biais de médias ou du bouche à oreille). La connaissance sur le sujet pourrait influencer en retour l'expérience vécue par le consommateur, puisqu'un individu informé n'aurait pas le même comportement de dévoilement de ses données personnelles qu'un novice. Ainsi, Culnan et Armstrong (1999) montrent qu'une expérience précédente positive accroît la confiance des individus à se dévoiler, de par la connaissance et la compréhension des bénéfices qu'elle procure.

En matière d'expérience, la distinction entre le fait qu'elle soit directe ou indirecte est primordiale. Il faut également tenir compte de la valence du résultat de cette expérience, positif ou négatif. Cette double distinction a une influence non négligeable sur les préoccupations de l'individu et sur sa réaction, en cas de sollicitation de ses données personnelles. Ainsi, Campbell (1997) montre que selon le type d'expérience dont on parle (directe ou indirecte), l'effet sur la préoccupation est différent. Alors que l'expérience personnelle (directe) influence les préoccupations en termes de collecte et d'erreurs, l'expérience indirecte est corrélée aux préoccupations concernant l'accès et l'utilisation secondaire des données. En ce qui concerne le sens (positif ou négatif) de cette expérience, il semblerait qu'une expérience personnelle positive n'ait aucun effet sur le comportement ultérieur de l'individu. A l'opposé, plus un individu a vécu des expériences personnelles négatives dans ce domaine, plus il risque de se montrer réticent à divulguer des informations à l'avenir (Stone et Stone 1990). De plus, une expérience personnelle négative avec une entreprise compromet ses relations futures avec toutes les firmes, car cette expérience diminue sa confiance dans la manière dont toute entreprise est susceptible d'utiliser les informations de ses clients. Enfin, une expérience indirecte négative, par exemple à travers des incidents reportés dans la presse, n'aurait qu'un effet très faible - voire nul – sur l'attitude face aux pratiques de collecte et d'utilisation de données (Phelps, Gonzenbach et Johnson 1994). En effet, un aspect fondamental de la psychologie sociale montre qu'un individu a plus confiance dans ses propres expériences que dans celles qui lui sont reportées, dans la mesure où l'information directe est plus coûteuse à obtenir (Robinson et Pearce 1983).

Une autre variable susceptible d'influencer la préoccupation des consommateurs en matière de protection de leurs données personnelles est leur attitude face au marketing direct,

technique employée par les entreprises pour leur faire parvenir des propositions commerciales ciblées mais pas toujours sollicitées ni désirées. Ainsi, Milne et Gordon (1993) font l'hypothèse que le désir d'éviter les sollicitations de marketing direct est un indicateur du désir de protéger sa vie privée. Pour ces auteurs, les consommateurs qui souhaitent recevoir moins de sollicitations utiliseront davantage les techniques de protection, et dévoileront moins facilement leurs informations personnelles. Leurs résultats montrent que les différents segments de consommateurs, élaborés sur la base des préférences en termes de protection de la vie privée (pragmatiques, lobbyistes ..) ont une attitude différente face aux techniques de marketing direct. De même, Culnan et Armstrong (1999) trouvent que l'expérience de marketing direct est la seule variable discriminante significative de l'acceptation de voir ses données personnelles utilisées pour établir un profil. Si l'influence de l'expérience en termes de marketing direct semble donc établie, le sens (positif ou négatif) d'un tel effet est toutefois davantage sujet à caution. Ainsi, alors que Campbell (1997) suggère que les individus qui sont les plus exposés aux offres de marketing direct ont un niveau de préoccupation élevé, Horne et Horne (1997) notent que les gros utilisateurs de marketing direct sont moins préoccupés par la protection de leurs données personnelles.

L'attitude des consommateurs face au marketing direct est fortement corrélée à leur expérience et à leur attitude en matière d'achat à distance. En effet, un individu qui apprécie de recevoir chez lui des offres commerciales ciblées sera plus enclin à y répondre favorablement, par le biais d'un achat à distance. Dès lors, l'expérience en matière d'achat à distance serait aussi liée au niveau de préoccupation des individus relatif à la collecte et à l'utilisation des données personnelles. Cette hypothèse se trouve confirmée par un grand nombre d'études menées dans ce domaine. Ainsi, Culnan (1995) montre que les individus à l'attitude positive vis-à-vis de l'utilisation secondaire de leurs données sont plus nombreux à percevoir l'achat à distance comme bénéfique. Milne et Rohm (2000) trouvent que l'expérience d'achat à distance est liée négativement au désir de retrait des listes. De même, l'analyse réalisée par Westin (1995) d'une enquête Harris-Equifax de 1994 souligne que les personnes ayant acheté à distance sont plus concernées par la protection de leur vie privée et donc plus enclins à refuser de fournir leurs données personnelles. Enfin, Phelps, Nowak et Ferrell (2000) indiquent que les gens ayant acheté à distance dans les 6 derniers mois sont moins préoccupés par la manière dont les entreprises peuvent utiliser leurs données que ceux n'ayant pas acheté. Pour ces auteurs, l'explication d'un tel résultat est liée au fait que l'acceptation par les consommateurs de fournir leurs données personnelles dans ce cadre est fonction des bénéfices qu'ils peuvent retirer de l'achat à distance. Le consentement serait

ainsi plus important quand les bénéfices attendus de l'achat à distance, tels que le confort, sont effectivement présents. Il faut cependant préciser que s'ils trouvent une relation entre le fait d'avoir acheté à distance (ou non) et le niveau de préoccupation, ils ne trouvent pas de relation significative entre cette dernière et la fréquence d'achat d'une part, et le montant dépensé, d'autre part. De plus, il faut également signaler que si le sens de la relation entre l'expérience d'achat à distance et le niveau de préoccupation est connu (négatif), sa direction l'est beaucoup moins. On peut en effet se demander lequel est l'antécédent de l'autre – achète t-on à distance parce qu'on est peu préoccupé par l'utilisation qu'on peut faire des données ou est-on peu préoccupé du fait qu'on a une expérience en la matière ? - ou encore si la relation entre les deux n'est pas concomitante ?

Le fait que l'expérience d'achat à distance soit liée à la préoccupation envers la protection des informations personnelles serait notamment vrai sur Internet Ainsi, dans l'ensemble, les personnes ayant déjà acheté en ligne seraient moins préoccupées, même si elles demandent à être informées et à pouvoir donner leur avis sur l'utilisation ultérieure de leurs données (Fletcher 2000). Pour cet auteur, ce besoin accru d'être informé pourrait d'ailleurs refléter une meilleure connaissance et compréhension des problèmes. Au-delà de l'achat en ligne, d'autres variables liées à l'utilisation de l'ordinateur en général et d'Internet en particulier seraient également influentes, comme l'anxiété vis-à-vis de l'ordinateur, le niveau d'expertise et la fréquence d'utilisation d'Internet. Ainsi, Stewart et Segars (2002) montrent que les consommateurs éprouvant une forte anxiété face aux ordinateurs auraient un niveau de préoccupation plus élevé. De même, il semble que plus l'expertise en ligne est élevée, moins la préoccupation est forte (Sheehan 2002), ce qui serait notamment dû à l'effet du contrôle perçu. Ce résultat est cependant contraire à celui trouvé par Nyshadham (2000). Pour Alexander (1998), une fréquence élevée d'usage du web serait associée à de plus bas niveaux de préoccupation et une expérience faible de l'Internet à de plus hauts niveaux d'anxiété. Ce résultat, issu d'une étude menée auprès d'étudiants, n'est cependant pas confirmé par Gauzente (2004) qui trouve que le niveau de préoccupation est influencé par plusieurs variables d'usage du web, sauf la fréquence d'utilisation. Le type de connexion (bas /haut débit) paraît, par contre, peu influent dans la majorité des situations.

Après avoir étudié l'impact des facteurs sociodémographiques et de ceux liés à la connaissance, à l'expérience et à l'attitude du consommateur, voyons à présent celui des variables psychologiques (2.3.3) puis idéologiques (2.3.4).

2.3.3 Les variables d'ordre psychologique

La préoccupation des consommateurs en matière de protection de leurs informations personnelles ainsi que leurs perceptions et leurs réactions au moment de la sollicitation pourraient aussi être influencées par des différences individuelles d'ordre psychologique telles que la tendance à se dévoiler (mesurée par le rapport introversion/extraversion), la capacité - ou la propension - à faire confiance, le degré d'individualisme et d'autres variables de personnalité comme la tendance à la paranoïa ou à la critique sociale. Si l'influence de ces variables a déjà été étudiée par certains chercheurs, le nombre d'études analysant l'impact de tels facteurs est relativement limité, ce qui oblige à rester prudent en ce qui concerne les conclusions qu'elles permettent de tirer.

Plusieurs auteurs estiment que la préoccupation des consommateurs en matière de protection de données personnelles est liée à leur orientation en matière d'introversion/extraversion. Cosby (1973) suggère ainsi que les individus introvertis sont moins enclins à se dévoiler que les extravertis. Stone (1986) estime que si les introvertis donnent moins d'informations les concernant, c'est qu'ils ont un plus grand besoin d'intimité et qu'ils ont donc tendance à percevoir les collectes de données comme envahissantes. Pour Berscheid (1977), le besoin d'intimité peut être évalué à travers une échelle d'introversion/extraversion (comme celle développée par Eysenck[69]) et les différences de niveau pourraient être liées à des facteurs génétiques et d'apprentissage. L'ensemble de ces hypothèses semblent confirmées par les études menées à ce sujet. Ainsi, Marschall (1974) trouve une relation entre les orientations de vie privée et une mesure d'introversion/extraversion. Stone (1986) constate, quant à lui, une relation positive entre l'introversion et les perceptions d'invasion de vie privée.

Bien que la logique et la théorie laissent supposer l'existence d'un lien entre la capacité à faire confiance, la tendance à l'individualisme et les réactions des individus eu égard aux collectes d'informations personnelles, l'étude menée par Hann et al. (2002) montre que les compromis coûts/bénéfices envisagés par les consommateurs lors de la sollicitation de leurs données ne varient pas selon ces caractéristiques personnelles. D'autres variables d'ordre psychologique comme la tendance à la méfiance, à la paranoïa et à la critique sociale auraient quant à elles un lien positif avec le niveau de préoccupation envers la protection des données. Ainsi, Smith

[69] Citée par Stone et Stone (1990)

et al. (1996) trouvent que les individus critiques, ainsi que ceux ayant une tendance à la méfiance ou à la paranoïa, seraient plus préoccupés que leurs opposés respectifs.

2.3.4 Les variables d'ordre idéologique

En ce qui concerne les variables individuelles d'ordre idéologique, leur influence sur la préoccupation envers la protection des informations personnelles et sur la réaction en cas de sollicitation de données est beaucoup moins abordée au sein de la littérature. Sont surtout étudiés les facteurs tels que la valeur accordée à l'intimité et l'adhérence aux normes, pour lesquels nous détaillons ci-dessous les principaux résultats trouvés.

La littérature dans le domaine du respect de la vie privée souligne que si le besoin d'intimité évolue d'une situation à l'autre - il dépendrait de facteurs tels que l'apprentissage, le milieu culturel et la réactivité psychologique de la personne - la valeur qui lui est accordée est par contre relativement stable au cours du temps. L'effet de cette dernière sur la réaction du consommateur en cas de sollicitation de ses informations personnelles a été étudié par Stone et al. (1983) et Stone (1986). Ces auteurs trouvent ainsi une relation entre la valeur accordée à l'intimité et l'invasion perçue au moment de la collecte des données. Ils montrent que plus un individu accorde de la valeur au fait de pouvoir contrôler ses données, moins il pense avoir de contrôle en cas de sollicitation effective, moins il est effectivement satisfait du niveau de contrôle perçu et moins il est enclin à dévoiler des informations personnelles.

En ce qui concerne l'adhérence aux normes et droits en vigueur, Milne et al. (1996) incluent cette variable dans leur modèle en expliquant que l'attitude d'un individu en matière de marketing direct et de protection de ses informations personnelles serait différente selon qu'il adhère (ou non) à ces normes. Cependant, l'effet de ce facteur n'est pas réellement testé, ce qui rend impossible de tirer des conclusions dans ce domaine.

Pour conclure, l'étude de l'influence des facteurs individuels sur la préoccupation des consommateurs eu égard à la protection de leurs données personnelles et sur leurs réactions en cas de sollicitation laisse apparaître un grand nombre de résultats contradictoires, empêchant de conclure de manière certaine à un effet - direct - de ceux-ci sur les variables dépendantes étudiées. Il n'en demeure pas moins vrai que tous les individus n'ont pas la même perception du phénomène. Plusieurs remarques peuvent donc être formulées à cet égard.

Tout d'abord, on pourrait envisager d'étudier l'impact des facteurs individuels en regroupant les individus sur la base de leur attitude générale par rapport au problème de la collecte et de l'utilisation des données personnelles, plutôt que d'étudier l'impact direct des caractéristiques individuelles telles que les données sociodémographiques, psychologiques ou idéologiques. C'est ce que suggère Salerno (2001) lorsqu'il explique qu'un des antécédents du degré d'inquiétude en cas de sollicitation de données concerne le type de consommateur. Plusieurs chercheurs ont élaboré, à cet égard, une typologie des individus, dont la plus célèbre est celle proposée par Westin (1996) qui identifie 3 principaux groupes (les fondamentalistes, les non concernés et les pragmatiques)[70].

Ensuite, il est possible, comme cela a déjà été suggéré dans les paragraphes précédents, que l'influence de ces facteurs soit indirecte, par le biais d'un effet médiateur et/ou modérateur. Cela concernerait notamment l'effet de ces variables sur les réactions attitudinales et comportementales de l'individu en cas de sollicitation de ses données personnelles, qui serait plutôt indirect. Dans ce cas en effet, les caractéristiques individuelles ne seraient pas directement liées au comportement de réponse mais tendraient à produire une pré-disposition psychologique, qui affecterait la décision (Lancelot-Miltgen 2003).

Enfin, il se pourrait que les différences individuelles soient « captées » par la divergence de perception des individus en situation de sollicitation, ce qui expliquerait l'absence ou la faible influence de ces variables. C'est la raison pour laquelle, dans ce domaine comme dans beaucoup d'autres, plusieurs chercheurs étudient l'impact des facteurs individuels, soit en évaluant *a posteriori* leur niveau de corrélation avec les variables dépendantes étudiées (Campbell 1997), soit en contrôlant leur effet sur ces dernières (Milne et Gordon 1993).

Quelle que soit la manière d'étudier leur impact, il semble que, dans de nombreux champs du marketing et dans d'autres disciplines aussi, les variables individuelles ne parviennent pas, à elles seules, à expliquer toute la variation observée au niveau du comportement. Ward et Robertson (1973)[71] signalent ainsi que les variables situationnelles pourraient expliquer une part plus large de la variance que les variables liées à l'acteur. Nous étudions, dans la section suivante, l'impact de ces facteurs situationnels (2.4). Nous commençons par identifier l'importance de la situation en général (2.4.1), avant d'aborder les différentes catégories de variables situationnelles identifiées dans la littérature.

[70] Nous reviendrons sur cette typologie dans le chapitre 6.
[71] Cités par Belk (1975)

2.4 Les antécédents situationnels

2.4.1 L'importance de la situation en général

L'impact de la situation est au cœur des études modernes du comportement du consommateur. Ainsi, les recherches sur le processus de décision, qu'elles concernent le choix d'un produit, d'une marque ou d'un circuit de distribution montrent la forte dépendance au contexte de ce processus de choix (Huber, Payne et Puto 1982, Schkade et Kleinmuntz 1994). Quel que soit le type de comportement étudié, il ne faut pas confondre l'objet vis-à-vis duquel la personne se comporte et la situation dans laquelle elle se trouve lorsqu'on l'étudie. Parce que le comportement vis-à-vis d'un objet est ce que l'on analyse en priorité lorsque l'on étudie le comportement du consommateur, l'objet est considéré comme une source possible d'influence, distincte de la situation elle-même (Belk 1975). De même, la personne est également distincte de la situation puisque celle-ci existe en dehors de la conscience de l'individu. Par individu, on entend toutes les caractéristiques stables et personnelles qui lui sont liées. Quand la caractéristique est plus transitoire (comme le fait d'être de mauvaise humeur, d'avoir mal à la tête, …), on la considère généralement comme faisant partie de la situation. De même pour l'objet, on lui attribue toutes les caractéristiques générales et stables qui lui sont liées, tandis que les caractéristiques spécifiques au lieu, à l'endroit ou au moment sont attribuées à la situation.

Dans le cas présent, nous nous intéressons au comportement de réponse du consommateur face à une sollicitation de ses données personnelles. Nous étudions donc le choix de l'individu vis-à-vis d'un objet particulier : le formulaire à remplir. Les caractéristiques de la sollicitation y compris celles concernant le formulaire à compléter (informations demandées, longueur du formulaire, ..) seront considérées comme faisant partie de la situation, dans la mesure où elles sont, à chaque fois, spécifiques au cas étudié. Ainsi, par exemple, un formulaire à compléter dans le cadre d'une commande est généralement différent de celui qui est proposé pour participer à un jeu-concours. De même, la politique de respect de la vie privée mise en place par l'entreprise sollicitant les informations, bien que relativement stable et générale, sera également considérée comme une donnée situationnelle. Elle est en effet liée à l'organisme qui demande les données, qui fait lui-même partie de la situation. De plus, elle est censée être mise à la connaissance des consommateurs chaque fois que ces derniers se trouvent soumis à une collecte de données, c'est-à-dire sur toute page comportant un formulaire à remplir. Au final, nous appellerons donc – parfois abusivement si l'on s'en tient à la stricte définition de

ce terme - caractéristique situationnelle, tout élément qui n'est ni environnemental ni individuel, qu'il soit ou non transitoire.

Nous considérons, à l'instar de Belk (1975), qu'une situation peut être définie comme un point dans l'espace et dans le temps. Celle-ci représente alors une rencontre momentanée avec certains éléments de l'environnement, à un moment particulier. La situation que nous analyserons ici est celle qui conduit un consommateur à se voir sollicité par une entreprise (ou plutôt un site Internet) pour fournir des données personnelles, par le biais d'un formulaire à remplir. Nous nous intéressons donc à la sollicitation et, plus particulièrement, à tous les éléments qui sont susceptibles de la caractériser. Nous considérons alors ces éléments comme des variables situationnelles, définies comme « un ensemble de facteurs liés à un moment et un endroit donné qui [..] exerce une influence manifeste sur le comportement » (Belk 1974). Contrairement aux caractéristiques individuelles et environnementales, qui influencent surtout la préoccupation pour le RVP et, par ricochet, le comportement de réponse, les variables situationnelles auraient donc un effet direct sur la réaction de l'individu. Nous présentons, dans les paragraphes suivants, les différentes caractéristiques situationnelles liées à une collecte de données, qu'elle se fasse ou non sur Internet, le média pouvant être considéré comme un des éléments de la situation. Nous analysons ainsi leur impact possible sur les réactions attitudinales et comportementales des consommateurs, en cas de sollicitation.

De nombreux chercheurs fournissent une liste des variables liées à la situation qui sont susceptibles, selon eux, d'influencer la réaction de l'individu en cas de sollicitation de ses données personnelles. Toutefois, certains d'entre eux - comme Salerno (2001) - n'étudient pas, de manière empirique, l'impact réel de ces variables, se contentant de les énumérer. De plus, ces listes sont souvent incomplètes puisqu'elles se focalisent généralement sur quelques variables. Dans l'ensemble cependant, les recherches menées dans le domaine de la protection de la vie privée indiquent que les consommateurs sont moins enclins à percevoir les pratiques des entreprises comme menaçant leur intimité quand les conditions suivantes sont respectées : 1) les informations sont collectées dans le cadre d'une relation existante (familiarité avec l'entreprise) ; 2) les informations collectées sont utiles aux besoins de la transaction en cours ; 3) les individus ont la possibilité de contrôler l'utilisation future des informations ; 4) ils pensent que les informations seront utilisées pour faire des inférences justes à leur propos (Bies 1993, Stone et Stone 1990, Tolchinsky et al. 1981, Woodman et al. 1982).

Comme il n'existe pas, au sein de la littérature, de liste recensant l'ensemble des variables situationnelles, nous avons décidé d'établir, sur la base des études déjà menées sur ce sujet, une typologie des principaux facteurs situationnels influents. Nous nous sommes aidés pour cela de la classification établie par Belk (1975) qui distingue 5 caractéristiques situationnelles majeures : l'« environnement » physique, l'environnement social, la perspective temporelle, le contexte du rôle à accomplir et les « états antécédents du consommateur » (comme l'humeur). L'environnement « physique » correspond, dans notre cas, à tout ce qui a trait au dispositif de collecte tandis que l'environnement social est représenté par l'entreprise à l'origine de la collecte. Par ailleurs, la perspective temporelle, le contexte du rôle à accomplir et l'état du consommateur, peu développés dans la littérature, seront regroupés au sein d'une seule et même catégorie, sous le nom de « circonstances de collecte ».

La typologie à laquelle nous avons aboutie et que nous présentons à la figure 4.3 nous conduit ainsi à distinguer 4 principaux types de facteurs situationnels, liés :

1. à la politique de RVP[72] mise en place par la firme collectant les données ;
2. au demandeur c'est-à-dire à l'entreprise qui sollicite les données ;
3. au dispositif de collecte ;
4. aux « circonstances de collecte ».

Nous avons ainsi séparé la politique de respect de la vie privée, de l'entreprise qui l'a mise en place d'une part, et du reste du dispositif de collecte, d'autre part. Ce choix s'explique tout d'abord par la différence de niveau d'analyse entre ces trois types de facteurs : mesures générales de protection des données (1), relation passée et présente avec l'entreprise (2), caractéristiques de l'instrument à travers lequel la collecte s'opère (3). Il s'explique aussi par l'importance accordée aux variables de respect de la vie privée au sein de la littérature, qui justifie, à nos yeux, de les séparer des deux autres catégories. Nous revenons dans les paragraphes suivants sur les spécificités liées à chacun des facteurs identifiés.

Nous abordons maintenant, tour à tour, et dans cet ordre, l'influence de chacun de ces facteurs, en insistant sur les résultats les plus significatifs dans ces différents domaines. Nous débutons par les facteurs liés au respect de la vie privée (2.4.2), avant d'étudier ceux liés au demandeur (2.4.3). Nous poursuivons par les facteurs liés au dispositif de collecte (2.4.4) et par ceux liés aux « circonstances de collecte » (2.4.5).

[72] Pour rappel, nous utilisons le sigle RVP pour désigner la notion de respect de la vie privée.

Figure 4.3 Typologie des facteurs situationnels influents

2.4.2 Les facteurs liés à la politique de Respect de la Vie Privée

Parmi les facteurs situationnels liés à la politique de respect de la vie privée[73] mise en place par l'entreprise sollicitant les données, nous distinguons la mise en œuvre des principes éthiques et/ou réglementaires[74] en vigueur (notamment à travers des chartes ou des mentions de RVP) et le recours à des labels, visant à garantir la confidentialité des données collectées, voire (pour certains d'entre eux) un niveau de sécurité suffisant[75]. Nous précisons l'influence de ces deux formes de politique de respect de la vie privée, qui s'avèrent d'ailleurs, dans les faits, souvent complémentaires l'une de l'autre.

Les principes éthiques et/ou réglementaires

Nous ne reviendrons pas ici sur l'explication de ces différents principes qui ont déjà été exposés en détail au chapitre 2. Nous nous attachons plutôt à identifier leur impact sur la réaction des consommateurs face à une sollicitation de leurs données personnelles (notamment sur Internet), tel que la littérature le laisse supposer. Nous précisons toutefois que la mise en œuvre de ces principes peut être portée à la connaissance des consommateurs par deux biais : 1) au moyen d'une charte (en anglais, *privacy notice*) que l'on peut définir comme « une description compréhensive des pratiques de l'entreprise (du site) en matière de collecte et d'utilisation des données consommateurs, placée à un endroit du site et accessible à l'aide d'un clic ou d'un hyperlien » ; 2) par des mentions RVP (en anglais, *information practice statement*) qui correspondent à « toute phrase décrivant une pratique particulière en la matière et pouvant apparaître à n'importe quel endroit du site ».
Nous évoquerons, tout d'abord, l'effet des différents principes en général, avant d'étudier, de manière plus succincte, l'impact des chartes.

Parmi les principes les plus étudiés dans la littérature (et les plus attendus par les consommateurs), on trouve principalement l'information (*notice*) et le choix (*choice*). Le fait, pour une entreprise de respecter ces deux règles, que ce soit par obligation (principe réglementaire) ou par choix (principe éthique), aurait ainsi une influence non négligeable sur la réaction des consommateurs qui se trouvent confrontés à une sollicitation de leurs données.

[73] Nous reviendrons en détail sur les modalités de mise en œuvre d'une telle politique dans le chapitre 10
[74] Nous avons expliqué dans le chapitre 2, la distinction qui s'opérait entre les Etats-Unis qui prônent l'auto-régulation (principes éthiques) et l'Europe qui préconise une vision étatique (principes réglementaires).
[75] La présence de labels et les garanties en matière de sécurité seront ainsi étudiées conjointement.

Plus précisément, il apparaît que le fait d'informer le consommateur de la collecte de ses informations et de la manière dont celles-ci seront utilisées par la suite, contribuerait à le rassurer et pourrait ainsi l'encourager à fournir les données demandées (Kehoe, Pitkow et Morton 1997). Ce résultat est d'ailleurs confirmé par les recherches de Culnan et Armstrong (1999), Andrade, Kaltcheva et Weitz (2002) et Dinev et Hart (2003). A l'opposé, d'autres travaux montrent que lorsque le consommateur n'est au courant ni de la collecte (Culnan 1995) ni de l'utilisation qui peut être faite de ses données (Nowak et Phelps 1992), il se révèle davantage préoccupé. Wang et Petrison (1993) indiquent, pour leur part, que le manque de connaissance du consommateur quant à l'utilisation de ses données a causé de nombreuses objections. Le fait, par exemple, d'utiliser les informations à d'autres fins que celles pour lesquelles elles ont été collectées à l'origine (notamment pour définir des profils de comportement) est souvent vu comme une invasion de la vie privée et comme une utilisation illégitime des données de la part de l'entreprise (Cespedes et Smith 1993). Enfin, Miyazaki et Fernandez (2000) étudient l'impact de l'information donnée au consommateur sur sa perception des risques et sur ses intentions d'achat. Selon eux, ce facteur n'a pas d'impact sur le risque perçu mais influence de manière significative l'intention d'achat.

Les résultats sont quasi identiques pour le second facteur, à savoir le choix laissé au consommateur de décider ce qui sera fait de ses données. Celui-ci est d'ailleurs souvent étudié en combinaison avec le précédent (information). Ainsi, Nowak et Phelps (1995) montrent que les consommateurs ne sont pas autant préoccupés par la question du RVP quand l'entreprise obtient d'eux la permission (de façon explicite ou par défaut) de collecter et d'utiliser leurs informations. A l'inverse, quand ils n'en n'ont pas été informés ou qu'ils n'ont pas donné leur autorisation préalable, leur niveau de préoccupation s'accroît (Cespedes et Smith 1993).

L'impact des chartes (vues comme une manière concrète de signifier au consommateur les principes adoptés) est plus difficile à démontrer. Si Eddy, Stone et Stone (1999) trouvent que les chartes demandant l'autorisation d'utiliser les données collectées ou qui restreignent cette utilisation à un niveau interne (par l'entreprise les ayant collectées) sont perçues comme moins envahissantes et plus justes (en anglais *fair*), il est difficile d'étendre ce résultat au niveau des comportements. Les recherches menées sur le sujet montrent en effet que la majorité des consommateurs ne lisent pas ces chartes qu'ils trouvent souvent difficiles d'accès et peu lisibles. Une étude menée par Gauzente, Benetteau et Dubreuil (2002) auprès de sites français révèle ainsi que 80% d'entre eux disposent d'une charte, souvent peu accessible. Et

ces auteurs de conclure, fort à propos d'ailleurs, qu'aujourd'hui, c'est moins la présence d'une charte que son accessibilité (et nous pourrions ajouter sa lisibilité) qui est importante.

La labellisation

La labellisation est un autre moyen utilisé par les entreprises (surtout sur Internet) pour signifier aux consommateurs les moyens mis en œuvre pour garantir la confidentialité et la sécurisation des données au cours ou à l'issue de la collecte. Benassi (1999) souligne les effets bénéfiques de cette méthode qui associe, selon lui, une image de confiance aux sites qui la mettent en place. Cranor, Reagle et Ackerman (1999) indiquent, pour leur part, qu'un programme joint de charte et de labellisation apporte un niveau de confiance identique à celui qui pourrait être apporté par la Loi[76], tout en soulignant que les internautes sont généralement peu informés sur les labels existants. Pour Killingsworth (1999), les programmes de labellisation ont un plus (par rapport à la simple apposition d'une charte), de par la crédibilité apportée par le tiers de confiance. Enfin, Miyazaki et Krishnamurthy (2002) montrent que la présence d'un label conduit, d'une part, à des perceptions plus favorables de la politique de RVP du site et, d'autre part, à une probabilité de dévoilement plus élevée.

Dans l'ensemble, les facteurs liés à la politique de respect de la vie privée de l'entreprise (chartes, mentions, labels) sont donc susceptibles de rassurer le consommateur, de réduire ses préoccupations et ainsi de l'encourager à répondre.
Voyons à présent l'impact des facteurs liés plus spécifiquement au demandeur et notamment à l'état de la relation qu'il entretient avec le consommateur. Nous souhaitons en effet savoir si les principes du dévoilement, étudiés au chapitre 3, s'appliquent aussi au niveau commercial.

2.4.3 Les facteurs liés au demandeur

Parmi les facteurs liés au demandeur, c'est-à-dire à l'entreprise sollicitant les données, nous distinguons ceux relatifs à sa notoriété (ou réputation), à l'ancienneté de la relation avec le consommateur sollicité (notion de familiarité[77]) et à la qualité de celle-ci. Nous insisterons plus volontiers sur les deux premiers, le dernier étant peu abordé dans la littérature.

[76] Cette étude est menée aux Etats-Unis où le RVP est essentiellement assuré par une politique d'auto-régulation.
[77] Nous ne précisons pas ici ce que nous entendons sous ce terme mais le ferons dans les chapitres suivants.

Nous précisons dès à présent que, si de nombreux auteurs soulignent l'impact potentiel de cette catégorie de facteurs, peu d'études empiriques viennent confirmer cela dans les faits. S'il est possible de supposer leur influence sur la réponse du consommateur face à une sollicitation de ses données personnelles, notamment en rapprochant ce cas de la manière dont le consommateur réagit dans le domaine interpersonnel[78], force est de constater que les preuves d'un tel effet sont parfois minces. Les recherches visant à étudier cette influence seraient donc particulièrement bienvenues. A partir des quelques études déjà menées sur le sujet, il est néanmoins possible de dresser un premier bilan, que nous présentons ci-dessous.

Plusieurs chercheurs, dont Cespedes et Smith (1993), indiquent qu'en matière de collecte de données personnelles, les consommateurs ont des seuils de tolérance basés (entre autres) sur l'entité (sous-entendu l'entreprise) qui sollicite les informations, sans préciser toutefois sur quels critères ils se fondent. D'autres, tels Hine et Eve (1998), précisent que la perception d'invasion de la vie privée dépend des relations établies entre l'entreprise et le consommateur. Pour Wang et Petrison (1993), l'attitude du consommateur quant à l'utilisation de ses données dépend de la réputation perçue de l'entreprise. Dans le domaine du e-commerce, il a déjà été prouvé que la réputation était positivement associée à la confiance accordée au vendeur (Jarvenpaa et tractinsky 1999). Ceci pourrait donc se confirmer dans le domaine de la collecte d'informations, d'autant qu'Andrade, Kaltcheva et Weitz (2002) montrent que la réputation tendrait à diminuer la préoccupation du consommateur quant au fait de se dévoiler.
Pour Culnan et Armstrong (1999), au-delà de la réputation, c'est surtout le fait que la sollicitation ait lieu dans le cadre d'une relation existante qui contribuerait à rassurer le consommateur. L'impact de l'ancienneté de la relation (ou de la familiarité) est d'ailleurs souligné par de nombreux auteurs (Pitkow et Kehoe 1996, Rogers 1996), sans que celui-ci n'ait été réellement démontré. Ainsi, Salerno (2001) indique que, dans ce domaine, les consommateurs distinguent les entreprises avec lesquelles ils sont familiers de celles qu'ils ne connaissent pas, sans en apporter la preuve empirique. Zhang, Wang et Chen (2001) viennent en partie combler ce manque en soulignant que le fait d'être client de la firme inciterait les individus à répondre alors que le fait de ne pas l'être constituerait un frein majeur.

En ce qui concerne l'impact de la qualité de la relation, seul l'effet de la confiance a été réellement étudié et démontré. Sweat (2000) souligne ainsi que lorsque la confiance est établie entre les parties, les questions de RVP deviennent moins préoccupantes. Ceci est

[78] Nous faisons référence ici à la littérature sur le dévoilement de soi, abordée au chapitre 3.

confirmé par Dinev et Hart (2003) qui montrent que la confiance réduit la préoccupation concernant l'utilisation ultérieure des données.

Pour conclure sur l'impact des facteurs liés au demandeur, si celui-ci semble avéré, on peut dire qu'il manque encore des études empiriques pour le confirmer.
Nous étudions à présent l'influence des facteurs liés au dispositif de collecte.

2.4.4 Les facteurs liés au dispositif de collecte

Par dispositif de collecte, nous entendons tout ce qui concerne le média de collecte, le mode de collecte (directe vs. discrète), le design de l'instrument de collecte, et les données sollicitées. Nous n'insisterons pas outre mesure sur les deux premiers, l'influence du second ayant d'ailleurs déjà été évoquée au chapitre 1. Nous rappelons en effet que nous nous situons ici dans le cas d'une sollicitation de données personnelles par le biais d'un formulaire à remplir (mode de collecte directe) en ligne (média Internet). Ces deux facteurs ne pourront donc pas exercer une influence quelconque dans le cadre de notre travail puisqu'ils sont fixés au départ. Nous analyserons, de manière plus approfondie, l'influence des deux autres facteurs, en faisant appel aux principaux résultats déjà établis dans ce domaine. Il faut, à cet égard, souligner que la majorité de ces variables ont déjà été étudiées dans le cadre des recherches concernant la réponse aux enquêtes par sondage (par le biais d'un questionnaire), ces dernières constituant une forme particulière de collecte de données clients, même si la majorité des informations recueillies ne sont pas de nature personnelle. Par contre, leur impact sur la réponse du consommateur face à une sollicitation de données personnelles, dans le cadre d'un formulaire à remplir[79] a été, jusqu'à présent, relativement peu étudié.

Le média de collecte

Peu de recherches concernant la protection de la vie privée se sont penchées sur l'analyse de l'influence du média de collecte sur la réaction de l'individu en cas de sollicitation de ses données personnelles. La majorité des travaux anciens concernent en effet la collecte d'informations par téléphone ou par courrier tandis que les plus récents se focalisent sur l'étude d'Internet, sans le comparer aux autres médias. Deux exceptions majeures soulignent

[79] Nous distinguons donc le questionnaire qui mesure l'avis et l'opinion des consommateurs sur certains points précis tout en identifiant le répondant à travers certains critères d'appartenance (âge, sexe, CSP, origine géographique) et le formulaire qui contient des informations de nature personnelle puisque systématiquement liées à un identifiant précis (nom, n° de client, …).

cependant l'intérêt de prendre en compte ce facteur. Ainsi, Cranor Reagle et Ackerman (1999) trouvent une différence de sensibilité des individus selon le média de collecte évoqué. De même, Milne et Rohm (2000) montrent que le désir du consommateur de retirer son nom des bases de données est lié au canal utilisé par les entreprises pour le solliciter. Ce désir serait plus élevé pour l'email comparativement au courrier mais moins élevé que pour le téléphone.
L'ensemble des recherches menées dans le domaine des enquêtes par sondage confirme ce rôle du média dans la motivation des individus à répondre. Ainsi, les études permettant aux répondants de choisir le canal de transmission des réponses démontrent qu'ils ont généralement un média préféré (Weible et Wallace 1998). De même, d'après Tse (1998), les individus qui ont un fort désir de confidentialité pourraient préférer répondre par courrier dans la mesure où les réponses par email sont plus facilement « traçables ».

Le mode de collecte

Concernant le mode de collecte, qu'on pourrait aussi appeler « forme de la prise de contact », si de nombreux auteurs évoquent l'impact de ce facteur, peu de recherches empiriques visent à le mesurer. Toutefois, la littérature - et la logique - laissent supposer que les collectes discrètes (par exemple, sous couvert de cookies) mettront davantage mal à l'aise les individus que les collectes directes, par le biais d'un formulaire à remplir. Il est cependant impossible d'analyser l'effet de cette variable sur le comportement de réponse lui-même puisque, dans le cas d'une collecte discrète, les données sont récupérées sans que le consommateur n'ait à agir (absence de comportement de réponse). Les chercheurs doivent donc étudier son impact sur d'autres variables dépendantes, et notamment sur les réactions cognitives et affectives des individus. A travers un design expérimental, Nantel et Elie (2002) analysent ainsi l'effet du mode de collecte sur la confiance des internautes envers le site visité. Contrairement à ce qu'ils supposaient, ils ne parviennent pas à démontrer que l'utilisation d'un formulaire engendre plus de confiance qu'une collecte par le biais de cookies. Bien qu'étonnant, ce résultat peut cependant s'expliquer par l'existence d'effets d'interactions entre les différentes variables manipulées par les auteurs et ne conduit pas à remettre en cause, outre mesure, l'effet de ce facteur.

Le design de l'instrument de collecte

Apparaissent dans cette rubrique les éléments relatifs à l'instrument de collecte lui-même tels que son format, le texte qui s'y trouve (i.e. police, taille des caractères …), les couleurs et les

images utilisées. Puisque nous étudions le cas d'Internet, nous pouvons aussi ajouter le nombre de questions par page qui concerne, selon nous, davantage le design que les données demandées. Nous n'incluons par contre ni la longueur du formulaire (que nous étudierons à travers la quantité de données sollicitées) ni le sujet abordé, qui fait davantage écho au but de la collecte et/ou au niveau de sensibilité des questions posées. L'ensemble des éléments du design semblant influencer la décision d'individus de répondre à des enquêtes, y compris en ligne, il apparaît opportun d'étudier leur effet dans le cadre de la sollicitation de données personnelles, par le biais d'un formulaire à remplir. Aucune recherche en ce sens n'a cependant été menée jusqu'à présent, ce qui offre une piste pour des travaux futurs. L'étude de l'influence de ces facteurs sur la réponse aux questionnaires – notamment dans le cadre des enquêtes en ligne - offre d'ailleurs des résultats nuancés. Ainsi, alors que Vehovar, Manfreda et Batagelj (2000) trouvent qu'ils ont un impact limité sur la décision initiale de participer - ils auraient par contre un impact non négligeable sur le taux d'abandon - Ganassali et Moscarola (2002) soulignent qu'ils influencent de manière significative le taux de réponse et surtout la nature des réponses enregistrées. Parmi les éléments testés par Vehovar, Manfreda et Batagelj (2000), on trouve notamment la présentation des questions (en une seule page ou sur plusieurs), la présence d'images, le fait que la réponse soit optionnelle ou obligatoire et la nature des instructions données (faiblement ou fortement détaillées). Ils trouvent ainsi que le taux d'abandon est plus élevé si : 1) le questionnaire contient des images au temps de chargement élevé ; 2) les réponses sont obligatoires et 3) les instructions sont trop détaillées. Ganassali et Moscarola (2002) analysent, quant à eux, la « richesse du questionnaire » à travers des éléments tels que la présence d'images et le niveau d'interactivité (questions formulées en fonction de la réponse aux questions précédentes) et étudient leur impact sur le niveau d'effort consenti par le répondant, mesuré par le nombre de questions répondues (taux de réponse) et la longueur des réponses aux questions ouvertes (richesse des réponses). Ils montrent alors comment les attraits symboliques (image et interactivité) ont un effet sensible sur la dimension créative de l'effort (réponse aux questions ouvertes). Ces résultats invitent donc à supposer l'effet possible de tels facteurs sur la réponse à la sollicitation de données personnelles, a fortiori lorsque celle-ci a lieu en ligne.

Les données sollicitées

Les facteurs liés aux données sollicitées concernent à la fois le type, le niveau de sensibilité et la quantité. Bien que proches les termes « type de données » et « niveau de sensibilité » ne doivent pas être confondus. En effet, quand le type caractérise le fait que la donnée

appartienne à l'une ou l'autre des catégories identifiées par la littérature (cf. chapitre 1), la sensibilité est liée au niveau de préoccupation qu'un individu ressent pour cette donnée dans une situation spécifique (Weible 1993). En ce sens, la sensibilité est à la fois individuelle (différente selon les individus) et situationnelle (dépendante de la situation) (Lancelot-Miltgen, 2003). Une même donnée (le numéro de téléphone) peut ainsi être vue comme privée par certaines personnes tandis que d'autres la considèrent comme publique (au prétexte qu'on la trouve dans l'annuaire). De plus, un même individu peut ne pas vouloir la dévoiler dans certains cas (par exemple à une entreprise inconnue) et la fournir facilement dans d'autres (par exemple en cas de commande, pour pouvoir être informé du moment de la livraison). Si ces deux variables (type et sensibilité) sont liées, elles n'en demeurent pas moins distinctes. Ainsi, alors que les données d'identification sont *a priori* moins sensibles que d'autres (car quasi publiques), on peut trouver des données considérées comme sensibles au sein de cet ensemble (comme le numéro de sécurité sociale, par exemple).

L'influence de ces deux facteurs s'avère particulièrement importante dans le contexte de la sollicitation de données à caractère personnel. En atteste le nombre de recherches qui citent ces critères comme des éléments clés dans la décision de l'individu de fournir ces informations (Vidmar et Flaherty 1985 ; Jones 1991 ; Gandy 1993 ; Wang et Petrison 1993 ; Weible 1993 ; Cranor, Reagle et Ackerman 1999 ; Zhang, Wang et Chen 2001, Phelps et al. 2000, Salerno 2001). Toutefois, la majorité – si ce n'est la totalité - des auteurs qui étudient l'impact des données, le font à travers le type de données demandées. Seule exception à ce principe, Milne (1997) qui, à travers un design expérimental, analyse l'effet de la sensibilité en manipulant la présence (ou l'absence) du numéro de téléphone dans la liste des informations se trouvant sur le formulaire à compléter. Il trouve d'ailleurs un résultat contraire à ce que la littérature prédit dans ce domaine, dans la mesure où l'hypothèse formulée, selon laquelle les consommateurs seront plus enclins à remplir le formulaire quand on ne leur demande pas leur numéro de téléphone, n'est pas validée. La particularité de l'échantillon interrogé (volontaires) et le caractère incitatif du mode de recrutement choisi (le fait de répondre donnait la possibilité de participer à une loterie permettant de gagner 100 dollars) expliquerait cependant, au moins en partie, ce résultat.

L'effet de la quantité de données demandées n'a été - pour l'instant - étudié que dans le contexte des enquêtes par sondage. Les résultats des recherches menées dans ce domaine sont cependant peu concluants. En effet, si comme le soulignent Kanuk et Berenson (1975), il est logique de penser que le taux de réponse sera plus faible pour un questionnaire long (vs.

court), les études ne confirment pas toujours cette hypothèse (Roscoe, Lang et Sheth 1975) et certaines vont même jusqu'à la contredire (Berdie 1973 ; Champion et Sear 1969), tandis que d'autres encore ne trouvent aucun effet (Fox, Crask et Kim 1988 ; Linsky 1975 ; Scott 1961). Il est cependant possible qu'une des explications à de tels résultats contradictoires tienne à la diversité des critères utilisés pour apprécier la longueur du questionnaire (notamment dans le cadre d'enquêtes par courrier). Ainsi, des chercheurs étudient le nombre de questions (comme Heberlein et Baumgartner 1978, Dillman 1991) alors que d'autres s'intéressent au nombre de pages (Childers et Ferrell 1979) ou encore au format des feuilles (Friedman 1975).

Comme aucune recherche dans le domaine de la sollicitation de données personnelles n'est venue corroborer l'influence de la quantité de données demandées, il semble intéressant d'envisager étudier l'impact d'un tel facteur, dans le cadre de notre travail.
Voyons à présent le dernier type de facteur situationnel, lié aux « circonstances de collecte ».

2.4.5 Les facteurs liés aux « circonstances de collecte »

Les variables liées aux « circonstances de collecte » ont été peu étudiées au sein de la littérature, ce qui ne permet pas toujours de conclure de manière formelle à leur influence. De plus, considérant le cas d'un individu à qui on demande de remplir un formulaire en ligne, on se situe dans le cadre d'une démarche active, l'individu s'étant d'abord rendu sur le site avant de cliquer sur un ou plusieurs liens l'ayant conduit sur la page contenant le formulaire. L'influence de facteurs liés à la perspective temporelle ou à l'« état » du consommateur sur le fait qu'il remplisse ou non le formulaire pourrait alors s'avérer, si ce n'est minime, moindre que si l'on s'intéressait à l'ensemble du processus de décision, depuis le choix de se rendre sur ce site. Nous ne développerons donc pas outre mesure l'influence de ces deux types de facteurs. Nous insisterons par contre plus volontiers sur l'influence des compensations offertes en échange des données, la littérature ayant souligné l'importance du rapport coûts/bénéfices lors de la prise de décision.

La perspective temporelle

Nous incluons, dans ce cadre, l'occupation du consommateur quand il se trouve sollicité pour fournir ses données personnelles (et donc le fait qu'il puisse considérer cela comme une intrusion dans une activité en cours) et le temps dont il dispose pour ce faire. L'importance du premier facteur est soulignée par Hine et Eve (1998) qui s'intéressent toutefois davantage au

caractère intrusif d'une offre de marketing direct qu'à une sollicitation de données. Nous estimons, pour notre part, que ce critère, si important lorsque la sollicitation s'opère par téléphone ou en face à face, serait négligeable dans le cadre des formulaires auto-administrés, notamment sur Internet. Le consommateur-internaute nous semble en effet rarement susceptible d'être dérangé par l'intrusion d'un formulaire puisque, sauf à considérer le cas de pop-ups[80], c'est lui qui a choisi de se rendre sur ce site et sur cette page.

L'impact du temps disponible est, quant à lui, souligné par Zhang, Wang et Chen (2001) sans avoir été pour autant démontré empiriquement. On peut toutefois penser, là encore, que le consommateur-internaute ayant choisi de se rendre sur Internet dispose d'un minimum de temps devant lui pour remplir, le cas échant, le formulaire. Ce facteur pourrait, par contre, s'avérer important si le nombre de données demandées dépasse un seuil « acceptable »[81].

L'état émotionnel de l'individu

Un autre facteur situationnel important est l'état émotionnel dans lequel se trouve l'individu au moment où il est confronté au formulaire à remplir. L'importance de ce facteur est soulignée par Childers et Skinner (1996), dans le cadre de la réponse à un questionnaire. On peut donc penser qu'il s'applique aussi au cas des formulaires en ligne.

Les compensations et/ou avantages liés à l'échange

Le dernier facteur situationnel évoqué concerne l'impact des compensations et/ou avantages proposés par l'entreprise en échange des données. Cette idée de compensation est d'ailleurs largement développée au sein de la littérature (Vidmar et Flaherty 1985, Goodwin 1991, Milne et Gordon 1993). Phelps et Nowak (2000) montrent ainsi que les bénéfices offerts par l'entreprise en échange des données fournies constituent un des antécédents aux préoccupations des consommateurs au sujet du respect de leur vie privée. Zhang, Wang et Chen (2001) indiquent, pour leur part, que le fait de recevoir « quelque chose » en échange est une raison (ou motivation) amenant les individus à accepter de répondre. Westin (1967) indique, quant à lui, que ces avantages (qui peuvent être de nature diverse, i.e. argent, produits et services, accès à des informations, …) conduisent certains consommateurs à effectuer des arbitrages tandis que d'autres y sont, soit insensibles, soit les acceptent automatiquement.

[80] Un pop-up est une petite fenêtre, en général publicitaire et non désirée, s'ouvrant automatiquement en même temps qu'une nouvelle page web.

[81] Le problème vient de la difficulté à mesurer ce que le consommateur considère comme un seuil « acceptable ».

Enfin, Andrade, Kaltcheva et Weitz (2002) soulignent que, contrairement à ce qu'on aurait pu penser, l'offre d'une récompense réduit la préoccupation du consommateur quant au fait de fournir ses données. Ces derniers résultats prouvent que, même si l'impact des compensations est avéré, celui-ci n'est pas systématique et ne va pas toujours dans le sens attendu.

Nous résumons l'influence des facteurs situationnels sur les réactions des individus face à une sollicitation de leurs données personnelles à travers le tableau suivant (cf. tableau 4.3).

Tableau 4.3 Principaux facteurs situationnels influençant la réaction à une sollicitation

Facteurs situationnels	Auteurs (année)	Résultats et commentaires
Facteurs liés à la politique de Respect de la Vie Privée		
Niveau d'information	Kehoe, Pitkow et Morton (1997) ; Culnan et Armstrong (1999), Andrade, Kaltcheva et Weitz (2002) ; Dinev et Hart (2003)	Le fait d'informer le consommateur contribue à le rassurer (en réduisant ses préoccupations) et à l'encourager à fournir les données demandées. Cela influencerait aussi ses intentions d'achat.
Niveau de choix	Cespedes et Smith (1993) ; Nowak et Phelps (1995)	Le fait d'obtenir la permission réduit les préoccupations en termes de RVP.
Présence d'une charte	Eddy, Stone et Stone (1999) ; Gauzente, Benetteau et Dubreuil (2002)	Les consommateurs y sont favorables mais peu les consultent réellement. Il est donc difficile de dire si elles ont une influence réelle sur les comportements.
Politique de labellisation	Miyazaki et Krishnamurthy (2002)	La présence d'un label conduit à des perceptions plus favorables et à une probabilité de dévoilement plus élevée. Toutefois, les consommateurs sont souvent peu informés des labels existants et du niveau de protection apporté.
Conclusion : les facteurs liés à la politique de RVP ont une influence très importante même si leur mise en œuvre est souvent peu connue et peu claire pour les consommateurs		
Facteurs liés au demandeur (entreprise sollicitant les données)		
Réputation	Wang et Petrison (1993) ; Jarvenpaa et Tractinsky (1999) ; Andrade, Kaltcheva et Weitz (2002)	La réputation est positivement associée à la confiance. Elle influence l'attitude du consommateur quant à l'utilisation de ses données et réduit les préoccupations quant au fait de se dévoiler.
Familiarité (ancienneté de la relation)	Pitkow et Kehoe (1996) ; Rogers (1996) ; Culnan et Armstrong (1999) ; Zhang, Wang et Chen (2001)	Les consommateurs distinguent les entreprises avec lesquelles ils sont familiers de celles qu'ils ne connaissent pas : le fait d'être client inciterait à répondre tandis que ne pas être client constituerait un frein majeur.
Confiance (qualité de la relation)	Sweat (2000) ; Dinev et Hart (2003)	Lorsque la confiance est établie, les questions de RVP deviennent moins préoccupantes.
Conclusion : l'influence des facteurs liés au demandeur sur le niveau des préoccupations des individus est clairement établie. Leur influence sur les réactions comportementales demande à être confirmée, et notamment celle de la familiarité (ancienneté de la relation).		

Facteurs situationnels	Auteurs (année)	Résultats et commentaires (SUITE)
Facteurs liés au dispositif de collecte		
Média de collecte	Weible et Wallace (1998) ; Cranor, Reagle et Ackerman (1999) ; Milne et Rohm (2000)	Les consommateurs ont généralement un média préféré et présentent des différences de sensibilité selon le média de collecte évoqué.
Mode de collecte	Nantel et Elie (2002)	Les auteurs ne parviennent pas à démontrer que la collecte directe (par le biais d'un formulaire) engendre plus de confiance qu'une collecte discrète (par cookies). Cela pourrait être dû à l'existence d'effets d'interaction avec les autres variables manipulées.
Design de l'instrument de collecte	Vehovar, Manfreda et Batagelj (2000) ; Ganassali et Moscarola (2002)	Dans le cadre de la réponse à une enquête, il aurait peu d'impact sur la décision initiale de participer mais en aurait sur le taux d'abandon. Il aurait aussi un effet sur le taux de réponse (nombre de questions répondues) et sur la richesse des réponses obtenues.
Sensibilité des données demandées	Milne (1997)	Le fait que l'auteur trouve un résultat contraire à ce que la littérature prédit (i.e. les consommateurs donnent plus facilement des données peu sensibles) s'expliquerait par les spécificités de l'échantillon et du mode de recrutement.
Quantité de données demandées	Kanuk et Berenson (1975) ; Heberlein et Baumgartner (1978) ; Dillman (1991)	Les résultats concernant l'enquête sont contradictoires. La logique qui voudrait que le taux de réponse soit plus faible pour un questionnaire long n'est pas toujours avérée. L'effet a été peu étudié dans le cadre d'une sollicitation de données personnelles.
Conclusion : si ces facteurs ont une influence non négligeable sur les réactions face à une sollicitation, démontrer dans quel sens s'établit leur effet est plus difficile, du fait de la multiplicité des critères à prendre en compte		
Facteurs liés aux « circonstances de collecte »		
Activité au moment de la sollicitation	Hine et Eve (1998)	Si le caractère intrusif d'une offre de marketing direct a été démontré, celui lié à une sollicitation de données personnelles n'a jamais été étudié.
Temps disponible	Zhang, Wang et Chen (2001)	Son impact a été souligné sans avoir été pour autant démontré empiriquement.
Etat émotionnel	Childers et Skinner (1996)	L'importance de ce facteur a été démontrée dans le cadre de la réponse à un questionnaire mais pas dans celui d'une sollicitation de données personnelles.
Compensations et avantages liés à l'échange	Phelps et Nowak (2000) ; Zhang, Wang et Chen (2001) ; Andrade, Kaltcheva et Weitz (2002)	Les avantages proposés en échange des données constituent un antécédent aux préoccupations. Toutefois, si leur impact sur les réactions face à une sollicitation est avéré, il n'est pas systématique et ne va pas toujours dans le sens attendu
Conclusion : l'étude de l'impact des facteurs liés aux circonstances de collecte, dans le cadre d'une sollicitation de données personnelles, est encore balbutiante. De plus, l'influence est souvent davantage soulignée que démontrée empiriquement. Enfin, l'analyse de l'effet des compensations s'avère prometteuse puisqu'il a été montré qu'il n'allait pas toujours dans le sens attendu.		

Nous venons d'évoquer l'ensemble des facteurs (i.e. environnementaux, individuels et situationnels) susceptibles d'avoir une influence sur les préoccupations et les réactions des consommateurs en cas de sollicitation de leurs données personnelles. Nous terminons ce chapitre en présentant un paradigme majeur dans les travaux portant sur la préoccupation pour le respect de la vie privée : celui du contrat social.

3 La sollicitation de données vue comme un contrat social

De nombreux auteurs américains, travaillant sur le respect de la vie privée (dont Milne et Gordon 1993, Culnan 1995), suggèrent de considérer la sollicitation de l'entreprise d'une part, et la fourniture de données par le consommateur d'autre part, comme un contrat social. Nous présenterons tout d'abord les fondements de la théorie du contrat social (3.1), avant de montrer en quoi cette théorie peut s'appliquer aux collectes de données personnelles (3.2).

3.1 La théorie du contrat social

Les tentatives pour trouver un équilibre entre les différents constituants d'une société ont été décrites comme des contrats sociaux implicites dans lesquels les individus s'engagent en échange de bénéfices économiques ou sociaux (Laufer, Proshansky et Wolfe 1976 ; Laufer et Wolfe 1977). Le contrat social serait donc une forme spécifique d'échange. Une des principales caractéristiques différenciant les types d'échange est le temps. Les échanges qui impliquent la promesse de bénéfices futurs peuvent être vus comme des contrats. MacNeil (1980) conceptualise les contrats sur un continuum allant de discret (simple transaction) à relationnel. Plusieurs attributs permettent alors de déterminer le type de contrat : le nombre de parties, le fait qu'il s'agisse d'une transaction unique ou multiple, de nature commerciale (argent contre marchandises) ou non, explicite ou implicite. Les contrats implicites, non commerciaux avec de multiples transactions sont considérés comme des contrats sociaux, au sens de MacNeil (1980).

La perspective du contrat social met l'accent sur le rôle des normes gouvernant les comportements. Ainsi, un contrat social serait initié quand on s'attend à ce que des normes sociales (ou obligations admises) gouvernent le comportement des parties impliquées. Une obligation admise signifie que les deux parties comprennent les risques et les bénéfices attendus de la transaction. Il a été montré que les normes (définies comme les attentes

partagées en termes de comportements) gouvernaient, en effet, les relations d'échange (Heide et John 1992). Ces normes diffèrent, à la fois, selon leur spécificité et le public auquel elles s'appliquent. Un aspect à prendre en compte est leur tendance à diminuer les comportements opportunistes et à imposer des réactions visant à considérer la relation comme un tout (Heide et John 1992). Si certaines sont issues de la loi, la majorité provient de croyances internes ou des conditions sociales. Aux Etats-Unis, la DMA (*Direct Marketing Association*) a ainsi publié un guide visant à établir des normes dans le domaine du marketing direct. En Europe, et en particulier en France, des initiatives similaires ont été mises en place, y compris concernant les pratiques de collecte de données sur Internet. Toutefois, les entreprises n'ont pas l'obligation de s'y conformer et tous les consommateurs ne considèrent pas ces normes comme appropriées. Les conflits interviennent donc, quand la compréhension des normes de comportement, par les parties, diffère. L'importance des normes peut aussi évoluer dans le temps (Morris 1966), ce qui peut amener à modifier, en retour, la nature des contrats. A cet égard, on peut penser que les consommateurs sont de plus en plus informés des normes implicites qui gouvernent les pratiques de traitement de leurs données personnelles. Ils sont donc plus attentifs en cas de sollicitation d'informations et leur évaluation du rapport coûts / bénéfices est également plus juste.

La théorie du contrat social fait l'hypothèse que les participants à l'échange (entreprise et consommateurs ici) sont engagés dans une relation réciproque gouvernée par un contrat social. Celui-ci est fondé sur des normes qui gouvernent le comportement des parties à l'échange. De plus, pour que le contrat soit réussi, ces normes doivent être respectées[82]. L'idée centrale du contrat social est donc d'assurer la loyauté des procédures, lors de la définition des termes du contrat. Les moyens d'assurer cette loyauté sont de faire en sorte que toute personne dont les intérêts sont susceptibles d'être affectés par le contrat y soit associée et d'exiger un consensus lors de son adoption. Il existe donc 3 éléments communs à l'établissement d'un contrat social : le consentement de l'individu, un accord entre agents moraux et un moyen par lequel cet accord est obtenu (Dunfee, Smith et Ross 1999). De plus, les termes de ce contrat sont basés sur une hypothèse forte : les contractants sont supposés être informés et soucieux d'appliquer les règles de la rationalité morale.

Selon certains auteurs, ce type de contrat pourrait s'appliquer aux situations de sollicitations de données personnelles auprès des consommateurs, que celles-ci se fassent, ou non, en ligne.

[82] Dans le cas d'une sollicitation de données personnelles, cela peut être un engagement de l'entreprise à utiliser les données que le consommateur aura accepté de fournir de manière « raisonnable ».

3.2 La théorie du contrat social appliquée à la sollicitation de données

Plusieurs chercheurs suggèrent de considérer l'échange de données entre l'entreprise et le consommateur comme un contrat social (Milne et Gordon 1993, Culnan, 1995, Milne 1997). Plus récemment, Dunfee, Smith et Ross (1999) ont également proposé de situer la protection de la vie privée en termes de contrat social. Ils encouragent en effet à se référer à cette théorie qui fournit une fondation morale au marketing, de par sa « correspondance avec la notion d'échange relationnel, centrale dans la pensée et les pratiques marketing ». Pour l'ensemble de ces auteurs, dès lors que le consommateur accepte de fournir des informations, il constituerait un contrat social implicite avec l'entreprise collectant les données. Ce contrat est implicite, car l'utilisation qui sera faite des données n'est pas toujours clairement établie. Il est non commercial car il n'y a généralement pas de compensation monétaire en échange des données. De plus, il est établi à long terme (plusieurs transactions à venir) car si le consommateur accepte de fournir des informations, c'est qu'il pense que la relation va durer. Enfin, les droits des consommateurs sont considérés comme violés si l'entreprise se démarque du comportement attendu.

Les transactions entre entreprises et consommateurs impliquent deux types de contrats : 1) un contrat économique caractérisé par l'échange de produits ou services en contrepartie d'argent (et des formes légales de régulation) et 2) un contrat social caractérisé par un échange non monétaire d'informations (gouverné par des règles implicites). A travers l'échange économique, le consommateur cherche à maximiser son utilité personnelle, ce qui n'est pas possible pour le contrat social puisque les résultats ne sont pas prévisibles. De plus, les recherches menées dans le domaine de la sollicitation de données clients montrent que ces derniers sont peu enclins à échanger leurs informations contre des bénéfices économiques, comme de l'argent ou une baisse de prix. Ils sont plutôt à la recherche d'un contrat social implicite exécuté dans le cadre d'une relation coopérative, basée sur la confiance (Hoffman et Novak 1997). Ainsi, concernant les sollicitations sur Internet, la majorité des internautes disent qu'ils ne fourniront des informations aux sites que si ces derniers les informent quant à la manière dont les renseignements collectés seront utilisés par la suite.

La décision du consommateur de fournir des informations à l'entreprise est basée sur l'évaluation des différents attributs du contrat. Cette évaluation est guidée par des normes contractuelles qui proviennent des consommateurs, des entreprises mais aussi, parfois, de

l'influence de ceux-ci sur les actions des pouvoirs publics. Pour accepter d'entrer dans une relation sociale contractuelle avec une entreprise, le consommateur doit donc percevoir (sciemment ou non) que les bénéfices retirés sont supérieurs aux coûts (Culnan, 1995)[83]. En effet, les individus valorisent les bénéfices économiques et sociaux dérivant de la fourniture de leurs données, mais ne veulent pas pour autant sacrifier la protection de leur vie privée (Culnan 1995, Goodwin 1991).

Pour Milne et Gordon (1993), quatre critères sont particulièrement importants dans la décision du consommateur de participer à un contrat social : 1) le ciblage (offre qui soit intéressante pour eux) ; 2) le volume (nombre de données collectées) ; 3) la permission (liée à la manière dont l'information collectée sera utilisée) ; 4) la compensation (ce qu'ils peuvent espérer en échange). Les consommateurs peuvent aussi considérer des questions non liées à la vie privée dans leur analyse des coûts et bénéfices. Fournier (1999) distingue ainsi la sensibilité du consommateur relative à la protection de la vie privée d'une part, et celle liée aux problèmes d'environnement, d'autre part.

Pour les spécialistes de cette théorie, un contrat social est donc créé quand le consommateur fournit des informations à une entreprise, afin qu'elle le serve mieux. Selon Culnan (1995), un contrat social est « juste »[84] quand 3 conditions sont respectées : 1) le consommateur doit être informé de la collecte ; 2) il doit savoir que ses données peuvent être partagées avec des tiers ; 3) il doit avoir l'opportunité de le refuser, d'en limiter l'usage ou de retirer son nom du fichier. Le contrat est donc considéré comme nul si le consommateur n'est pas informé de la collecte d'informations le concernant, si l'entreprise utilise ces informations sans avoir obtenu son accord au préalable ou si le consommateur n'a pas la possibilité de refuser l'utilisation de ses données. Les deux hypothèses sous-jacentes sont les suivantes : 1) la plupart des individus souhaite avoir plus de contrôle sur l'utilisation de leurs données ; 2) le fait d'en avoir, réduira leurs préoccupations dans ce domaine et les incitera alors à se montrer plus coopératifs.

Le choix, pour une firme, d'adopter une politique de respect de la vie privée et de la publier[85] est donc un moyen d'établir un contrat social loyal puisque, ce faisant, elle signale au consommateur qu'il peut avoir confiance (Spence 1974). Ceci serait même particulièrement important quand ce dernier n'a pas d'expérience préalable avec l'entreprise le sollicitant.

---

[83] On retrouve ici les fondements issus de la théorie de l'échange social, présentés brièvement au chapitre 3.

[84] Le terme anglais « *fair* » est plus explicite que tous les termes français pouvant être utilisés pour le traduire (i.e. juste, moral, loyal, équitable, etc, …)

[85] Aux Etats-Unis, cette pratique consiste à adopter les *Fair Information Principles* préconisés par le FTC et présentés au chapitre 2.

Pour conclure, la théorie du contrat social offre donc un moyen de mieux comprendre les tensions existant entre les entreprises et les consommateurs à propos du RVP[86]. Cette théorie suppose que les membres d'une communauté se comportent de manière juste et que leurs pratiques soient gouvernées par des normes partagées par tous. Elle est donc particulièrement appropriée pour comprendre les pratiques des entreprises en matière de traitement de données clients. Dans ce cadre, le respect de la vie privée du consommateur impliquerait donc la mise en œuvre d'un contrat social implicite à travers lequel l'individu échangerait une partie de son intimité en échange de bénéfices économiques ou sociaux. De plus, Milne et Gordon (1993) soulignent que ce contrat social pourrait être vu une série de compromis entre les coûts et bénéfices résultant de l'engagement dans une relation avec l'entreprise.

Un échange impliquant des données personnelles est de ceux qui concernent une relation dont les obligations futures ne sont pas spécifiées à l'avance et qui est basée sur la confiance existant entre les parties (Blau 1964). Il est donc différent des échanges purement économiques dans la mesure où il basé sur des liens sociaux et sur des échanges informels de « biens intangibles » tels que des idées et du pouvoir. Dès lors, l'impression du consommateur concernant la sollicitation devient critique pour déterminer si l'échange aura (ou non) lieu. Dans un contexte commercial, ce sentiment pourrait être capturé par l'existence de facteurs menant à la confiance.

[86] Nous avons souligné, dans le chapitre 1, les conflits d'intérêt existants entre les entreprises et les consommateurs concernant le respect de la vie privée et la propriété des données clients.

Conclusion du chapitre 4

La réaction du consommateur face à une sollicitation de ses données personnelles est à la fois compliquée à appréhender et difficile à mesurer. Elle fait en effet appel à des mécanismes qui interagissent conjointement et les facteurs susceptibles de l'influencer sont nombreux.

Jusqu'à présent, les chercheurs ont majoritairement étudié l'influence de la sollicitation au niveau attitudinal. Ils analysent alors l'impact de celle-ci sur la préoccupation des consommateurs quant à la protection de leurs données personnelles. Relativement peu d'études se sont intéressées au comportement de réponse, alors même qu'il s'agit d'une variable majeure. De plus, l'écart pouvant exister entre les convictions personnelles des individus et leur manière de réagir concrètement quand ils sont sollicités renforce la nécessité de prendre en compte tous les aspects du phénomène.

Comme précisé ci-dessus, les facteurs susceptibles d'influencer les réactions attitudinales et comportementales des individus face à une sollicitation de données personnelles sont variés. De façon classique, nous distinguons trois catégories d'antécédents, de nature environnementale, individuelle ou situationnelle. Dans notre recherche, l'impact des facteurs environnementaux sera neutralisé puisque nous travaillons au sein d'un environnement identique pour tous les individus. L'impact des facteurs individuels est sujet à caution et les recherches passées ont du mal à prouver de façon irréfutable l'influence de certains d'entre eux. Nous focaliserons donc notre attention sur l'analyse des facteurs situationnels, dont l'impact est à la fois peu étudié et substantiel.

Le paradigme du contrat social est largement mobilisé dans les recherches portant sur la réponse des individus face à une sollicitation de leurs données personnelles. Il considère en effet l'échange de données entre le consommateur et l'entreprise comme un contrat fondé sur le respect de normes implicites entre les parties. La sollicitation de données, vue comme un contrat social, permet alors de mieux prendre en compte les intérêts de chacun et donc de résoudre, en partie, les conflits existants, tels que ceux évoqués au chapitre 1.

CONCLUSION DE LA PARTIE I

Cette première partie, qui se présente essentiellement comme une revue de la littérature, a permis de répondre aux interrogations soulevées par l'étude des réactions des consommateurs face à une sollicitation de leurs données personnelles.

Dans le premier chapitre, nous avons insisté sur les différents enjeux liés à la collecte de données clients (en particulier d'ordre stratégique et éthique) et avons montré les conflits qui pouvaient exister entre les besoins en informations des entreprises et les attentes des consommateurs en matière de respect de la vie privée.

Le second chapitre a été l'occasion de mieux définir cette notion de vie privée et de montrer de quelle manière elle pouvait être affectée par la collecte de données à caractère personnel.

Dans le chapitre 3, nous avons présenté les attributs majeurs du dévoilement de soi (nombre d'informations dévoilées et degré d'intimité de celles-ci) et avons montré en quoi le fait de se dévoiler était un processus conduit par la motivation. Nous avons ensuite listé l'ensemble des facteurs susceptibles d'influencer un tel processus, en insistant plus particulièrement sur l'importance de la cible et de la relation établie avec cette dernière.

Enfin, le quatrième chapitre a permis de faire le point sur les principaux résultats établis dans la littérature concernant les réactions des consommateurs confrontés à une sollicitation de leurs données personnelles. Nous avons en particulier mis l'accent sur l'impact respectif des facteurs environnementaux, individuels et situationnels.

PARTIE II

EXPLORER ET BATIR

Sollicitation de données personnelles et réactions des consommateurs

Des études exploratoires à la méthodologie de la recherche

INTRODUCTION DE LA PARTIE II

Après avoir présenté la revue de la littérature dans la partie I, cette deuxième partie (intitulée « EXPLORER ET BATIR ») vise à élaborer le cadre conceptuel général de cette recherche permettant d'évaluer les réponses attitudinales et comportementales des consommateurs lorsqu'ils sont confrontés à une sollicitation de leurs données personnelles.

Nous commencerons par investiguer, de manière qualitative (chapitre 5) puis descriptive (chapitre 6), la façon dont les consommateurs perçoivent ce type de sollicitation d'une part, et y répondent, d'autre part. Les résultats de ces deux études exploratoires, auxquels s'ajoutent les conclusions de la littérature (présentées en partie I), nous permettront d'établir le cadre conceptuel.

Après avoir exposé l'ancrage théorique de cette recherche, nous présenterons le modèle conceptuel général auquel nous avons abouti (chapitre 7) qui donnera lieu à la formulation d'un certain nombre de propositions et hypothèses de recherche que nous serons amenés à tester dans la suite de ce travail doctoral.

Nous terminons cette deuxième partie par la méthodologie de la recherche (chapitre 8) et le développement des instruments de mesure correspondants aux différentes variables du modèle conceptuel (chapitre 9). Nous exposerons ainsi les raisons ayant conduit à adopter un design expérimental, avant d'aborder les questions de validité. Nous poursuivrons par la présentation des qualités psychométriques des échelles de mesure développées et/ou reprises pour les besoins de la recherche.

Partie II – EXPLORER ET BATIR

Sollicitation de données personnelles
et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

CHAPITRE 5
PROCESSUS DE REPONSE FACE A UNE SOLLICITATION DE DONNÉES PERSONNELLES : EXPLORATION QUALITATIVE (ÉTUDE 1)

« Il y a un manque de réglementation, surtout dans le cadre d'Internet.
Les pouvoirs publics devraient se pencher sur la question. Pour l'instant, c'est le désert ».

Un des participants à l'étude qualitative

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

Chapitre 1 : Les enjeux liés à la collecte de données clients

Chapitre 2 : Préambule sur la notion de vie privée

Chapitre 3 : Dévoilement de soi et divulgation de données personnelles

Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)

Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)

Chapitre 7 : Formulation du cadre conceptuel général de la recherche

Chapitre 8 : Méthodologie de la recherche

Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)

Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)

Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 5

L'objet du présent chapitre est de présenter les résultats d'une étude exploratoire qualitative menée auprès d'internautes français. Les justifications de l'intérêt d'une telle étude sont multiples. Nous les indiquons brièvement ci-dessous.

Malgré les investigations diverses dont la préoccupation pour la protection des données à caractère personnel a fait l'objet[87], on ne sait toujours pas vraiment de quelle manière les individus conçoivent le respect de la vie privée dans leur vie quotidienne ou dans leurs rapports avec les entreprises, ni quelle est l'origine exacte de leurs réactions face aux pratiques de collecte et d'utilisation de leurs données personnelles.
De nombreux auteurs travaillant sur le thème de la vie privée, et plus particulièrement sur la vie privée informationnelle (contrôle sur la collecte et la diffusion des informations personnelles) soulignent le manque d'études qualitatives offrant une meilleure compréhension du phénomène. Il semblait donc utile d'étudier plus en profondeur les perceptions des consommateurs à ce sujet, de manière à comprendre comment les préoccupations se construisent dans l'inconscient des individus et comment elles se répercutent quand, confrontés à une sollicitation de données personnelles, ils doivent décider de fournir ou non les renseignements demandés.
De plus, l'absence d'études françaises sur ce thème renforce la nécessité d'étudier les mécanismes profonds sous-jacents, avant d'adopter une démarche plus quantitative. Nous avons donc choisi d'interroger plusieurs consommateurs français, afin de comprendre comment ils procédaient pour décider de délivrer ou non des informations personnelles.

La méthodologie et les résultats de cette étude qualitative exploratoire sont présentés ci-après. Nous commençons par expliciter les objectifs de l'étude ainsi que les méthodes utilisées pour collecter puis analyser les données obtenues (1). Une fois les résultats décrits et analysés (2), nous les mettons en perspective (3) à travers, notamment, la présentation d'un modèle conceptuel provisoire (cf. figure 5.1).

[87] Celles-ci ont été présentées notamment dans le cadre de la première partie de ce travail doctoral qui consistait à faire le point sur les avancées de la littérature dans ce domaine.

1 Méthodologie de l'étude qualitative

1.1 Les objectifs de l'étude

Cette étude exploratoire a plusieurs objectifs.

Nous cherchons tout d'abord à analyser la construction faite par les consommateurs des concepts de donnée personnelle (DP) et de vie privée (VP). Il s'agit ainsi de dégager les croyances saillantes des individus quant aux pratiques de traitement de données. En cela, nous souhaitons entrevoir l'opinion d'internautes français concernant la sollicitation de leurs données personnelles en général, et sur Internet en particulier. Le recours à une méthodologie qualitative offre, à cet égard, la possibilité de générer de vraies conclusions (au sens de non polluées), grâce à la prise en compte des mécanismes émotionnels rapportés dans les discours.

Le second objectif vise à mieux comprendre comment le consommateur perçoit la sollicitation de ses données et de quelle façon il y répond. La collecte peut en effet être considérée comme intrusive ou non, selon la manière dont la sollicitation est vécue, analysée et évaluée. Cette évaluation dépendrait de différents critères que nous cherchons ici à identifier. Nous voulons aussi mieux apprécier l'impact des facteurs identifiés dans la littérature (individuels et situationnels notamment) sur l'ensemble du processus de réponse.

Le troisième objectif est d'aboutir à un « premier » modèle conceptuel visant à identifier les étapes du processus de réponse face à une sollicitation de données personnelles.

Enfin, le quatrième et dernier objectif vise à dégager des verbatims susceptibles d'être utilisés ultérieurement pour la création d'échelles.

1.2 La méthode de collecte des informations

1.2.1 Le recours à des entretiens semi-directifs

Sur un sujet aussi délicat que celui de la vie privée, le choix de la méthode s'avère crucial. Comme l'a fait remarquer Edwards (1993), les gens sont souvent hésitants à discuter de leur vie privée avec d'autres personnes. Ceci nous a conduits à éliminer très rapidement la technique des réunions de groupes, puisqu'il nous a semblé que la verbalisation ne serait pas forcément facilitée en présence d'autrui. De plus, ce n'est pas un thème pour lequel il existe un phénomène d'influence du groupe. Le recours à des analyses projectives aurait pu être intéressant. Cela aurait permis de mesurer l'inconscient des individus, ce qui peut sembler approprié sur un sujet aussi « intime ». En face à face, il n'est en effet pas certain que les gens parviennent toujours à exprimer réellement le fond de leurs pensées et surtout ce qui les

amène à réagir ainsi. Toutefois, bien que le sujet soit *a priori* « sensible », il ne nous semblait pas évident qu'il générerait de la part des interviewés des mécanismes de blocage, qu'il serait dès lors nécessaire de contourner. En effet, il s'agissait d'aborder la vie privée sous son aspect informationnel et « commercial », à savoir la collecte et l'utilisation des données personnelles par les entreprises ; ceci est moins sensible que de traiter ce même thème sous l'angle interpersonnel. De plus, comme toute technique, l'analyse projective n'est pas exempte d'inconvénients dont le principal est de s'assurer que le matériel élaboré n'induit pas, par sa composition propre, d'associations systématiquement biaisées (Evrard et al. 2003). Pour ces différentes raisons, nous n'avons donc pas utilisé de tests projectifs.

Afin de répondre de manière optimale aux objectifs que nous nous étions fixés, nous avons finalement opté pour une étude qualitative fondée sur des entretiens individuels semi-directifs aussi appelés entretiens « centrés » (Merton, Fiske et Kendall 1990). Ce type d'entretien permet en effet de faire ressortir les logiques individuelles de perception et, en travaillant sur une population suffisamment large, de construire des régularités dans les témoignages, permettant ensuite d'identifier des grands types d'attitudes et/ou de comportements. Il offre aussi l'avantage d'analyser dans le détail la motivation des individus et de rendre compte de mécanismes psychologiques très fins. Un autre bénéfice de cette méthode est la validité des données produites. Cossette (1994) remarque ainsi que les données ayant été générées spontanément par le répondant, elles sont davantage susceptibles de refléter sa pensée. A l'opposé, cette méthode limite souvent la fiabilité des données.

1.2.2 Le guide d'entretien

Dans l'entretien semi-directif, le chercheur utilise un guide structuré, appelé guide d'entretien, qui vise à aborder au cours de l'interview une série de thèmes définis préalablement. Ce guide aide à structurer l'interview (à respecter un cadre le plus standardisé possible) et à aiguiller la récolte d'informations, de manière à ne pas trop se disperser et à évacuer les discours « inutiles » (hors sujet) des répondants.

Avant d'établir la version définitive de ce guide, nous avons réfléchi (puis vérifié), de manière approfondie, à la pertinence des thèmes abordés, des questions posées et des relances possibles d'une part, et à la validité globale de l'outil d'autre part. Nous n'avons ainsi retenu que les questions à la fois pertinentes et utiles et avons vérifié que celles-ci nous permettraient de répondre aux objectifs que nous nous étions fixés. Le guide d'entretien ainsi élaboré a ensuite été soumis au jugement d'un chercheur, expert en analyse qualitative, qui nous a

suggéré quelques voies d'amélioration dont nous avons tenu compte pour établir la version finale. La principale recommandation de cet expert consistait à inclure des questions amenant l'interviewé à projeter certaines de ses réponses sur d'autres personnes. Cela constituait un moyen de contourner d'éventuels mécanismes de blocages, qui empêcheraient celui-ci de se dévoiler complètement. Nous avons ainsi inclus des questions du style « Connaissez-vous des personnes qui pourraient avoir des réactions différentes des vôtres ? ». Après avoir établi la version finale du guide d'entretien comprenant les différents points à aborder ainsi que les relances à prévoir (cf. Annexe 5.1), celui-ci a ensuite été pré-testé en face à face auprès de quelques personnes, afin de vérifier que les instructions et les questions étaient claires et ne prêtaient pas à confusion. Il faut préciser que les questions principales (identifiées par un ●) ont été systématiquement posées aux participants (sauf s'ils les abordaient d'eux-mêmes à travers leurs discours, auquel cas, nous les amenions à développer leur pensée) tandis que les questions d'investigation (symbolisées par ☞) constituaient surtout des relances, visant à amener l'interviewé à exprimer au maximum son opinion sur le sujet.

Ce guide d'entretien comprend deux thèmes principaux, qui ont été abordés successivement au cours des interviews. Le premier porte sur la collecte d'informations personnelles et le second sur la vie privée. Ce dernier est volontairement abordé à la fin puisqu'il est moins important que le premier (compte-tenu de nos objectifs de recherche) et moins facile d'accès (demande plus de réflexion). Il nous a donc paru préférable de commencer par un thème qui mettrait les participants en confiance.

1.2.3 Le déroulement de l'entretien

Nos interviews ont consisté en une conversation libre (de 1h10 à 1h30 environ). Ils ont eu lieu, soit au domicile des personnes, soit à leur travail pendant la pause déjeuner. Le déroulement des entretiens a suivi une procédure mise au point préalablement et appliquée à chaque fois, de manière à ne pas créer de disparité entre les interviewés. Cette procédure est présentée en Annexe 5.2. Chaque entretien débutait par une phase de présentation au cours de laquelle nous nous sommes attachés à établir le contact avec le répondant, à le rassurer et à lui indiquer les objectifs de l'étude. Nous lui avons expliqué la présence du magnétophone et lui avons demandé l'autorisation d'enregistrer ses propos. Nous lui avons aussi posé quelques questions « filtres » pour vérifier qu'il appartenait bien à la « cible » (population d'internautes) et nous assurer que nous aboutissions à un échantillon suffisamment diversifié. Dans certains cas toutefois, cette phase de filtrage avait été réalisée au préalable, lors de la

prise de rendez-vous téléphonique, de manière à éviter un déplacement inutile. Des questions portant sur l'utilisation d'Internet et le recours aux techniques d'achat à distance étaient posées de manière à dresser un portrait « complet » du répondant en matière d'exposition aux techniques de traitement de données. Poser ces questions nous donnait, en outre, la possibilité de tester les liens entre les réponses et le profil du répondant. Au total, ce mini questionnaire d'identification (cf. Annexe 5.3) ne dépassait pas 5 minutes et offrait l'avantage de mettre l'interviewé en confiance puisqu'il s'agissait de questions fermées donc faciles à répondre.
Après lui avoir demandé s'il avait des questions à poser ou des remarques à formuler à propos du déroulement de l'entretien, nous avons signalé au répondant le début de l'enregistrement et avons posé la première question. Ensuite, au cours de l'interview proprement dite, nous nous sommes « contentés » d'introduire les différentes questions (si elles n'étaient pas abordées de façon spontanée par les répondants), de recentrer leur discours et de reformuler leurs propos afin de les relancer. Nous nous sommes particulièrement attachés à respecter les principes de base de la non-directivité (qui s'appliquent également aux entretiens semi-directifs) à savoir : « une attention positive inconditionnelle » et « une attitude d'empathie » (Evrard et al. 2003). En résumé, cela consiste essentiellement à montrer au sujet qu'on l'écoute de manière attentive, qu'on comprend ce qu'il dit et qu'on est intéressé par son discours. A la fin de l'interview, nous avons bien évidemment remercié les personnes et avons pris le temps de discuter quelques instants avec elles, notamment avec celles que l'interview avaient rendu curieuses et avides d'en connaître un peu plus sur le sujet.

1.3 Le recrutement et le profil des participants

Concernant le choix des répondants, nous nous sommes attachés à interroger des personnes aux profils contrastés, afin de couvrir une large variété d'expériences en matière de collecte d'informations personnelles sur Internet. C'est la recherche de cette diversité qui a donc guidé notre sélection. Plusieurs auteurs ayant souligné l'impact des critères sociodémographiques et de l'expérience d'internet sur la préoccupation pour le respect de la vie privée et sur la réponse à une sollicitation de données personnelles (cf. chapitre 4), ces variables nous ont semblé utiles pour composer l'échantillon. Si l'objectif initial était de respecter le principe de saturation (Yin 1989) pour déterminer le nombre souhaitable d'interviews à mener, nous devons avouer que cette règle n'a pas pu être entièrement respectée, faute de temps disponible pour mener plus d'interviews[88] d'une part et de trouver des individus acceptant d'être interviewés d'autre part. Ainsi, plusieurs personnes ayant accepté initialement de participer se

[88] Cette étude a été réalisée dans le cadre du mémoire de DEA, donc dans un temps imparti limité

sont désistées au dernier moment, le plus souvent, à cause d'un emploi du temps surchargé. Il faut toutefois signaler que, sans être strictement parvenus au critère de saturation, les derniers entretiens menés montrent déjà quelques redondances avec les entretiens initiaux. Ceci laisse donc penser que la saturation serait intervenue rapidement et qu'en stoppant l'étude à ce stade, nous ne limitons pas outre mesure la validité et la fiabilité de celle-ci.

Au final, nous avons interrogé 9 individus dont 7 parisiens et 2 provinciaux. Les participants ont des profils différents en ce qui concerne l'âge, le sexe, le niveau d'études, la situation de famille, l'expérience en matière d'achat à distance et d'utilisation d'Internet. Il s'agit donc d'un échantillon plutôt diversifié sociodémographiquement parlant (cf. tableau 5.1).

Tableau 5.1 Profil sociodémographique des répondants

N°	Nom	Jour	Durée	Sexe	Age	Profession	Niveau d'études	Situation familiale
1	Frédérique	03 / 06	1h 10	F	31	Enseignante	Bac +3/4	Concubinage 1 enfant
2	Marie	04 / 06	1h 10	F	25	Etudiante	Bac + 4	célibataire
3	Michel	05 / 06	1h 20	H	52	Directeur technique	Bac	Marié 3 enfants
4	Estair	06 / 06	1h 20	F	24	Assistante de gestion	Bac + 4	concubinage
5	Mickaël	07 / 06	1h 20	H	23	Technicien de gestion	Bac + 2	concubinage
6	Elizabeth	15 / 06	1h 15	F	26	Chargée d'études, cadre	Bac + 5	concubinage
7	Marie Madeleine	17 / 06	1h 10	F	55	Mère au foyer	< Bac	Mariée 3 enfants
8	Cédric	21 / 06	1h 20	H	28	Chargé d'études	Bac + 5	Célibataire
9	Josette	01 / 07	1h 30	F	52	Technicienne d'exploitation	< Bac	Mariée 1 enfant

Le seul point commun entre ces individus est d'être tous des internautes, à des degrés divers d'ancienneté et d'utilisation (cf. tableau 5.2). Malgré notre volonté initiale d'interroger des individus au profil diversifié en termes d'expérience Internet (à la fois des débutants et des gens confirmés), nous avons eu du mal à recruter des « novices » (nous tombions en effet soit sur des non-internautes soit sur des personnes utilisant Internet depuis plus de 2 ans), ce qui induit un léger déséquilibre au profit des individus plus expérimentés, sans que cela puisse toutefois limiter la portée de nos résultats.

Tableau 5.2 Expérience des répondants en matière d'achat à distance et d'Internet

N°	Expérience d'achat à distance (hors Internet)	Expérience sur Internet	Remarques
1	Oui plusieurs fois / an multiples produits / services	5 ans plusieurs fois / semaine achats réguliers	Achat pour le confort, le gain de temps
2	Oui plusieurs fois / an beauté / hygiène	2 ans quasiment tous les jours 1 seul achat en ligne	Crainte de dévoiler son n° CB sur Internet
3	Oui plusieurs fois / an loisirs	2 ans quasiment tous les jours pas d'achat	Peur du piratage de sa CB, y voit un problème de génération
4	Oui plusieurs fois / mois multiples produits /services	3 ans quasiment tous les jours achats réguliers	Peur de la fraude sur sa CB
5	Oui 1 à 2 commande / mois loisirs, informatique	6 ans quasiment tous les jours achats réguliers	Passionné d'Internet et des jeux en réseau
6	Oui plusieurs fois / an beauté / hygiène	5/ 6 ans quasiment tous les jours achats réguliers	Utilisation d'Internet au travail (pas de connexion au domicile)
7	Oui moins d'1 fois / an vêtements	2 ans plusieurs fois / mois pas d'achat	Préférence pour l'achat en magasin
8	Non	3 ans quasiment tous les jours achats réguliers	Passionné d'informatique et d'Internet
9	Oui plusieurs fois / an vêtements	2 ans plusieurs fois / semaine pas d'achat	Disposée à fournir des données (sauf privées), peur du piratage de sa CB

1.4 Le choix de la méthode d'analyse : l'analyse de contenu

1.4.1 Justification et principes de l'analyse de contenu

Pour l'analyse des entretiens, nous avons eu recours à une analyse de contenu, forme d'analyse du discours développée dans les années 20 aux Etats-Unis. Selon Berelson (1952) (cité dans Evrard et al. 2003), l'analyse de contenu est « une technique de recherche pour la description objective, systématique et quantitative du contenu des communications ayant pour but de les interpréter ». Nous nous sommes exclusivement attachés à réaliser une analyse thématique (au sens de Bardin 2003) où le thème, c'est-à-dire l'unité de signification est l'unité de découpage, de codage puis d'analyse. En effet, au-delà du fait qu'elle constitue la

technique la plus fréquemment utilisée en marketing (Evrard et al. 2003), elle nous paraissait particulièrement adaptée à notre cas. Enfin, même s'il est possible de pousser l'interprétation jusqu'à analyser le contenu latent du discours (c'est-à-dire les éléments « symboliques » du matériel analysé), nous avons pour notre part, et comme le recommande Muchielli (1982), porté l'essentiel de notre attention au contenu manifeste des entretiens (matériel brut, c'est-à-dire de ce qui a été ouvertement dit par chaque répondant), dans la mesure où celui-ci constitue une réalité « indiscutable ».

Notre objectif était ici double, voire triple. Il s'agissait dans un premier temps, à partir des données collectées, de décrire l'état des croyances et des représentations des internautes français en matière de respect de leur vie privée sur Internet et, plus particulièrement, de traitement de leurs données personnelles. Dans un second temps, nous souhaitions également comparer les discours des différents répondants entre eux, afin de souligner leurs éventuelles différences et ressemblances. Cela nous permettait ainsi de mettre en évidence l'importance de certains facteurs, notamment individuels et situationnels, dans la réponse donnée par les internautes français, face à une sollicitation de leurs données personnelles. Ainsi, la mise en évidence de différences, dans la définition donnée par les répondants du concept de vie privée pourrait permettre d'expliquer certaines différences de comportement. Enfin, nous souhaitions également expliquer (autant que possible) ces comportements de réponse, en identifiant les critères pris en compte lors de la prise de décision.

1.4.2 Les étapes de l'analyse

L'analyse des données qualitatives comprend 6 étapes principales - précédées d'une étape préalable de préparation des données - que nous nous sommes attachés à respecter. Nous détaillons chacune de ces étapes en Annexe 5.4, afin que le lecteur ait une idée de la démarche utilisée. Nous ne présentons ci-dessous que les choix auxquels nous avons finalement abouti.

Notre position (absence de « *tabula rasa* ») nous a portés à opter pour une catégorisation plus souple que celle traditionnellement utilisée en analyse de contenu. La lecture de la littérature nous avait en effet conduits à « imaginer » un certain nombre de catégories dès le départ. L'analyse des entretiens nous a ensuite conduits à ajouter de nouvelles catégories, à modifier voire à regrouper certaines catégories définies initialement. L'avantage de procéder de cette manière est d'allier la rigueur de l'analyse à la richesse de termes « inattendus » qui émergent de cet examen et dont l'interprétation peut offrir des éclairages intéressants (Herbert 2005).

Il faut ensuite procéder à une réduction de l'ensemble en un nombre restreint de catégories distinctives. Il convient en effet de s'assurer de la pertinence des catégories développées (cf. liste des « qualités » à respecter, présentée en Annexe 5.4).

Notre analyse nous a ainsi conduits à distinguer 30 catégories au total (cf. tableau 5.3), dont 8 non prévues initialement, représentées en grisé. Ces catégories ont ensuite été regroupées, pour les besoins de l'interprétation, en 6 grands « thèmes » parmi lesquels, on trouve les facteurs (individuels et situationnels notamment) susceptibles d'influencer l'évaluation de la sollicitation d'une part et le comportement de réponse d'autre part.

Tableau 5.3 Liste des catégories et thèmes retenus

N°	CATEGORIES	THEMES		N°
1	Informations personnelles (définition)	Perceptions et définitions des principaux concepts		I
2	Informations personnelles (types, caractéristiques)			
3	Vie privée (définition)			
4	Vie privée (menaces / atteintes)			
5	Protection (besoin)			
6	Protection (mesures prises)			
7	Spécificités d'Internet	Spécificités du média Internet		II
8	Perceptions et attitude face aux pratiques de collecte	Attitude face au traitement des données personnelles		III
9	Perceptions et attitude face à l'utilisation des données			
10	Perceptions et attitude face à la constitution et au transfert de fichiers clients			
11	Types d'informations demandées	Facteurs situationnels	Facteurs influençant le compor-tement de réponse	IV
12	Dispositif de collecte			
13	Liens avec l'entreprise sollicitante			
14	Politique / mentions RVP sur le site			
15	Contexte / particularités du contact			
16	Circonstances			
17	Réglementation	Facteurs environnementaux		
18	Culture			
19	Technologie			
20	Bouche à oreille			
21	Personnalité, valeurs et perception de soi	Facteurs individuels		
22	Expériences, passé du consommateur			
23	Caractéristiques sociodémographiques			
24	Disposition du consommateur (implication / motivation)			
25	Contrôle /confidentialité perçu(e)	Evaluation de la sollicitation		V
26	Avantages / coûts perçus			
27	Utilité / pertinence de la collecte			
28	Sensibilité perçue des données demandées			
29	Notion d'échange			
30	Comportements de réponse	Comportements de réponse		VI

Nous n'avons pas jugé utile de procéder à un traitement « statistique » des données (étape 4), pour deux raisons majeures. Tout d'abord, notre analyse était essentiellement exploratoire et descriptive, sans volonté de quantifier les thèmes apparus. Ensuite, la faible taille de l'échantillon réduisait, selon nous, la pertinence et la « scientificité » d'une telle analyse[89]. Enfin, ce type de quantification nous apparaît plus opportun et en tout cas plus facile à réaliser avec des unités de sens formés par des mots plutôt que par des phrases, le découpage (parfois aléatoire) de celles-ci rendant, selon nous, difficile de tirer quelque enseignement utile des résultats obtenus. Nous nous sommes donc contentés de comptabiliser, à certaines occasions, le nombre d'interviewés ayant abordé tel ou tel thème puisque cela nous permettait de souligner (sans objectivité absolue toutefois) l'importance de ces sujets pour les répondants. Nous avons ainsi remarqué que certains d'entre eux n'avaient pas abordé le thème de la réglementation, ce qui nous a semblé représentatif du manque d'information d'une partie du public concernant les droits que la loi est censée leur conférer dans ce domaine.

L'analyse qualitative des résultats (étape 5) est une étape cruciale. Il s'agit ici d'identifier les liens existants entre les différentes catégories et/ou le processus d'enchaînement entre elles. Dans notre cas, nous souhaitions apprécier l'importance des différents facteurs d'influence (individuels et situationnels notamment) et formaliser leurs relations avec d'autres variables comme l'évaluation de la sollicitation et le comportement de réponse.

Il faut, pour l'essentiel, recourir à des allers retours perpétuels entre la littérature et les résultats obtenus afin de noter les similitudes et les contradictions éventuelles. L'utilisation de matrices, recommandée par de nombreux chercheurs dont Miles et Huberman (1991), est utile pour chercher des relations de cause à effet entre certaines variables. Tel fut le cas dans notre étude pour l'analyse des liens entre les facteurs explicatifs (individuels et situationnels notamment) et les variables dépendantes (l'évaluation de la sollicitation et le comportement de réponse). Nous avons réalisé des matrices[90] à visée à la fois descriptive et causale où nous avons fait figurer les facteurs en ligne et les variables dépendantes en colonne. Nous avons ensuite indiqué dans chaque case les verbatims permettant de suggérer l'existence de relations ou de dégager les relations récurrentes explicatives des variables dépendantes étudiées.

---

[89] Notre étude se voulait en effet plus en « profondeur » (nombre limité de personnes interrogées longuement) qu'en « largeur » (pour laquelle le nombre d'individus interviewés aurait été plus important). Les logiciels d'analyse textuelle visant à comptabiliser les occurrences d'apparition de certains thèmes, souvent fort utiles, n'ont pas été utilisés ici, du fait du temps limité qui nous était imparti à l'époque (pour rappel, cette étude a fait l'objet de notre mémoire de DEA) et sachant qu'il s'agissait d'une étude exploratoire suivie d'études quantitatives approfondies par la suite. Nous prévoyons toutefois, dans un second temps, de reprendre les analyses menées ici avec un logiciel de ce type, afin d'approfondir nos résultats.

[90] Ces matrices ne sont pas présentées ici, pour limiter la taille du document. Le recours à un logiciel d'analyse textuelle à ce niveau aurait peut-être permis de mieux comptabiliser les liens entre les différentes catégories.

La dernière étape (6) consiste à interpréter les résultats obtenus. Dans notre cas, la lecture de la littérature sur le thème de la vie privée avait montré l'existence de liens entre certains facteurs explicatifs et l'attitude ou le comportement d'un internaute face à la sollicitation de ses données personnelles, ce que nous avons cherché à vérifier et/ou à confirmer à travers l'analyse. D'autres liens étant davantage suggérés, nous avons eu la possibilité de vérifier leur applicabilité potentielle. Enfin, l'existence possible de certains liens a été plus intuitive dès l'instant où nous les avons « déduits » de la transposition d'études réalisées dans d'autres champs de recherche, qu'ils soient issus du marketing (marketing direct, e-commerce, ..) ou non (psychologie, sociologie). Cet ensemble nous avait conduits à imaginer un modèle provisoire identifiant les liens possibles entre les facteurs explicatifs et la réponse du consommateur. L'interprétation des résultats nous a permis de clarifier ce modèle, de l'enrichir et d'aboutir ainsi à un modèle conceptuel plus complet, intégrant les résultats issus de l'analyse qualitative, enrichis des conclusions issues de la littérature (cf. figure 5.1). Ce modèle sera présenté plus en détail dans la dernière section de ce chapitre.

1.5 Validité et fiabilité de l'étude qualitative

1.5.1 La validité de l'étude

Assurer la validité d'une recherche consiste à vérifier la pertinence et la rigueur des résultats (validité interne) et à évaluer leur niveau de généralisation (validité externe).

La validité interne concerne l'adéquation entre l'observation (ou les résultats) et la réalité. Elle constitue généralement le point fort de la recherche qualitative dans la mesure où il est à la fois possible de contrôler certains effets (comme l'effet d'histoire, de maturation, de mortalité, ..) et de corroborer les données et les résultats.

Une des manières de s'assurer de la validité consiste à soumettre les conclusions de l'analyse aux répondants pour voir s'il y a accord avec les interprétations établies par le chercheur. Nous n'avons malheureusement pas pu le faire dans notre cas du fait de l'éloignement géographique et du manque de disponibilité des participants (les résultats ayant été analysés durant l'été, beaucoup d'entre eux étaient partis en vacances). Un autre moyen de « valider » les résultats consiste à faire en sorte que l'analyse soit guidée par une approche théorique, ce qui fut le cas ici. Les résultats peuvent, en effet, être corroborés (confirmés ou infirmés) par les études déjà menées sur ce thème. Yin (1989) conseille ainsi de comparer les résultats empiriques mis en évidence, aux propositions théoriques issues de la littérature. Ainsi, si nous

avons retrouvé, au cours de notre analyse, de nombreuses similitudes avec des études menées par d'autres chercheurs, nous avons aussi parfois trouvé des résultats différents, qui s'expliquent notamment par l'importance du phénomène culturel sur ce sujet.

La validité externe constitue généralement un point faible de la recherche qualitative car il est plus difficile de généraliser les résultats qu'en recherche quantitative. On ne peut généraliser les résultats qu'aux contextes identiques à celui étudié. Or, celui-ci est souvent réduit. Ainsi, pour notre analyse, il se limite aux consommateurs français, internautes, correspondants au profil de nos répondants. Il est néanmoins possible d'améliorer la validité externe de l'étude en comparant les résultats avec la littérature (Eisenhardt, 1989). Ceci permet d'interpréter différemment les cas étudiés et d'identifier de quelle manière il est possible de dépasser les spécificités de chaque cas (ici, chaque internaute interrogé) pour généraliser les résultats à un univers plus vaste. Dans notre étude, sauf à démontrer que le choix des participants aurait été mal raisonné, la diversité des personnes interrogées autorise à envisager une généralisation à l'ensemble de la population des internautes français (avec toutes les précautions nécessaires toutefois). La faible taille de l'échantillon vient cependant limiter une telle ambition.

1.5.2 La fiabilité de l'étude

La fiabilité consiste à s'assurer que les différentes opérations de la recherche pourront être répétées avec les mêmes résultats par des chercheurs différents ou à des moments différents. Elle constitue généralement le point faible de la recherche qualitative puisqu'il n'existe pas de contrôle systématique des biais, ni de recours à des tests de fiabilité. De plus, ce type de recherche est rarement « réplicable » dans les mêmes conditions. Si le recours aux tests statistiques dans le cadre de l'analyse de contenu peut constituer une exception, tel ne fut pas le cas pour notre étude.

En recherche qualitative, la fiabilité de la recherche dépend essentiellement du matériau fourni par le chercheur, de sa capacité à transmettre aussi fidèlement que possible sa manière de conduire l'étude, de façon à permettre à d'autres auteurs de la répliquer. Comme le signalent Miles et Huberman (1991), elle est donc liée à la capacité du chercheur à décrire concrètement son processus de recherche. Nous nous sommes particulièrement attachés, dans cette partie méthodologie, à décrire aussi fidèlement que possible la manière dont nous avons procédé, espérant ainsi avoir atteint un niveau de fidélité acceptable.

Il existe plusieurs moyens d'accroître la fiabilité « interne » de la recherche. Le premier consiste à mener les entretiens en respectant les règles élémentaires dans ce domaine (pas d'orientation des propos, reformulation, ..) et en conservant, du début à la fin, une attitude d'empathie vis-à-vis du répondant. Ceci doit permettre de s'assurer que les données obtenues sont fidèles à la pensée des interviewés.

Parmi les principales stratégies permettant d'accroître la confiance dans les résultats, trois d'entre elles sont plus particulièrement utilisées par les chercheurs « qualitatifs ».
La recherche d'explications rivales est une première possibilité. Elle peut se faire de manière inductive ou logique. Il peut s'agir, par exemple, de rechercher d'autres moyens d'organiser les données, ce qui pourrait aboutir à des résultats différents. Au cours de notre analyse, nous avons ainsi cherché à plusieurs reprises à trouver d'autres explications possibles aux phénomènes et représentations décrites par les répondants.
Une autre « solution » consiste à trouver des « cas déviants », en recherchant des données susceptibles de prouver l'inverse de ce que les premiers résultats disent. Si nous n'avons pas eu la possibilité de mener davantage d'entretiens, il faut cependant noter que nous avons couvert une large étendue des préoccupations vis-à-vis de la collecte de données personnelles. Nous avons en effet recensé aussi bien des individus favorables que des individus réticents, voire absolument opposés. Malgré la faible taille de l'échantillon, nous avons donc réussi à identifier des individus plutôt « extrêmes » dont nous avons pu confronter l'opinion avec celle de personnes à la position « moyenne ».
La triangulation est aussi une technique couramment employée par les chercheurs qualitatifs pour consolider la crédibilité des résultats. Parmi les différents types de triangulations, nous avons surtout testé différentes théories permettant d'interpréter les résultats. De plus, nous avons systématiquement comparé nos résultats avec le maximum de sources possibles, notamment les derniers sondages et publications réalisés aux Etats-Unis.

2 Résultats de l'étude qualitative

Les résultats de l'analyse concernent les six points suivants : la définition des concepts (1) ; les spécificités perçues d'Internet (2) ; l'attitude des répondants face aux pratiques de traitement de données (3) ; les facteurs susceptibles d'influencer le comportement de réponse (4) ; les critères servant de base à l'évaluation de la sollicitation (5) et enfin les comportements de réponses (6). D'autres résultats moins cruciaux (dont la portée est plus accessoire puisqu'ils ne constituent pas réellement des découvertes par rapport à la

connaissance déjà établie sur le sujet ou qu'ils ne concernent pas le cœur de la problématique étudiée) ne seront pas détaillés[91]. Notre objectif est ici de signaler les principaux résultats susceptibles d'esquisser un début de réponse à nos questions de recherche et d'être utilisés ultérieurement dans le cadre de ce travail doctoral.

Nous présentons ci-après ces six points en insistant, plus particulièrement, sur les résultats qui sont réellement nouveaux ou qui sont essentiels à la compréhension[92]. Des tableaux donnant des exemples de verbatims pour chacun d'eux sont présentés ci-après ou en annexe 5.5.

2.1 Perceptions et définitions des principaux concepts en jeu

Nos entretiens montrent que les perceptions des internautes français, relatives aux concepts étudiés (donnée personnelle et vie privée notamment), sont parfois plus larges que les définitions issues de la littérature et de la réglementation en vigueur (cf. Annexe 5.5).

Les perceptions des répondants, relatives aux sollicitations commerciales dont ils peuvent faire l'objet, font également apparaître cinq catégories de menaces à ce sujet, manifestations de ce qu'ils considèrent comme une « invasion de leur vie privée ». Au delà de l'aspect dérangeant des relances commerciales et de l'utilisation parfois abusive des données, il faut souligner le côté éminemment affectif du phénomène, tel qu'il transparaît des entretiens, et qui est souvent peu pris en compte dans les recherches sur ce thème. Nos interviews montrent d'ailleurs que le degré d'invasion perçu est très différent selon les répondants et que cette menace concerne plus l'utilisation des informations (relances, transfert de fichiers) que la collecte elle-même, résultat déjà signalé dans la littérature (Cranor, Reagle et Ackerman 1999). Cette différence pourrait notamment dépendre de l'expérience passée de l'individu, mais aussi de sa personnalité et de sa propre perception des concepts précédents.

2.2 Les spécificités perçues d'Internet

Notre analyse montre que les gens considèrent Internet comme un média particulier. La virtualité et l'interactivité de cet outil modifient ainsi de façon substantielle leurs perceptions vis-à-vis des pratiques de collecte de données. Ceci se retrouve dans les avantages et les inconvénients qu'ils lui attribuent (cf. Annexe 5.5). Toutefois, l'attitude face à ce média de collecte s'avère, là aussi, différente selon les personnes. Car, si beaucoup ont l'impression que

[91] Nous renvoyons le lecteur intéressé à notre mémoire de DEA qui présente en détail l'ensemble des résultats.

[92] Pour de plus amples informations sur ces résultats, nous renvoyons le lecteur intéressé, soit à notre mémoire de DEA (Lancelot-Miltgen 2002), soit à une communication sur ce sujet, présentée au Congrès de l'Association Française du Marketing organisé à Tunis en 2003 (Lancelot-Miltgen 2003).

la collecte y est plus systématique ou plus intrusive, d'autres soulignent au contraire la sensation de contrôle que leur offre cet outil. Cette différence de perception paraît notamment liée aux variables sociodémographiques et à l'expérience de l'individu en la matière.

2.3 L'attitude face au traitement des données personnelles

Les personnes interrogées ont des attitudes très différentes à l'égard du traitement qui peut être fait de leurs données personnelles et des entreprises qui y ont recours. On distingue ainsi trois grands types de profils qui se différencient par leur manière d'être préoccupés (ou non) par le respect de leur vie privée (cf. tableau 5.4).

Tableau 5.4 Typologie des attitudes face au traitement des données personnelles

Groupe	Caractéristiques	Extraits d'interviews
Les inquiets (4 répondants)	- préoccupés par la protection de leur vie privée - rejet ou méfiance des pratiques de collecte et/ou utilisation de leurs données	*« dès que tu veux commander, il faut décliner tout un tas de renseignements et ça ça ne me plaît pas … surtout le numéro de carte bleue. Pour moi, c'est très inquiétant et c'est systématique. On a l'impression d'être pisté. On veut toujours nous demander un maximum de choses »*
Les peu concernés (2 répondants)	- peu préoccupés par les questions de respect de la vie privée - se dévoilent assez facilement	*« quand je suis sollicitée par des messages sur outlook, je leur réponds facilement, là je suis pour … je participe aux sondages parce que je mets à la place des gens. Je me dis que si personne ne répondait, les pauvres ils n'auraient pas de travail, donc là je suis assez conciliante »*
Les utilitaristes (3 répondants)	- comparent les coûts et les bénéfices de la divulgation - abandonnent leur intimité s'ils y trouvent un avantage personnel	*« ça dépend toujours de ce qu'il y a comme produit derrière et à quoi ça sert de donner ce renseignement »* *« si c'est un produit qui peut m'intéresser, là je donnerais mon nom, mon adresse … si c'est pour l'utiliser pour faire une publicité, là je pense que je ne donnerais pas »*

Parmi les répondants, certains témoignent d'un niveau d'inquiétude assez (voire très) élevé, qui se traduit par une attitude plutôt (voire totalement) négative, allant de la méfiance au rejet pur et simple de telles pratiques de la part des entreprises (que ce soit au niveau de la collecte ou de l'utilisation ultérieure des données collectées). Sur les 9 participants, 4 appartiennent plutôt à cette catégorie des « inquiets » qui se reconnaissent essentiellement par leur niveau élevé de préoccupation pour le respect de leur vie privée et par le besoin qu'ils ressentent de se protéger. 2 répondants sur les 9 jugent globalement ces pratiques peu dérangeantes et peuvent ainsi être qualifiés de « peu concernés ». Ils sont dans l'ensemble peu préoccupés par la préservation de leur vie privée et dévoilent plutôt facilement leurs données personnelles. Plusieurs avouent cependant que leur attitude favorable (et bienveillante) est parfois liée au

respect de certaines conditions (*« ça ne me dérange pas dès l'instant qu'il n'y a pas de questions privées »),* soulignant ainsi l'importance de la sensibilité des données demandées. On peut penser, à l'analyse de leurs discours, que ce sont des individus qui font facilement confiance, qui se confient facilement, qui aiment aider les autres (*« je me dis aussi il faut qu'ils améliorent … si ça peut apporter un plus au niveau de leur entreprise, au niveau du produit qu'ils vendent là je suis tout à fait d'accord »)* et qui voient plutôt le côté positif des choses, signe de l'impact probable des variables individuelles de personnalité et de valeurs.
Enfin, certains répondants (3 participants), que l'on peut qualifier d'« utilitaristes », voient avant tout dans la collecte de leurs données l'occasion de profiter d'avantages variés (comme des offres personnalisées, des bons de réduction, …). Ils acceptent donc de donner des informations uniquement s'ils estiment en retirer un bénéfice personnel. Pour prendre leur décision, ils procèdent à une comparaison des coûts et des bénéfices de la divulgation et sont prêts à perdre un peu d'intimité si ce sacrifice leur permet d'avoir un gain net substantiel.

Ce bilan des attitudes des internautes français face au traitement des données à caractère personnel est similaire aux classifications issues de la littérature, notamment à la typologie élaborée à l'origine par Westin (1996)[93] et qui fait toujours référence dans ce domaine. Nous tenons cependant à préciser que pour pouvoir établir une telle classification, il a fallu analyser les verbatims de manière approfondie, cette démarche étant rendue difficile par le caractère parfois contradictoire des discours. Il s'agit donc pour l'essentiel de tendances, qui ne présentent aucun caractère purement scientifique. Cette analyse montre par ailleurs que les variables individuelles, mais aussi les variables environnementales (les « inquiets » étant par exemple peu confiants dans la capacité du législateur à assurer une protection suffisante), pourraient avoir une influence importante sur ces attitudes. Enfin, il faut remarquer que l'attitude de l'individu face aux pratiques de traitement des données (thème retenu ici) correspond en tout point à sa préoccupation pour le respect de la vie privée, concept majoritairement utilisée par les chercheurs pour établir ce type de classification et que nous utiliserons donc également dans notre modèle conceptuel.

2.4 Les facteurs influençant le comportement de réponse

2.4.1 L'influence des facteurs environnementaux

Comme l'a montré la littérature, l'environnement aurait une emprise notable sur l'attitude de l'individu face au traitement des données, par le biais du niveau de connaissance des pratiques

[93] Nous reviendrons plus précisément sur cette typologie dans le chapitre 6.

des entreprises d'une part et des droits offerts aux consommateurs par la loi, d'autre part. Notre étude montre toutefois que ce n'est pas tant le fait de disposer d'une réglementation protectrice qui rassure le consommateur que celui de connaître les droits afférents à celle-ci. Parmi les facteurs influents, on relève le système de réglementation mais aussi la culture, la technologie et le bouche à oreille (cf. Annexe 5.5), déjà identifiés dans la littérature. Nous présentons donc ci-dessous les résultats qui se distinguent et/ou complètent cette dernière.

Milne (2000) avait suggéré l'effet possible de la technologie sur le comportement de l'individu en matière de transmission d'informations, mais ne l'a pas réellement démontré. Il considère que la réglementation agit surtout sur la stratégie de l'entreprise en matière de collecte, sans s'intéresser à son effet sur l'attitude des consommateurs. Nos entretiens confirment que technologie et réglementation ont une influence significative sur le niveau de préoccupation des individus, reflet de leur attitude face au traitement des données.
Concernant la culture (ou plutôt les « valeurs culturelles », au sens de Hofstede), son incidence sur la préoccupation des individus en matière de respect de la vie privée a déjà été démontrée (Milberg et al. 1995, Rustemli et Kokdemir 1993). Notre étude souligne que le passé historique et la façon dont celui-ci est perçu par l'individu joueraient aussi un rôle. Comme l'explique un répondant *« Je serais tenté de dire qu'historiquement les listes, ça a toujours une connotation pas très libérale. Que ce soit les listes noires d'Hollywood au temps du McCartisme, etc.. , en remontant par plein d'exemples historiques, généralement, ce n'est pas très bon signe. Donc, j'hésite, j'ai de toute façon des réticences ».*
Le bouche à oreille (alimenté par les connaissances et les médias) aurait aussi une influence significative, à la fois sur la connaissance des pratiques de collecte et de la réglementation, mais aussi sur le degré d'attention et de confiance avant et pendant la sollicitation.

2.4.2 L'influence des facteurs individuels

L'analyse des interviews confirme l'influence des facteurs individuels sur l'attitude de l'internaute en matière de traitement de données personnelles et leur effet possible sur la réponse de ce dernier, en cas de sollicitation. Sans surprise, parmi les facteurs identifiés, on retrouve les données sociodémographiques, l'expérience, la personnalité et les valeurs (cf. Annexe 5.5), catégories qui ont déjà été identifiées dans la littérature (cf. chapitre 4). Nous développons ci-dessous les principales nouveautés par rapport à la littérature et laissons de côté les résultats qui ne font que confirmer ce que les études passées ont déjà démontré.

Bien que supposé, l'effet des données sociodémographiques sur l'attitude des individus face à la collecte de données n'a jamais été réellement prouvé, tant les recherches en la matière offrent des résultats contradictoires. Nos analyses offrent une explication possible à ce phénomène. Ainsi, l'effet de l'âge pourrait être en partie lié à la maîtrise de la technologie, principalement d'Internet, qui s'avère différente selon les générations. Comme l'indique un des participants *(« les jeunes ont moins d'a priori parce qu'ils maîtrisent mieux la technologie »)*. Cette maîtrise étant plus délicate pour les gens âgés, cela expliquerait que ces derniers soient plus réticents à se dévoiler sur la Toile. Plusieurs répondants soulignent aussi l'impact possible de la CSP et font principalement la distinction entre les actifs et les inactifs. Ils relient donc cet aspect à la disponibilité dont l'individu peut faire preuve, lorsqu'il se trouve sollicité pour fournir des informations à caractère personnel. Ainsi, les inactifs sont plus disponibles pour fournir des informations et vont donc plus facilement répondre.

Notre analyse souligne aussi l'effet de l'expérience de l'individu sur son attitude face à la collecte et indique que celle-ci influencerait également son comportement. L'expérience en matière d'utilisation d'Internet, d'achat à distance et de marketing direct, a déjà été étudiée par des chercheurs, qui suggèrent que celle-ci aurait une incidence sur la perception de contrôle détenu par le consommateur sur ses données (Miyazaki et Fernandez 2001) ainsi que sur son désir de se faire rayer des listes (Milne et Rohm 2000). Notre étude montre que l'expérience passée de l'individu en matière de collecte aurait aussi une influence. Alors que Culnan (1993) a étudié l'effet de ce facteur sur l'attitude vis-à-vis de l'utilisation secondaire des données, nous remarquons qu'il agirait aussi sur l'attitude vis-à-vis de la collecte et sur le comportement de dévoilement. Comme l'explique un répondant *« j'accepte de donner mes coordonnées si je sais que ça va pas m'importuner. A force de voir comment ça marche, on imagine. Il suffit de connaître le système »*.

Pour certaines personnes, c'est manifestement la personnalité qui influence le plus leur attitude, comme le reconnaît ce répondant : *« ça rentre dans mon caractère, je ne suis pas expansive, je me dévoile pas donc certaines questions, ça va être une agression, je vais les vivre comme telle »*. Des chercheurs en psychologie avaient déjà constaté que les « choix de vie privée » (comme la préférence pour l'intimité) étaient souvent associés à certains traits de personnalité (Pedersen 1982). Notre analyse signale que des critères tels que l'image de soi et l'importance du regard des autres auraient aussi une action notable sur la capacité de l'individu à se dévoiler et donc sur son attitude face à la collecte.

Concernant les valeurs, la littérature a surtout étudié l'influence de l'attachement au respect

de la vie privée (Stone et al. 1983). Nous trouvons, pour notre part, que l'attachement au respect des libertés pourrait aussi jouer un rôle. Comme le signale un répondant *« c'est une sensibilité personnelle ; par principe, il est très sensible [..] aux libertés publiques »*.

2.4.3 L'influence des facteurs situationnels

Parmi les facteurs situationnels ayant une influence sur le comportement de l'individu en cas de sollicitation, le type de données demandées se révèle être particulièrement important. Nos interviews soulignent cependant que les individus ne placent pas toujours au même niveau de sensibilité des données qui sont pourtant proches et font *a priori* partie de la même catégorie (ex : l'âge et la date de naissance comme données sociodémographiques). Ceci semble dû au fait qu'ils ne leur attribuent pas les mêmes conséquences en cas d'utilisation ultérieure *(« l'âge je le donne encore. La date d'anniversaire, après on a la joie et le bonheur de voir un mail bon anniversaire. Une fois, ça va, mais 10, c'est plus pénible »).* De plus, comme la littérature le suggère, il apparaît que cette sensibilité est non seulement propre à l'individu, mais également dépendante, dans une certaine mesure, de la situation (Milne 1997 ; Cranor, Reagle et Ackerman 1999). Des répondants confient par exemple qu'une donnée considérée comme privée vis-à-vis d'une firme inconnue peut être considérée comme moins sensible s'il s'agit de la confier à un fournisseur habituel. De plus, au-delà du type d'informations demandées (et donc de leur niveau de sensibilité), d'autres critères tels que la manière de poser les questions, le degré de précision exigé et/ou la quantité totale d'informations sollicitées pourraient également influencer le processus de réponse.

D'autres facteurs situationnels ont une influence sur le comportement de l'individu en cas de sollicitation (cf. Annexe 5.5). Aux facteurs identifiés dans la littérature (politique de respect de la vie privée, type d'entreprise et dispositif de collecte), nos entretiens nous amènent à en proposer un autre, lié au contexte et/ou aux « particularités » du contact. Nous revenons brièvement sur les facteurs déjà détaillés au chapitre 4 avant d'insister sur ce nouveau facteur.

Alors que la littérature souligne l'importance de la politique de respect de la vie privée adoptée par le site (dispositif d'information et/ou de consentement notamment) dans la décision de l'individu de divulguer ou non ses données (Milne 1997), il apparaît que nos répondants ont, vis-à-vis de ces éléments, une attitude et un comportement parfois contradictoires, signe d'une difficulté à faire le lien entre ce que disent les entreprises et le niveau de protection réel. Si certains évoquent l'importance d'être informés sur l'utilisation

ultérieure de leurs données, beaucoup avouent ne pas toujours lire les notices d'informations. De même, alors que certains se disent préoccupés par le fait que des firmes ne respectent pas leur vie privée, aucun d'entre eux n'a jamais utilisé son droit d'opposition ou de réclamation. Il faut enfin remarquer que la sécurisation du dispositif et notamment celle du paiement apparaît primordiale, mais surtout en cas de commande. Beaucoup ont en effet admis ne faire aucun achat en ligne par crainte des risques de piratage de leurs codes bancaires, ce qui confirme les résultats de Miyazaki et Fernandez (2000) et de Jarvenpaa et Tractinsky (1999).

L'entreprise sollicitant les données et la relation que le consommateur entretient avec elle, constituent également un critère décisif dans la décision de dévoiler ou non des données personnelles. Les répondants identifient ainsi 4 « qualités » devant être possédées par la firme (le site) pour qu'ils acceptent de répondre : 1) pouvoir l'identifier rapidement (en cas de sollicitation par email) ; 2) posséder une certaine notoriété ou une bonne image de marque ; 3) la connaître personnellement (être déjà client) ; 4) la fiabilité et le sérieux qu'elle dégage, notamment à travers le design du site. Nos résultats confirment donc que les gens ne font pas toujours confiance aux firmes qui collectent des données et qu'ils distinguent celles qui leur sont familières et les autres (Sheehan et Hoy 2000).

Certaines caractéristiques du dispositif de collecte peuvent modifier l'évaluation que fait l'individu de la sollicitation et donc sa décision de divulguer (ou non) les données demandées. Parmi celles-ci, nous retrouvons : le média de collecte (Internet vs. médias « traditionnels »), le mode de collecte (direct par le biais d'un formulaire vs. discret par le biais de cookies), l'instrument de collecte et sa « disposition » (longueur, design, ..) et enfin, les circonstances de collecte (moment, temps disponible, humeur de l'individu, ..).

Il faut enfin signaler, comme le suggèrent nos entretiens, l'effet possible des « particularités » du contact et, en particulier, l'origine et l'objectif de celui-ci, que nous pouvons considérer comme faisant partie des « circonstances de collecte » (cf. chapitre 4). Les gens se révèlent ainsi plus favorables à la divulgation de leurs données quand ils sont à l'origine du contact (plutôt que l'entreprise) et s'il s'agit d'une prise de commande (plutôt qu'une simple visite). Le contexte dans lequel intervient la sollicitation, qui correspond au « mode de recrutement » de l'individu (inscription à un programme de fidélité, participation à un jeu concours, réponse à un sondage, ..) influencerait également la décision. Au-delà du type de compensations proposées dont l'effet a déjà été étudié dans la littérature, le contexte autour duquel intervient la sollicitation s'avère donc susceptible de jouer un rôle. Si l'idée que le contexte affecte les

perceptions et le comportement du consommateur a déjà été avancée, celle-ci n'a cependant jamais été étudiée par les chercheurs sur le respect de la vie privée. Notre étude souligne l'importance de ce facteur et encourage les travaux ultérieurs à se pencher sur cette question.

2.5 L'évaluation de la sollicitation

L'évaluation de la sollicitation constitue une étape particulière au sein du processus de réponse. C'est en effet à l'issue de cette phase que l'internaute va décider de fournir ou non les données demandées. L'analyse des discours fait ressortir quatre critères sur lesquels les individus fondent majoritairement leur évaluation : la confidentialité (et/ou le contrôle) perçue des données, la sensibilité perçue des données (déjà évoquée précédemment), la pertinence des données sollicitées (en fonction du contexte évoqué) et enfin, et peut être surtout, le gain net à répondre, évaluation de la différence entre les coûts et les bénéfices à se dévoiler.
Nous présentons successivement ci-dessous ces quatre critères d'évaluation.

2.5.1 La confidentialité perçue des données sollicitées

Pour bon nombre d'internautes qui se trouvent sollicités pour fournir des données sur un site, le contrôle sur ces informations concerne, d'une part le choix de ne pas répondre et, d'autre part, la possibilité de dire que les données ne devront pas être transmises à des tiers. Concernant le premier critère (liberté d'action), les interviewés sont quasi unanimes pour dire qu'il est généralement respecté *(« dans la majorité des cas, on a quand même le choix, on n'a pas un couteau sous la gorge, c'est encore de notre plein gré qu'on fait les choses »).* Dans l'ensemble, il ne s'agit donc pas d'un critère différenciant et sur lequel l'individu se fonde pour prendre sa décision. Les discours des répondants concernant le deuxième critère (notion de confidentialité) montrent qu'il s'agit là d'un point très important mais pour lequel certains s'avèrent sceptiques quand à la réalité des faits. Comme le déclare une des personnes interviewée : « *c'est une clause censée engager l'entreprise à ne pas partager les informations, ce n'est pas présenté comme un contrat, c'est une clause simple, c'est pas une certitude* ». Dès lors, la majorité des participants admettent y accorder une attention particulière *(« je vais faire attention .. les trucs à cocher .. vous nous autorisez à dévoiler ces infos à des partenaires* ») et utiliser ce critère pour évaluer la situation *(« si on a cette preuve qu'il y a la confidentialité … que ça va vraiment qu'à eux, qu'il n'y a pas d'interlocuteur pour piquer l'info … si j'ai cette assurance là, je donnerais peut être plus facilement »).* A l'opposé, tout manque de transparence de la part de l'entreprise entraîne une

incompréhension, voire une réaction virulente. Un des participants explique : *« la collecte est grave quand l'internaute n'est pas averti que son formulaire va transiter chez les concurrents, partenaires et filiales … on lui dit que ça reste confidentiel, c'est complètement faux. C'est grave car il y a mensonge et irrespect de la vie privée* ».

Si les internautes interrogés avouent ne pas lire les chartes de confidentialité qu'ils trouvent souvent longues, peu claires voire illisibles, ils se fient plus facilement aux mentions figurant en bas du formulaire ainsi qu'aux cases pré-cochées (ou non) signalant le choix concernant l'utilisation ultérieure des données divulguées. Comme l'explique ce répondant « *en général, maintenant il y a une petite case .. je décoche assez facilement à moins que ce soit quelque chose de moins commercial* ». Au delà du mode de consentement proposé par l'entreprise, les internautes aiment aussi être informés de l'utilisation qui sera faite de leurs données. Comme l'explique une de nos répondantes : *« si j'ai la liste des partenaires ou leur secteur d'activité, je préfère parce qu'on sait à quoi s'attendre, on peut prévoir un peu ce qu'on va recevoir, savoir si on sera déçue ou pas, importunée .. et donc avoir la liberté de choisir* ». Donner des informations sur l'utilisation ultérieure serait donc un moyen pour l'entreprise d'accroitre la confidentialité perçue des données divulguées.

2.5.2 La sensibilité perçue des données sollicitées

Les informations sollicitées par l'entreprise par le biais du formulaire à remplir semblent jouer un rôle majeur dans l'évaluation de la sollicitation et donc dans le comportement de réponse. Un répondant dit : *« si on me pose des questions plus précises, là tu t'impliques plus, tu te dévoiles plus ».* Certains individus peuvent ainsi être découragés de répondre s'ils estiment que les informations demandées sont trop sensibles *(« je donne sur des choses qui sont basiques. Mais tout ce qui est très personnel, ce n'est pas que je n'ai pas confiance, mais ça ne regarde pas les entreprises »).* D'autres indiquent clairement qu'il ne faut dépasser un certain seuil de curiosité, sous peine de les voir renoncer : *« s'il y a des questions qui me semblent aller trop dans la sphère de la vie privée, je ne réponds pas ».* L'ensemble de ces verbatims montre donc que la sensibilité perçue des données demandées est un critère majeur de l'évaluation de la sollicitation.

2.5.3 La pertinence perçue des données sollicitées

La pertinence est un autre critère utilisé par les consommateurs pour évaluer la sollicitation et prendre leur décision. La majorité des répondants expliquent être attentifs au fait que la

collecte et les données sollicitées soient utiles et/ou pertinentes par rapport à la transaction en cours. L'un d'eux déclare : *« les centres d'intérêt c'est un renseignement normal pour mieux cerner le client, mais le situer géographiquement c'est pas nécessaire »*. Les individus font donc clairement la distinction entre les informations qu'ils considèrent comme indispensables pour établir l'échange et celles qu'ils jugent superflues *(« je vais donner les informations qui sont nécessaires par rapport à ma demande. Si c'est pas nécessaire, je vois pas pourquoi je les donnerais »).* La présence de données jugées inutiles peut d'ailleurs conduire l'individu à imaginer les conséquences ultérieures d'une telle divulgation. Il apparaît ainsi, à l'analyse des discours, que moins les données exigées seront jugées pertinentes (exemple : l'âge lors d'une prise de commande), plus l'individu aura tendance à anticiper des conséquences négatives et plus il sera dès lors réticent à fournir les informations. Comme nous le verrons ultérieurement, cette évaluation peut cependant le conduire à des stratégies de réponse différentes. Il pourra soit décliner la participation, soit se limiter à l'essentiel, soit encore fournir des réponses erronées aux questions jugées non pertinentes. L'évaluation du gain net à répondre constitue un autre critère susceptible de l'aider à choisir entre ces différentes stratégies.

2.5.4 Les coûts et bénéfices perçus de la divulgation

La littérature signale que, pour prendre sa décision, le consommateur procède au calcul du rapport coût/bénéfice de la divulgation. Il s'agirait notamment de s'assurer que les données fournies seront utilisées de façon « juste » et qu'il ne souffrira pas de conséquence négative.
Notre analyse des discours montre que dans l'ensemble, les gens se montrent moins sensibles au désir de garder leurs informations personnelles qu'ils ne sont insatisfaits des bénéfices qu'ils retirent de la divulgation. Ainsi, la majorité des répondants estime que les avantages obtenus sont dans l'ensemble *« assez minces »*. On distingue deux grands types d'avantages pouvant être obtenus : des bénéfices hédoniques (*« ce genre de trucs, ça m'amuse de remplir les petites cases »)* d'une part, et utilitaires (*« je coche cadre dans l'optique ils vont m'envoyer des offres qui vont correspondre à mes goûts, aux goûts des cadres »)* d'autre part. On note aussi une opposition entre ceux qui concèdent le côté intéressé de leur démarche et ceux qui expliquent donner souvent des renseignements « gratuitement », par envie de *« faire une bonne action »*. Des répondants reconnaissent également l'idée d'échange avec la firme qui les sollicite, expliquant accepter de divulguer des données s'ils ont un retour *(« je l'envisage comme la contrepartie quand il y a des services gratuits. C'est plus un échange de bons procédés que vraiment un bénéfice que j'en retire de manière directe »).* Notre étude confirme donc que des gens sont prêts à se dévoiler en échange d'un bénéfice, conformément

à la théorie du contrat social (Milne et Gordon 1993 ; Phelps, Nowak et Ferrell 2000).

Concernant les inconvénients et risques perçus, il apparaît que ceux-ci sont plus attribués aux conséquences de la collecte (notamment aux relances de marketing direct) qu'à l'acte de divulgation lui-même. Certains internautes évoquent ainsi la peur d'être catégorisés, l'impression d'être pistés et le sentiment de perte de contrôle qu'ils ressentent du simple fait d'avoir dévoilé ces informations *(« se retrouver dans un fichier ultra détaillé qui peut être éventuellement revendu, ça ne me plaît pas particulièrement »)*. Parmi les risques évoqués, il convient cependant de distinguer ceux qui sont purement liés à la menace d'intrusion dans la vie privée et ceux, plus spécifiques à Internet, touchant aux problèmes de protection (invasion de virus, sécurité des transactions). Différentes personnes évoquent également les risques qui existent en cas de diffusion des données à des tiers : *« L'inconvénient c'est l'utilisation qui est faite des fichiers. On ne sait jamais entre quelles mains ça tombe. N'importe qui peut récupérer ces informations. Si ça reste des fichiers qui ont des desseins commerciaux, [..] ce n'est pas trop dramatique. Ça pourrait aussi avoir des fins politiques et là c'est plus gênant »*. Notre analyse suggère donc que les internautes comparent volontiers les coûts et bénéfices qu'ils peuvent retirer de la divulgation de leurs données et qu'ils utilisent majoritairement ce critère pour prendre leur décision. L'un d'eux remarque : « *si j'estime qu'ils vont m'envoyer de la pub, je laisse tomber, si ça m'intéresse de recevoir des choses, je donne* ».

Pour conclure, les internautes sont plus coopératifs lorsque ces quatre conditions sont réunies. Celles-ci se rapprochent de celles trouvées dans la littérature (Stone et stone 1990, Tolchinsky et al. 1981, Woodman et al. 1982), même si elles n'ont jamais été réellement testées empiriquement. De plus, aucun auteur n'avait jusqu'à présent réuni ces quatre critères au sein d'un même « modèle ».

Un consommateur accepterait donc de donner des informations à caractère personnel quand :

- il a l'assurance que les informations divulguées seront gardées confidentielles (non transmises à des tiers) : confidentialité perçue
- les informations demandées ne dépassent pas un certain seuil de sensibilité : (faible) sensibilité perçue
- les informations demandées sont jugées utiles et/ou pertinentes par rapport à la transaction en cours : pertinence perçue
- il pense qu'il a plus à gagner qu'à perdre en divulguant ses données : estimation du gain net perçu à répondre

Ces quatre critères constituent la phase d'évaluation de la sollicitation, préalable à la prise de décision finale, qui correspond aux comportements de réponse présentés maintenant.

2.6 Les comportements de réponse identifiés

Plusieurs types de réponses sont utilisés par les internautes en cas de sollicitation de leurs données personnelles. Notre étude a permis de distinguer 5 stratégies régulièrement adoptées par nos répondants (cf. tableau 5.5) : la Collaboration, l'Evitement, le Refus ou Renoncement, l'Affrontement et le Détournement d'identité[94]. Hormis cette dernière, ces stratégies ont déjà été identifiées dans la littérature, sans avoir été pour autant ni labellisées ni explicitées. Notre analyse montre également qu'un même individu peut recourir à l'une ou l'autre de ces stratégies selon la sollicitation, signe de l'influence notable des variables situationnelles.

Tableau 5.5 Typologie des stratégies de réponse

Comportement	Caractéristiques	Extraits d'interviews
Collaboration	Acceptation ponctuelle et/ou habituelle de fournir des données personnelles	*« en général je suis assez cool, je donne pas mal de renseignements »* *« si ça me paraît pas excessif et à partir du moment où ce n'est pas nominatif, ça me pose pas de problème »*
Evitement	Rester vague dans ses réponses	*« généralement je mets des choses assez générales. C'est pas la peur de Big Brother mais j'essaye de ne pas trop rentrer dans les détails .. dans tous les cas, je me montre évasif, vague ».*
	Ne pas répondre aux questions facultatives	*« je réponds en squizzant tout ce qui est optionnel [.] ce sont souvent ces cases là qui sont sources d'échanges »* *« si c'est facultatif, je laisse, je remplis pas. Je donne que ce qui est indispensable pour avoir le dialogue »*
Refus ou renoncement	Refuser catégoriquement et dès la sollicitation de fournir des données personnelles	*« si on demande comme information obligatoire le numéro de téléphone, c'est d'entrée non »* *« on n'est pas obligé de remplir des formulaires. Quand on me le propose, je dis non »*
	Renoncer à divulguer ses données personnelles au cours ou à l'issue de la démarche de collecte	*« si ça devient trop personnel et qu'il faut forcément répondre, je change de site »* *« il y a des fois, à la fin du questionnaire, on vous met « vous n'avez pas répondu à toutes les questions ». S'ils insistent de trop, je ne réponds plus »*
Affrontement	Piéger l'entreprise qui a sollicité les données en lui donnant des réponses erronées. Opérer par le biais du mensonge et/ou de la ruse	*« s'il y a des champs obligatoires qui me pèsent un peu de remplir, quand je trouve que c'est trop agressif, ça peut m'arriver de mettre des choses comme ça. C'est presque un prêté pour un rendu. S'ils veulent un fichier, ils l'auront. Il y aura plein de conneries dedans mais ils seront contents »* *« j'ai un ami, pour lui c'est un problème. C'est de la paranoïa aiguë. Il change de nom, d'adresse et de PC pour pas que les entreprises fassent de recoupement »*
Détournement d'identité	Emprunter l'identité d'autrui afin de déverser sur lui les risques liés à la divulgation	*« j'ai donné les renseignements au nom de mes parents et le numéro de carte bleue de ma mère »* *« j'ai donné le numéro de ma femme, pas le mien »*

[94] Pour de plus amples informations sur cette typologie développée par nos soins, nous renvoyons les lecteurs intéressés à notre mémoire de DEA et/ou à la communication qui s'en est suivie (Congrès AFM, 2003)

L'analyse des discours montre que la majorité des internautes français sont peu revendicatifs. Néanmoins, le fait que les stratégies de réponse les plus souvent utilisées par les répondants soient l'évitement et le refus semble indiquer que les pratiques utilisées par les entreprises pour collecter leurs données sont suffisamment gênantes pour justifier un comportement de protection. Notre analyse témoigne même de l'existence d'un lien entre l'appartenance des individus au groupe des « peu concernés » (PC), « inquiets » (I) ou « utilitaristes » (U) (cf. paragraphe 2.3) et le type de comportement utilisé par chacun d'eux. Ainsi, seuls les peu concernés et les utilitaristes acceptent de collaborer. De même, pour se protéger, les « inquiets » peuvent aller jusqu'à utiliser l'identité de quelqu'un d'autre. Enfin, la stratégie qui consiste à piéger les entreprises, notamment par le biais du mensonge, semble globalement peu utilisée, sauf par Cédric, qui revendique clairement son côté utilitariste. Ainsi, l'attrait du gain proposé en échange de ses données peut parfois le faire mentir, notamment s'il considère que les données demandées sont sensibles. Ce résultat vient donc contredire l'analyse de Regan (1995) qui pensait qu'il n'existait pas de corrélation entre le niveau de préoccupation pour le respect de la vie privée et le comportement en résultant.

Tableau 5.6 Liens entre préoccupation pour le RVP et stratégies de réponses

Participant	Préoccupation RVP	Collaboration	Evitement	Refus ou renoncement	Affrontement	Détournement d'identité
Frédérique	U	X	X	X		
Marie	I		X	X		X
Michel	I		X			X
Estair	I		X	X		
Michael	I		X			
Elizabeth	PC	X				
Marie-madeleine	U	X		X		
Cédric	U		X	X	X	
Josette	PC	X	X	X		

3 Mise en perspective des résultats

3.1 Les conclusions de l'analyse

3.1.1 Présentation d'un modèle conceptuel provisoire

L'analyse de nos entretiens révèle que les trois principales catégories de facteurs explicatifs étudiés (individuels, environnementaux et situationnels) n'influencent pas de la même manière les attitudes et les comportements des internautes en matière de fourniture d'informations personnelles. Il apparaît que les facteurs individuels et environnementaux se

manifesteraient principalement avant la sollicitation et auraient une influence sur l'attitude générale de l'individu face aux pratiques de traitement des données (mesurée par la préoccupation pour le RVP). Les variables situationnelles interviendraient, quant à elles, au moment de la sollicitation, et auraient une influence indirecte sur la réponse de l'individu, par l'intermédiaire des caractéristiques perçues (évaluation de la sollicitation). Cette interprétation nous conduit à bâtir un modèle conceptuel provisoire qui tient compte des résultats de l'analyse qualitative menée, tout en intégrant les allers-retours avec la littérature. Ce modèle (cf. figure 5.1) vise à souligner le caractère processuel de la réponse de l'individu face à la collecte de ses données puisqu'il présente successivement 3 phases (avant, pendant et après la sollicitation). Nous présenterons ci-après les deux premières, en indiquant notamment les « variables » susceptibles d'intervenir pour chacune d'elles[95].

Phase 1 : avant la sollicitation d'informations personnelles

La première phase du processus de réponse face à une collecte de données personnelles se situe avant la sollicitation et fait principalement intervenir les facteurs individuels (bloc 1) et environnementaux (bloc 2). Ces deux types de facteurs influenceraient l'attitude de l'individu face aux pratiques de traitement des données (bloc 4). L'effet des facteurs environnementaux serait toutefois essentiellement indirect, par le biais du niveau de connaissance des pratiques des entreprises en matière de traitement des informations, et, des droits des consommateurs en matière de respect de la vie privée (bloc 3). A noter que l'attitude face aux pratiques de traitement des données peut se rapprocher du concept de préoccupation pour le respect de la vie privée, couramment étudié au sein de la littérature.

Phase 2 : au cours de la sollicitation

La seconde phase, étape essentielle du processus (milieu de la figure), se situe au moment de la collecte et va de l'exposition à la sollicitation jusqu'à la prise de décision par la personne de divulguer ou non ses données. Interviennent ici les facteurs situationnels (bloc 5), qui viennent influencer l'évaluation qu'effectue l'individu de la situation de collecte (bloc 6). Cette évaluation porte sur les perceptions relatives aux caractéristiques majeures de la sollicitation, à savoir le niveau de confidentialité perçue, la sensibilité perçue des données demandées, la pertinence des données sollicitées et enfin le gain net à dévoiler ces données (rapport entre les coûts et bénéfices perçus à répondre). Cette phase d'évaluation pourrait parfois se voir occultée, du fait de la personnalité même de l'individu *(« je ne sais même pas*

[95] Pour ce qui concerne la dernière phase, qui ne constitue pas le cœur de notre sujet, nous renvoyons les lecteurs intéressés aux lectures déjà citées précédemment.

pourquoi j'ai participé à ce truc là, je l'ai fait sans réfléchir »). Elle pourrait aussi être rendue plus holistique, parce qu'il peut se produire un effet d'apprentissage (après plusieurs expériences avec la même entreprise par exemple). Dans la majorité des cas cependant, l'ensemble des critères serait évalué par les consommateurs. Cette évaluation conduirait ensuite l'individu à choisir, parmi les stratégies de réponses envisagées (bloc 7), celle qui convient le mieux à la situation. Signalons que l'attitude face au traitement des données (issue de la phase 1) pourrait également influencer cette décision.

Tout au long de la phase de sollicitation, interviendraient d'autres facteurs liés à la fois aux prédispositions personnelles et aux caractéristiques situationnelles, jouant le rôle de modérateurs (bloc 8). Parmi eux, on trouve principalement l'implication. En effet, plus la personne est impliquée par l'activité (ou le produit) de l'entreprise, plus elle a de chance d'accepter de répondre *(« là je suis prête à donner mon adresse email parce que le produit m'intéresse* »), même si toutes les conditions exigées ne sont pas remplies.

3.1.2 Principaux enseignements de l'étude

Cette étude exploratoire permet de tirer six enseignements majeurs concernant la manière dont l'individu perçoit la sollicitation de ses données et dont il prend sa décision.

On constate, en premier lieu, que les consommateurs français interrogés sont peu conscients des pratiques utilisées par les entreprises pour collecter leurs données d'une part et des droits que leur confère la réglementation à cet égard d'autre part. Toutefois, alors que nous aurions pu penser que cela augmenterait leur niveau de préoccupation, cela ne semble pas être le cas, car ils semblent avoir confiance dans la capacité du législateur à assurer leur protection.

Le deuxième constat concerne la diversité des attitudes des consommateurs français eu égard aux pratiques de traitement de leurs données personnelles (de hostiles pour certains à très conciliantes pour d'autres) et son incidence sur le choix des stratégies à privilégier en réponse aux sollicitations dont ils font l'objet.

Le troisième constat concerne l'impact différent des facteurs individuels et situationnels. Les premiers (conjointement aux données environnementales) interviendraient principalement en amont (avant la sollicitation). Ils influenceraient la préoccupation de l'individu pour le RVP, qui aurait elle-même un effet sur la réponse face à la sollicitation de données personnelles, dans une situation spécifique. Les seconds auraient un effet sur la décision de répondre ou non, par le biais des perceptions de l'individu relatives aux caractéristiques de la sollicitation.

Figure 5.1 Modèle processuel de réponse à une sollicitation de données personnelles

Le quatrième constat s'intéresse aux éléments analysés par l'individu pour prendre sa décision. Parmi les caractéristiques perçues de la situation de collecte, quatre d'entre elles semblent plus particulièrement évaluées : le niveau de contrôle sur l'utilisation ultérieure des données (notion de confidentialité), la sensibilité des données sollicitées, la pertinence de la collecte (ou plutôt des données collectées), les coûts et bénéfices liés à la divulgation, aboutissant au calcul du gain net à se dévoiler. Les questions relatives à la sécurité (problème de virus, paiement en ligne, …), spécifiques à Internet, sont incluses dans la notion de coût.
Parmi ces facteurs perçus et évalués, la sensibilité des données sollicitées paraît jouer un rôle majeur. Celle-ci est cependant à la fois subjective, personnelle et situationnelle (une donnée peut être considérée par un même individu comme sensible dans un cas et peu sensible dans l'autre) et donc très difficile à définir d'une part et à mesurer, d'autre part.
Le cinquième constat est lié à l'existence de plusieurs stratégies de réponse (collaboration, évitement, refus, affrontement, détournement d'identité) qui sont adoptées tour à tour par les individus, en fonction des circonstances. De plus, il apparaît que parmi ces stratégies, deux d'entre elles sont plus particulièrement utilisées par les internautes français : l'évitement (rester vague et ne pas répondre aux questions facultatives) et le refus (ne pas répondre).
Le sixième et dernier constat s'intéresse à l'impact du média de collecte. Etre sur Internet pourrait en effet modifier l'attitude et le comportement des consommateurs à l'égard de la collecte de leurs données personnelles. Cependant, les perceptions des répondants à ce sujet s'avèrent là aussi très hétérogènes et parfois contradictoires. En effet, si beaucoup soulignent la sensation de contrôle que leur offre ce média (choix et liberté de répondre), d'autres indiquent au contraire que la collecte de données y est plus systématique et plus intrusive (plus de questions personnelles notamment). Cette différence de perception serait notamment liée aux variables sociodémographiques et aux effets d'expérience.

3.2 Comparaison avec les recherches antérieures

L'importance du phénomène culturel, dans le contexte du respect de la vie privée et de la réponse à une sollicitation de données personnelles, invite à comparer les différences de réactions entre le public français et américain, afin d'identifier les éventuelles convergences et divergences de vues. La comparaison de nos résultats avec ceux issus des recherches précédentes laisse apparaître un certain nombre de points communs et quelques différences que nous allons indiquer rapidement ci-dessous.

Les points communs concernent principalement trois aspects :

- l'impact de la sensibilité des données demandées : reconnu par la littérature et confirmé ici,
- la différence de préoccupation des individus en matière de traitement des informations à caractère personnel : les typologies présentées dans la littérature semblent pouvoir être reproduites sur le public français ;
- l'existence de différentes stratégies de réponse, pouvant être employées tour à tour par les individus selon les circonstances, signe que la majorité d'entre eux ne raisonnent pas toujours de manière binaire (je donne/ je ne donne pas) : reconnu par la littérature et confirmé ici.

Au contraire, parmi les principales différences, on peut signaler :

- le fait que les internautes français soient assez peu conscients des pratiques utilisées par les entreprises pour collecter et utiliser leurs données personnelles sur Internet. Contrairement aux internautes américains, ils seraient surtout peu informés et/ou conscients des pratiques de collecte « discrète », par le biais de cookies notamment ;
- le fait que les internautes français paraissent peu informés de leurs droits en matière de RVP. Ainsi, les répondants ont peu évoqué le rôle des organismes chargés de faire respecter la réglementation en vigueur et n'ont pas exprimé, non plus, l'envie de déposer plainte contre des entreprises, dont les pratiques sont pourtant parfois à la limite de la légalité.
- le fait que les stratégies de réponse employées par les Français diffèrent sensiblement de celles des Américains (ces derniers étant plus enclins à se plaindre auprès de l'entreprise ou des organismes de défense et à en parler à leur entourage). Les types de réponse le plus souvent employés par les Français s'apparentent plus à un comportement de protection, signe d'un niveau de préoccupation certain, bien que non exprimé ouvertement.

Au-delà des points communs qui ressortent de cette analyse, l'ensemble des différences qui ont été soulignées sont le signe de l'importance du contexte culturel sur un sujet aussi délicat que celui du respect de la vie privée, ce que la littérature a déjà suggéré. Ceci conforte aussi l'intérêt que nous avons vu à mener une recherche spécifique au public français. L'étude présentée ici est ainsi la première réalisée sur ce thème en France, permettant alors de mieux comprendre les perceptions et réactions des internautes français à l'égard de la protection de leur vie privée. Elle montre notamment en quoi l'évaluation de la sollicitation (sur la base de 4 perceptions majeures) est une étape préalable à la prise de décision.

Conclusion du chapitre 5

Cette étude exploratoire qualitative visait à étudier en profondeur les freins et les motivations des individus lorsqu'ils se trouvent confrontés à une sollicitation de leurs données personnelles, en particulier sur Internet.

Après avoir expliqué les principaux objectifs de cette étude, nous avons justifié le choix de recourir à des entretiens individuels semi-directifs et avons exposé de quelle manière ceux-ci s'étaient déroulés. Nous avons ensuite présenté le profil de nos répondants et avons expliqué pourquoi nous avons opté pour une analyse du contenu des discours. Nous avons enfin indiqué sur quels critères nous nous fondions pour assurer la validité et la fiabilité des résultats obtenus.

Dans une seconde partie, nous avons exposé l'ensemble des résultats auxquels nous avons abouti, indiquant notamment quelles étaient les perceptions des individus interrogés concernant les concepts de « donnée personnelle » et de « vie privée ». Nous avons également souligné en quoi le média Internet soulevait des questions particulières, avant d'exposer les facteurs susceptibles d'influencer le processus de réponse, qu'ils soient de nature environnementale, individuelle ou situationnelle. Nous indiquions ensuite quels sont les critères utilisés par les consommateurs pour évaluer la sollicitation et pour aboutir au comportement choisi.

Dans une troisième et dernière partie, nous avons mis l'ensemble de ces résultats en perspective, en présentant notamment un cadre conceptuel provisoire indiquant le processus adopté par le consommateur pour prendre sa décision, lorsqu'il se trouve confronté à une sollicitation de ses données. Nous avons également insisté sur les principaux enseignements de cette étude, tout en les comparant avec les résultats issus des études antérieures sur le sujet. Enfin, nous avons clos ce chapitre en rappelant les principaux apports de cette étude.

Nous présenterons, dans le chapitre suivant, les résultats d'une seconde étude exploratoire, à vocation quantitative et descriptive cette fois. Nous dévoilerons notamment une typologie des préoccupations des internautes français, sur la base des classifications identifiées au préalable dans la littérature.

Partie II – EXPLORER ET BATIR

Sollicitation de données personnelles
et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

CHAPITRE 6
LA GESTION DES DONNÉES PERSONNELLES CHEZ LE CONSOMMATEUR-INTERNAUTE : EXPLORATION DESCRIPTIVE (ÉTUDE 2)

« Je ne transmets pas mes données personnelles à des entreprises que je n'ai pas sollicité.
Quand on me propose de transmettre mes données personnelles à des partenaires,
je ne le fais pas car je ne les connais pas à l'avance ».

Un des répondants à l'étude (anonyme)

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

Chapitre 1 : Les enjeux liés à la collecte de données clients

Chapitre 2 : Préambule sur la notion de vie privée

Chapitre 3 : Dévoilement de soi et divulgation de données personnelles

Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)

Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)

Chapitre 7 : Formulation du cadre conceptuel général de la recherche

Chapitre 8 : Méthodologie de la recherche

Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)

Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)

Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 6

Tous les individus ne perçoivent pas la vie privée de manière identique. La revue de la littérature le démontre et notre étude qualitative le confirme. Une des principales raisons à cela est le fait qu'il s'agisse d'un phénomène hautement contextuel (Schoeman 1984). Plusieurs recherches prouvent d'ailleurs qu'il est possible de scinder la population en plusieurs sous-parties selon le statut qu'elles accordent au Respect de la Vie Privée. La majorité (si ce n'est la quasi-totalité) des résultats obtenus jusqu'à présent provient d'échantillons de consommateurs ou d'internautes nord-américains. Or, force est de constater que les politiques mises en place par les entreprises ainsi que l'encadrement juridique diffèrent notablement d'un pays à l'autre (Gurau, Ranchhod et Gauzente 2003).

En France, l'étude de la perception du respect de la vie privée et de ses effets sur le comportement de partage de données personnelles demeure encore limitée, voire quasi inexistante. On sait ainsi peu de choses quant à la manière dont les consommateurs français gèrent leur identité en ligne, voire même comment ils se comportent quand ils sont sollicités pour fournir des données à caractère personnel sur le web. De même, on peut se demander si les internautes français peuvent, à l'instar de leurs homologues américains, être segmentés en groupes distincts, selon leur niveau de préoccupation envers le RVP[96].

C'est l'objet de cette étude exploratoire descriptive que d'identifier les habitudes des internautes français en matière de gestion de leurs données personnelles en ligne et de voir s'il est possible de segmenter cette population en plusieurs groupes distincts.

Nous indiquerons, dans une première partie, les objectifs de cette étude ainsi que l'organisme avec lequel nous avons collaboré pour la mener à bien (1). Nous présenterons ensuite la littérature en matière de typologies de préoccupations pour le respect de la vie privée (2), avant d'évoquer la méthodologie mise en œuvre dans le cadre de cette étude (3). Nous donnerons enfin les résultats auxquels nous avons abouti (4) avant de les mettre en perspective, en indiquant les limites et les principaux enseignements de cette étude (5).

[96] Pour mémoire ce sigle est utilisé pour évoquer le terme de « respect de la vie privée ». Les deux acceptions (sigle et terme complet) seront employées tour à tour, tout au long de ce document.

1 Présentation de l'étude

Nous présentons l'étude en indiquant ses principaux objectifs (1.1) et en expliquant le partenariat mise en œuvre à cette occasion (1.2)

1.1 Les objectifs de l'étude

Notre premier objectif, à travers cette étude, était de décrire les habitudes adoptées par les internautes français en réponse aux sollicitations de données personnelles dont ils peuvent faire l'objet lorsqu'ils surfent sur le Net. Il s'agissait aussi de voir si et en quoi ces habitudes peuvent être reliées à l'attitude des individus en matière de RVP. Cette étude visait ainsi à mieux comprendre l'incidence des préoccupations concernant le respect de la vie privée sur les comportements de partage de données personnelles.

Le second objectif était d'étudier le type d'avantages et/ou de bénéfices susceptibles de favoriser la divulgation de données personnelles avec, pour but, de dresser une typologie des internautes français en la matière. En cela, nous cherchions plus particulièrement à évaluer deux points importants :

1) les individus (internautes français) diffèrent ils quant aux avantages recherchés en échange de la divulgation de leurs données personnelles en ligne ?
2) Peut-on expliquer ces différences ?

Le troisième objectif, plus mineur que les deux précédents, était de pré-tester certaines hypothèses identifiées à travers la revue de la littérature et/ou l'analyse qualitative préalable. Il s'agissait notamment d'évaluer l'impact possible du contexte de sollicitation d'une part et du type d'interlocuteur demandant les informations d'autre part, sur le comportement de partage de données.

1.2 Présentation du partenariat avec la FING

Cette étude a été menée en collaboration avec la FING (Fédération Internet Nouvelle Génération). Il s'agit d'une association Loi 1901 qui compte plus de 160 membres, parmi lesquels on trouve principalement des médias, des entreprises industrielles et de service, des laboratoires de recherche, des établissements de formation, des collectivités territoriales voire

des administrations. L'objectif de la FING est de « repérer, stimuler, valoriser l'innovation dans les services et les usages des technologies de l'information et de la communication »[97].

Il existe, au sein de cette association, plusieurs groupes de travail dont un axé sur l'identité numérique, chargé spécialement de réfléchir et de produire des travaux sur la gestion des données personnelles en ligne et le recours à l'anonymat et au pseudonymat. Dans ce cadre, la FING envisageait, depuis quelque temps déjà, de réaliser une grande enquête en ligne ayant pour but d'identifier et de quantifier les pratiques de transmission de données personnelles des internautes français. Les responsables du groupe de travail sur l'identité numérique (dont Daniel Kaplan, délégué général de la FING) avaient d'ailleurs commencé à bâtir un questionnaire à cet effet. Ayant eu vent de mes travaux dans ce domaine, ils m'ont proposé de collaborer avec eux pour la réalisation de cette étude. Ils m'ont alors chargée d'améliorer leur version initiale du questionnaire et d'en analyser les résultats. Notre collaboration fut d'autant plus fructueuse que chacun en retirait des bénéfices. Pour ma part, il s'agissait là d'un terrain particulièrement intéressant qui me permettait, dans une logique exploratoire, d'élaborer une analyse descriptive des pratiques des internautes français en matière de transmission de données personnelles. Ce travail constituait donc un excellent complément à l'étude qualitative présentée précédemment. Comme tout partenariat cependant, il y avait des limites, la principale étant liée au fait que je n'ai pu ajouter qu'un nombre limité de questions et qu'il ne m'était pas possible d'inclure des échelles de mesure sophistiquées.

La suite de la présentation est organisée autour de quatre axes. Celle-ci débute par l'examen des études empiriques portant sur la segmentation des consommateurs en matière de respect de la vie privée. Le cadre méthodologique de l'étude est exposé dans un deuxième temps. Les résultats sont ensuite présentés avant de discuter, en dernier lieu, des apports et des limites soulevées par ce travail.

[97] Pour de plus amples détails sur le fonctionnement, les objectifs et les travaux de cette association, nous renvoyons les lecteurs intéressés à son site Internet : www.fing.org

2 Revue de littérature sur les typologies de préoccupation pour le RVP

Nous verrons tout d'abord les typologies réalisées dans ce domaine (2.1) avant d'évoquer les limites inhérentes à ces travaux (2.2). Nous indiquerons, dans un second temps, les principales variables d'influence identifiées par ces auteurs (2.3).

2.1 Les typologies existantes

Les recherches passées révèlent que les individus diffèrent quant à leur sensibilité vis-à-vis du partage de leurs données personnelles, et, que leurs préférences changent en fonction de variables individuelles, environnementales et situationnelles (Hoffman et al. 1999 ; Laufer et Wolfe 1977 ; Stone et Stone 1990), ce que confirme l'étude qualitative. Plusieurs auteurs ont alors cherché à segmenter la population sur la base des préoccupations pour le respect de la vie privée. Nous présentons la première étude réalisée dans ce domaine (2.1.1), qui est toujours considérée comme la typologie de référence par de nombreux auteurs, avant d'indiquer les autres typologies existantes (2.1.2).

2.1.1 La typologie d'origine (Westin, 1996)

Le premier auteur à avoir établi l'existence de différences de préoccupations pour le Respect de la Vie Privée, au sein de la population, est Alan Westin (1996). Au cours d'un sondage réalisé auprès de consommateurs américains, il met en évidence 3 groupes d'individus qui se distinguent de par leur attitude face aux pratiques de collecte et d'utilisation de données personnelles dans un contexte traditionnel (i.e. offline). Westin (1996) catégorise les sujets, ayant participé à l'enquête, sur la base de leurs réponses à 3 questions (affirmations) :

- « les consommateurs ont perdu tout contrôle quant à la manière dont leurs données personnelles sont collectées et utilisées par les entreprises »
- « la majorité des entreprises utilisent les données collectées auprès des consommateurs de manière honnête et confidentielle »
- « les lois et les mesures prises par les entreprises fournissent un niveau de protection acceptable pour le respect de la vie privée des consommateurs aujourd'hui »

Les individus dont les réponses à ces 3 questions soulignent la préoccupation pour le Respect de leur Vie Privée sont classés dans la catégorie des « fondamentalistes ». Ils représentent

25% des consommateurs interrogés. Ce sont des personnes qui estiment, d'une part, que les consommateurs ne contrôlent pas suffisamment leurs données personnelles et d'autre part, que la majorité des entreprises utilisent les données de leurs clients de manière abusive. Ceux dont les réponses dévoilent au contraire une faible préoccupation pour la protection de leurs données sont qualifiés de « non préoccupés » (20%). Ces individus se dévoilent facilement et, dans l'ensemble, font confiance aux firmes qui leur demandent des informations personnelles. Les répondants qui se situent entre ces deux extrêmes sont appelés « les pragmatiques » (55%). Ce sont des consommateurs qui recherchent avant tout leur propre intérêt. Ils analysent les bénéfices proposés en échange de leurs données ainsi que les risques encourus et les garanties offertes et se décident en conséquence. Ils n'acceptent de se dévoiler que s'ils ont confiance dans l'entreprise qui les sollicite. Cette typologie souligne l'importance relative de la catégorie des « pragmatiques » qui constituent, aujourd'hui encore, le cœur de cible de nombreuses entreprises. Ces individus semblent en effet, si ce ne sont les seuls, au moins les plus faciles, à convaincre, les autres se décidant surtout sur la base de leurs convictions personnelles.

Westin sera amené à réaliser plusieurs études de cette nature par la suite, lui permettant d'avoir une vue longitudinale de la répartition de la population américaine en termes de préoccupation pour le respect de la vie privée. Nous présentons en Annexe 6.1 les résultats obtenus entre 1996 et 2003.

2.1.2 Les autres typologies existantes

Par la suite, plusieurs chercheurs ont essayé d'élaborer une classification des préoccupations des consommateurs en termes de respect de la vie privée, en se basant sur celle de Westin (1996). La majorité de ces travaux porte sur Internet alors que la typologie d'origine concernait un contexte offline. Dans la mesure où chacun de ces auteurs a pu se baser sur les résultats des études précédentes pour établir ses hypothèses, nous exposerons les résultats issus de ces recherches dans l'ordre chronologique. Nous présentons une synthèse des principales conclusions issues de ces travaux dans le tableau 6.1.

Cranor, Reagle et Ackerman (1999) sont les premiers à réaliser cette analyse auprès d'internautes américains. Ils établissent leur typologie sur la base de l'attitude des répondants face au respect de la vie privée, ainsi que sur leurs intentions comportementales de réponse face à une collecte de données en ligne. Leur analyse aboutit à la formation de 3 groupes,

similaires à ceux de Westin (1996) : les « fondamentalistes » (17%), les « pragmatiques » (56%) et enfin, les « peu préoccupés » (27%). Ces derniers, bien que proches des « non préoccupés » identifiés par Westin, s'en différencient par leur souci de valoriser leur vie privée, si peu soit-il. Ainsi, s'ils sont plutôt enclins à fournir leurs données, ils sont aussi soucieux des conséquences que cela peut avoir et s'avèrent occasionnellement favorables aux mesures leur permettant de se protéger (comme la possibilité de retirer leur nom des listes).

La première étude réalisée dans un contexte européen est celle menée par Spiekermann, Grossklags et Berendt (2001). Ayant interrogé des membres d'une université allemande (dont 95% d'étudiants), ces auteurs identifient 4 segments de consommateurs qui se différencient d'après leur préoccupation en matière de respect de la vie privée sur Internet. A côté des « fondamentalistes » (30%) et des « peu préoccupés » (24%), ils trouvent deux catégories d'indécis. D'un côté, les « *identity concerned* » (20%), préoccupés par le fait de devoir fournir des données d'identification et, de l'autre, les « *profile concerned* » (26%), concernés par les questions portant sur leur profil sociodémographique ou leurs centres d'intérêt.

L'objectif de l'étude menée par Sheehan (2002) était de vérifier si la typologie de Westin (1996) s'appliquait aussi au contexte d'Internet. S'appuyant sur un échantillon représentatif de la population des internautes américains, l'auteur montre que la segmentation en 3 groupes est peu adaptée et qu'il est préférable de scinder la catégorie des « pragmatiques » (plus de 80% des individus) en deux. Elle distingue alors 4 groupes d'individus : les « non concernés » (16%), les « circonspects » (38%), les « prudents » (43%) et enfin les « alarmés » (3%).

A la différence des autres chercheurs, Hann et al. (2003) mènent leur étude auprès d'échantillons d'étudiants appartenant à deux pays différents (Etats-Unis et Singapour). Ceci permet d'accroître la validité de leurs résultats, qui convergent quelle que soit la nationalité du répondant[98]. Si ces auteurs identifient 3 segments d'internautes dans les deux pays, ceux-ci diffèrent en grande partie de ceux trouvés par Westin (1996), dans la mesure où deux nouvelles catégories apparaissent. A côté des « gardiens de la vie privée », qui font écho aux « fondamentalistes » et qui concernent respectivement 72% des répondants américains et 84% de ceux venant de Singapour, ils distinguent en effet et les « vendeurs de données » et les « chercheurs de confort ». Les premiers qui, comme leur nom l'indique, sont prêts à vendre leurs données, quelles que soient les mesures de protection offertes par les sites, se

[98] il faut cependant souligner la taille inégale de chacun de ces échantillons, l'analyse étant menée auprès de 78 étudiants américains et 165 issus de Singapour

différencient des seconds qui valorisent le confort (personnalisation, gain de temps) plutôt que les récompenses monétaires. La répartition très inégale entre ces 3 groupes (plus des trois quarts des individus appartenant au segment des « gardiens de la vie privée ») laisse cependant supposer la possibilité de scinder l'un d'entre eux en plusieurs sous-catégories, ce que les auteurs eux-mêmes reconnaissent.

Enfin, en se basant sur des questions concernant l'attitude face au respect de la vie privée en ligne, Jensen et al. (2005) trouvent également 3 segments, similaires à ceux obtenus par Westin (1996) : « les fondamentalistes » (34%), « les pragmatiques » (43%) et « les non préoccupés » (23%). Ils font cependant remarquer que la catégorie des fondamentalistes ne formerait pas un ensemble cohérent puisqu'ils relèvent, au sein de ce groupe, des personnes ayant des attitudes différentes face aux menaces qui pèsent sur leur vie privée. Pour ces auteurs, certains fondamentalistes ne seraient donc pas aussi préoccupés qu'on pourrait le croire.

2.2 Limites des typologies existantes

La relative convergence des résultats issus de ces recherches n'autorise guère de doute quant à l'existence de différences au sein de la population, d'une part, et en une segmentation en 3 ou 4 groupes, d'autre part. Il convient toutefois de souligner que des limites inhérentes à ces travaux apparaissent, afin de relativiser certaines conclusions et tenter, le cas échéant, d'y remédier. Ainsi, dans certains cas, les échantillons sont soit composés d'étudiants (Hann et al. 2003), soit de faible taille (93 réponses utilisées dans l'étude menée par Jensen, Potts et Jensen 2005), ce qui réduit la possibilité de généraliser des résultats. De plus, certains auteurs, comme Cranor et al. (1999), se contentent de prendre la typologie de Westin (1996) pour modèle[99] et font leurs analyses de manière à obtenir des résultats similaires. Ils ne cherchent donc pas à établir leur propre typologie mais plutôt à retrouver celle de Westin (1996).

Nous nous efforcerons pour notre part, tout en nous inspirant des travaux existants, de nous en détacher. L'idée est d'établir notre propre typologie, sans chercher à calquer nos résultats sur ceux établis précédemment dans la littérature. Une des principales raisons à cela est que nous travaillons sur la population française alors que la majorité des typologies existantes concernent la population américaine. Une des questions posées à l'origine par Westin (1996)

[99] Ils posent notamment les mêmes questions que celles proposées par Westin et présentées au paragraphe 2.1.1.

Tableau 6.1 Synthèse des principales typologies de consommateurs dans la littérature sur la préoccupation pour le Respect de la Vie Privée

Auteur (année)	Population étudiée (taille de l'échantillon)	Base de segmentation (variables analysées)	Types de consommateurs trouvés	Variables discriminantes
Westin (1996)	Consommateurs américains	3 questions générales sur l'attitude face au RVP	3 segments trouvés : - les « fondamentalistes » (25%) - les « pragmatiques » (55%) - les « non préoccupés » (20%)	- sexe : femmes plus préoccupées que les hommes et ce quel que soit le domaine (collecte, utilisation secondaire des données, accès, …)
Cranor Reagle Ackerman (1999)	Internautes panélistes américains 405 questionnaires (381 conservés pour l'analyse)	Questions sur l'attitude générale face au RVP et intentions de réponse à des sollicitations de données personnelles (dans différentes conditions)	3 segments trouvés : - les « fondamentalistes » (17%) - les « pragmatiques » (56%) - les « peu préoccupés » (27%)	- aucune différence sociodémographique - sensibilité perçue des données similaire dans tous les segments - les « fondamentalistes » sont les plus opposés au transfert automatique de leurs données
Spiekermann Grossklags et Berendt (2001)	Membres d'une université allemande (95% étudiants) 206 questionnaires (171 conservés pour l'analyse)	Questions sur la préoccupation pour le RVP (protection des données personnelles)	4 segments trouvés : - les « peu préoccupés » (24%) - les « préoccupés par leur identité » (20%) - les « préoccupés par leur profil » (26%) - les « fondamentalistes » (30%)	- les « peu préoccupés » sont ceux qui refusent le moins de se dévoiler - continuité de la communication (flow) : les « préoccupés par leur identité » ont un « flow » faible, ceux « préoccupés par leur profil » élevé
Sheehan (2002)	Internautes américains 3724 questionnaires (889 complétés) échantillon représentatif des internautes américains	Niveau de préoccupation face à 15 situations de collecte de données personnelles	4 segments trouvés : - les « alarmés » (3%) - les « prudents » (43%) - les « circonspects » (38%) - les « non préoccupés » (16%)	- sexe et niveau de revenu non discriminants - âge : non préoccupés et alarmés plus âgés - niveau d'études : non préoccupés et circonspects sont moins « éduqués » - utilisation d'Internet non discriminante - comportements de divulgation discriminants : l'enregistrement sur des sites, la fourniture d'informations incomplètes ou erronées
Hann, Hui, Lee et Png (2003)	Etudiants américains (84) et de Singapour (184) 268 questionnaires (243 classés dont 78 américains)	Questions sur les bénéfices acceptés en échange des données et sur les types de préoccupations envers la protection des données	3 segments trouvés : - les « gardiens de la VP » (72% pour US et 83% pour Singapour) - les « vendeurs de données » (20% et 8,5%) - les « chercheurs de confort » (8% et 8,5%)	- dans l'échantillon US, les « vendeurs de données » ont une plus grande expérience de dévoilement de données personnelles - dans les échantillons US et Singapour, les « chercheurs de confort » sont moins préoccupés par l'utilisation de cookies
Jensen Potts et Jensen (2005)	Internautes volontaires (175 majoritairement américains dont 93 conservés)	5 questions reflétant l'attitude face au RVP	3 segments trouvés : - les fondamentalistes (34%) - les « pragmatiques » (43%) - les « non préoccupés » (23%)	- sexe : les femmes sont plus préoccupées que les hommes pour les 5 questions - expérience d'achat en ligne non discriminante - pas d'effet sur la probabilité de lire une charte

et par certains de ses successeurs pour établir leur classification concerne d'ailleurs la manière avec laquelle « la loi et les mesures de protection prises par les entreprises assurent un niveau de protection suffisant ». Cet aspect du problème qui s'avère pertinent dans le cadre américain nous semble perdre, au moins en partie, son sens, dans le cas qui nous intéresse ici. La législation européenne (Directive Européenne, 1995) qui s'applique en France est en effet beaucoup plus protectrice que la réglementation américaine dans ce domaine.

2.3 Les variables d'influence

Dans la majorité des recherches de cette nature, pour décrire le profil des groupes et assurer une meilleure validité de leurs résultats, les auteurs explorent si l'appartenance des consommateurs aux différentes classes n'est pas liée à d'autres variables, qu'elles soient de nature sociodémographique, expérientielle, perceptuelle, attitudinale ou comportementale. L'étude de l'impact de ces facteurs sur les classifications obtenues fait l'objet du paragraphe suivant. Nous aborderons tour à tour chacune des catégories de variables évoquées ci-dessus.

Si certains auteurs concluent à des différences sociodémographiques entre les segments de consommateurs identifiés à travers leur typologie, d'autres soulignent au contraire l'absence de pouvoir discriminant de ces variables. Parmi les facteurs susceptibles d'occasionner des différences intergroupes significatives, on trouve principalement l'âge, le niveau d'études et le sexe. Toutefois, tandis qu'une seule étude, parmi celles évoquées, confirme l'impact possible des deux premiers (Sheehan, 2002), les résultats concernant le sexe sont largement controversés. Ainsi, quand Westin (1996) ou Jensen et al. (2005) pensent que les femmes sont plus préoccupées que les hommes (tendance à faire partie des « fondamentalistes »), les résultats obtenus par Cranor et al. (1999) et Spiekerman et al. (2001) ne montrent aucune différence significative selon le sexe, ni même selon aucun critère sociodémographique.

Différents chercheurs se sont également interrogés sur l'impact de facteurs d'expérience tels que l'expérience d'Internet ou l'expérience passée en matière de divulgation de données personnelles. Si Sheehan (2002) ne déduit aucune différence selon l'ancienneté ou l'utilisation qui est faite d'Internet (i.e. surf, lecture des emails), Jensen et al. (2005) montrent, quant à eux, l'absence d'effet de l'expérience d'achat sur le web. Ce résultat doit cependant être analysé avec précaution dans la mesure où la majorité de ces recherches a été menée auprès d'individus plus expérimentés que la moyenne. A contrario, Hann et al. (2003) invitent à penser que l'expérience passée en matière de dévoilement de soi influence l'appartenance

aux différents groupes. Ils montrent en effet que, sur leur échantillon américain, les « vendeurs de données » ont une plus grande expérience de partage de données que les deux autres groupes.

Le troisième facteur susceptible de discriminer les segments est d'ordre perceptuel ou attitudinal. Bien que la construction de la typologie laisse supposer l'existence d'un tel résultat, les conclusions issues des différents travaux ne vont pourtant pas toujours dans ce sens. Si Hann et al. (2003) disent que les « chercheurs de confort » sont moins préoccupés que les autres par l'utilisation de cookies[100], Cranor et al. (1999) suggèrent, pour leur part, que la sensibilité perçue des données est similaire, quel que soit le groupe d'appartenance.

Enfin, deux recherches abordent plus particulièrement l'aspect comportemental. Alors que Jensen et al. (2005) trouvent que l'appartenance à un groupe n'influence pas la probabilité de lire une charte de confidentialité, Sheehan (2002) montre que chaque segment répond aux demandes de données de manière significativement différente. Ainsi, les personnes les moins préoccupées acceptent deux fois plus souvent que les autres de s'enregistrer sur un site pour pouvoir obtenir des informations et fournissent moins fréquemment des données incomplètes. A l'opposé, les plus préoccupées ont une plus forte tendance à fournir des données erronées.

3 Méthodologie de l'étude

Nous présentons, tout d'abord, le mode de recueil des données (3.1) en évoquant notamment le développement du questionnaire (3.1.1) et la description de l'échantillon (3.1.2). Nous abordons ensuite le choix des variables à introduire (3.2.1) et de la méthode d'analyse (3.2.2).

3.1 Le recueil des données

3.1.1 Développement du questionnaire

L'objectif de cette étude était d'appréhender l'attitude des internautes français en matière de protection de la vie privée, d'évaluer les pratiques de transmission de données personnelles sur Internet et d'établir une typologie en la matière, dans une optique essentiellement descriptive. A la demande de la FING, une partie du questionnaire concernait le pseudonymat mais ne sera pas traitée dans le cadre de ce travail doctoral.

[100] Terme déjà défini au chapitre 1 comme étant un ensemble de fichiers envoyés par le serveur sur l'ordinateur de l'internaute de manière à capter des données personnelles et à suivre son comportement de navigation

Outre la réalisation de la typologie, nous nous intéressions donc aussi à plusieurs points :

- Quel est le comportement habituel des internautes français en matière de transmission de données personnelles sur Internet ?
- Quelle est leur attitude face aux pratiques de traitement et au respect de la vie privée ?
- Quel est le niveau de risque perçu en cas de transmission de données sur un site Internet ?
- En quoi la connaissance et la familiarité avec le site influencent les pratiques de transmission ?
- Quel est l'impact du contexte de collecte sur l'acceptation de fournir des données en ligne ?
- Quel est l'effet du type d'avantage proposé par le site en échange des données transmises ?

L'instrument de collecte développé s'inscrit dans ce cadrage. Il a été élaboré en étroite collaboration avec les responsables de la FING. Certains items souhaitant être intégrés par ces derniers ont ainsi été conservés, malgré leur absence de « scientificité » (formulation négative, deux items en un, item très long). Plusieurs allers-retours ont d'ailleurs été nécessaires pour aboutir à une version finale qui satisfasse les deux parties. Ce questionnaire a ensuite été pré-testé auprès d'une dizaine de personnes, afin de vérifier la clarté des questions et des instructions. Le questionnaire a ensuite été programmé par le webmaster de la FING puis implanté sur le serveur de l'association.

Le questionnaire final, administré en ligne, comprend sept écrans (ou pages) au total, dont le premier consacré à l'introduction et le dernier aux remerciements et à l'identification du répondant (cf. Annexe 6.2). Les pages 2 et 6 reprennent les questions permettant de caractériser les répondants selon leur utilisation d'Internet (ancienneté d'usage, temps consacré au surf sur le web et à la lecture de la messagerie, expérience d'achat en ligne) et leur profil sociodémographique (âge, sexe, pays, région, profession et niveau d'études).
Sept blocs d'items, correspondant aux pages 3 et 5[101], répondent plus particulièrement à nos objectifs de recherche. Le premier bloc concerne le comportement habituel de l'internaute face aux demandes d'informations personnelles des entreprises (Q9). Les deuxièmes et troisième blocs réunissent plusieurs mesures comportementales (en déclaratif, toutefois) : il s'agit de l'appréhension du nombre d'informations données (plus précisément refusées) selon le type de contexte dans lequel elles sont sollicitées d'une part (Q10) et le type d'interlocuteur susceptible de les demander (Q12) d'autre part (la notion de familiarité avec l'interlocuteur est également intégrée, en référence à Rogers 1996). Le quatrième et le cinquième bloc regroupent un ensemble de questions relatives aux comportements de protection (Q13) et aux

[101] La page 4 correspond aux questions sur le pseudonymat et ne sera donc pas détaillée ici

avantages recherchés en échange des données (Q14). Les deux derniers blocs concernent le risque perçu en mono item (Q16), suivi de questions sur l'attitude face au respect de la vie privée et sur le besoin d'information ressenti par l'internaute (Q22).
Le tableau fourni en Annexe 6.3 reprend les principaux items utilisés dans chacun des blocs.

Puisqu'il s'agit d'une étude exploratoire, essentiellement descriptive, visant à analyser les pratiques « déclarées » des internautes français en matière de transmission de données, nous avons repris les formats de questions utilisés habituellement dans ce type de recherches. Ainsi, concernant le comportement habituel de transmission (COMP, Q9), nous cherchions à évaluer la manière dont les participants ont l'habitude de répondre aux sollicitations de leurs données personnelles en ligne. Cinq comportements, identifiés à travers l'étude qualitative comme étant majoritairement utilisés par les internautes français étaient proposés : la collaboration (donner sa véritable identité), l'évitement (ne donner que le strict nécessaire), le détournement d'identité (fait de fournir l'identité d'autrui), le refus (ne pas répondre) et l'affrontement (mentir). Nous y avons ajouté, à la demande de la FING, un item correspondant à l'utilisation d'un pseudonyme, qui n'était pas ressorti de l'étude qualitative mais qui constitue une réponse possible. Les répondants devaient alors indiquer la fréquence (de 1 à 5, de jamais à toujours) avec laquelle ils adoptaient chacun de ces comportements.

L'influence du contexte (CONTEXTE, Q10) et du type d'interlocuteur (INTERLOC, Q12) sur la transmission d'informations personnelles a été mesurée, en croisant le type de données que le répondant refuserait de fournir (parmi 10 proposées), selon les circonstances dans lesquelles elles lui étaient demandées d'une part, le type d'organisme et le niveau de contact établi avec lui, d'autre part. 6 types de contextes (entrer sur un site, obtenir un document, discuter sur un forum, s'abonner à une newsletter, acheter en ligne et participer à un jeu-concours) et 5 types d'interlocuteurs (organisme public, entreprise réputée mais non familière, entreprise familière, entreprise peu connue, association) étaient ainsi proposés[102]. Les répondants devaient alors remplir une grille avec en colonne 10 données personnelles susceptibles de leur être demandées (email, nom, numéro de téléphone fixe et mobile, sexe, adresse postale, âge, situation familiale et professionnelle, numéro de carte bancaire)[103] et en ligne le type de contexte ou d'interlocuteur proposés. Les cases cochées signifiaient que la

[102] Le choix de ces contextes et interlocuteurs a été fait en collaboration avec la FING, sur la base des résultats issus de notre étude qualitative. Ils correspondent aux « éléments » les plus fréquemment rencontrés en ligne.
[103] Là encore, le choix de ces données a été réalisé en collaboration avec la FING, sur la base des résultats de notre étude qualitative. Elles correspondent également aux données les plus fréquemment demandées sur le web.

personne n'aurait pas fourni cette donnée dans ce contexte ou à ce type d'interlocuteur. Une grille avec beaucoup de cases cochées était donc le signe d'un comportement de méfiance.

La notion de protection (PROTEK, Q13) était intégrée à travers 4 items visant à mesurer la fréquence (de 1 à 5, de jamais à toujours) avec laquelle les répondants adoptaient différents comportements prouvant leur besoin de se protéger des invasions subies. Une analyse factorielle menée sur ces 4 items montre que les deux derniers (PROTEK 3 et 4) sont mal représentés (respectivement 0,334 et 0,341). Ils sont donc exclus des analyses ultérieures.
La portée de la compensation offerte par le site en échange des données (AVANT, Q14) a ensuite été évaluée à travers l'intention du répondant de fournir davantage d'informations pour chaque type de bénéfice possible. Dix avantages, issus de la revue de la littérature et des résultats de l'étude qualitative, étaient proposés. L'intention de fournir des données en échange de chacun de ces avantages a été mesurée en 4 points, de 1 (pas du tout) à 4 (oui certainement). Nous avons en effet volontairement supprimé le point neutre, de manière à obliger le répondant à se prononcer. Le risque perçu (RISK, Q16) était aussi mesuré par un item sur une échelle de Likert en 5 points (de 1 pas du tout importants à 5 très importants).

Enfin, les questions concernant l'attitude face au respect de la vie privée (RVP, 3 items) et le besoin d'information (INFO, 4 items) ont été mesurées sur des échelles de Likert en 5 points (de pas du tout d'accord à tout à fait d'accord). Une analyse factorielle menée sur ces 7 items montre que les deux derniers items de l'échelle RVP (RVP 2 et 3) sont très mal représentés (respectivement 0,386 et 0,272). Ils sont donc supprimés de l'analyse. Une nouvelle analyse factorielle sur les 5 items restants laisse apparaître deux facteurs : le premier regroupe l'ensemble des items correspondant au concept de besoin d'information (INFO) tandis que le second comprend uniquement l'item RVP 1.

La nature largement exploratoire de cette étude, l'objectif fixé (qui n'incluait aucunement le développement d'échelles de mesure) et le partenariat avec un organisme dépourvu de vocation académique expliquent en grande partie le faible nombre d'items dévolu à certaines échelles, que ce soit au départ (par exemple celle du risque perçu) ou à l'issue de l'analyse factorielle exploratoire (PROTEK et RVP). Certains des items ont ainsi été proposés à l'origine par la FING et éventuellement reformulés par nos soins. Ils ne sont donc pas issus de la littérature et n'avaient pas été pré-testés. En réalité, seule l'échelle mesurant le niveau d'information souhaité (INFO) a été reprise de la littérature anglo-saxonne et traduite en

français[104]. L'analyse factorielle exploratoire confirme d'ailleurs l'unidimensionnalité du concept sous-jacent et montre la validité de cette échelle (Alpha de Cronbach : 0,911).

3.1.2 Description de l'échantillon

Une fois développé puis pré-testé auprès d'une dizaine d'individus, le questionnaire a donc été administré en ligne auprès de 1364 internautes. Les répondants ont été sollicités par le biais d'annonces figurant sur les pages d'accueil des sites fing.org et secuser.com (la FING avait développé un partenariat avec ce site de manière à accroître et diversifier l'échantillon) et sur les newsletters envoyées par ces deux sites à leurs abonnés. Ne connaissant pas le nombre de personnes ayant été soumis à ces annonces, il nous est impossible de calculer un taux de participation à cette enquête. Nous pouvons néanmoins supposer que, cette étude ayant été réalisée sous l'égide de la FING, un nombre important de membres de cette association et/ou d'abonnés à sa newsletter, y ont participé.

Le tableau 6.2 ci-après résume les principales caractéristiques des personnes interrogées. On peut remarquer qu'un grand nombre de répondants (79%) habite en France, le reste (20%) étant issu d'un pays francophone. Dès lors, même si l'objectif initial était d'établir la typologie auprès des Français, nous décidons d'inclure les francophones (hors France) dans l'analyse afin de mesurer l'incidence de la nationalité sur la classification obtenue. En ce qui concerne l'utilisation d'Internet, les répondants disposent en moyenne de 4 adresses email par individu. 83% d'entre eux ont déjà acheté en ligne, un quart achetant une à plusieurs fois par mois, les autres (58%) se contentant d'une à plusieurs fois par an.

Tableau 6.2 Description de l'échantillon interrogé pour l'étude 2

Variables descriptives	N = 1364
Sexe	Hommes 78%
Age	20-40 ans : 35% 40-60 ans : 42%
CSP	Ingénieurs/techniciens (14%) Commerciaux/employés/ouvriers(15%) Inactifs (24%)
Niveau d'études	Niveau bac+2 : 30% Niveau bac+4 : 29%
Nationalité	France : 80% Pays francophone : 19,2%
Durée moyenne d'utilisation d'Internet à titre professionnel	5 ans

[104] Il s'agit de l'échelle de Farag et Krishnan (2003) dont nous reparlerons au chapitre 9.

Durée moyenne d'utilisation d'Internet à titre privé	6 ans
Temps consacré à la consultation des mails	1 à 2h / jour pour 42%
Temps consacré au surf	1 à 2h / jour pour 72%
Expérience d'achat sur Internet	83%
Paiement en ligne	90%

Cet échantillon, tel qu'il apparaît à la lecture de ces chiffres, est donc beaucoup plus masculin que la moyenne des internautes français (53% d'hommes, source Ipsos 2004). Il comporte également moins de jeunes de moins de 20 ans et plus de 40-60 ans. La répartition est par contre comparable en ce qui concerne la région d'habitation (Paris/Province) et la catégorie socioprofessionnelle (CSP +, CSP – et inactifs). A contrario, si près de la moitié des internautes français ont déjà acheté en ligne (Source Ipsos et Médiametrie 2005), ils sont plus de 80% dans notre échantillon, ce qui indique le côté éminemment technophile de nos répondants. Ceci s'explique principalement par notre source de recrutement qui concerne des individus intéressés par les applications de l'Internet et par les problèmes de sécurité en ligne. Le niveau d'études élevé de notre échantillon, couplé à une forte expérience de l'achat sur Internet, est conforme aux résultats établis par Westin (1998) selon lesquels les acheteurs en ligne ont un niveau d'études supérieur à celui des non acheteurs.

Dans l'ensemble, notre échantillon n'est donc pas statistiquement représentatif de la population des internautes français puisqu'il est composé d'utilisateurs aguerris de l'Internet. Ces derniers représentent cependant de possibles leaders d'opinion dans le domaine des NTIC, dont il est important de comprendre la position en matière de traitement de données personnelles afin d'évaluer l'impact de ce phénomène sur la population Internet dans son ensemble et sur les internautes du futur en particulier. En effet, comme le nombre d'internautes augmente de manière exponentielle, de plus en plus d'individus se trouveront dans la situation de ceux faisant partie de notre échantillon actuel dans les 10 ans à venir.
Par ailleurs, nos résultats révèlent que 85% des répondants estiment qu'une entreprise ne peut pas utiliser les données de ses clients selon son gré. S'il ne nous est pas possible de comparer ce chiffre avec celui de l'ensemble de la population française (aucune donnée n'étant disponible de ce point de vue), nous remarquons toutefois qu'il est comparable aux résultats trouvés par plusieurs auteurs américains sur le sujet. Ainsi, Westin (1998) trouve que 81% des internautes américains sont concernés par les menaces pesant sur l'utilisation de leurs données personnelles en ligne, tandis que Cranor et al. (1999) estiment ce chiffre à 87%. On peut donc imaginer que les internautes interrogés aient des attitudes comparables à celles des internautes

américains à propos de la collecte de données. Etudier leur point de vue et leurs pratiques et les comparer aux résultats déjà établis dans la littérature permettront donc de dresser un premier bilan de la situation française et de faire des projections sur l'évolution future.

3.2 La méthode d'analyse

Nous présentons et justifions tout d'abord le choix des variables sur lesquelles ont porté notre classification (3.2.1), avant d'expliquer la méthode d'analyse retenue (3.2.2).

3.2.1 Le choix des variables à introduire

Comme nous l'avons déjà indiqué, l'objectif majeur de cette étude était de réaliser une typologie des internautes français sur la base de leur comportement face à une sollicitation de données personnelles en ligne. L'analyse typologique a pour but de classer les individus dans différents groupes, selon leur similarité/dissimilarité dans l'évaluation du phénomène étudié. Le principe de base est de regrouper les objets étudiés (ici, les individus) en classes homogènes, de telle sorte que les objets à l'intérieur d'une même classe soient très semblables et que ceux appartenant à des classes différentes soient très dissemblables (Thiétart et al. 2003). Entreprendre une telle démarche suppose implicitement qu'un système de classification existe au sein de la population étudiée, le but de l'analyse étant de le mettre à jour. La littérature nous permettant de supposer l'existence d'une telle hétérogénéité, notre objectif était donc de vérifier cette hypothèse, dans le cadre de la population française.

Dans toute analyse typologique, plusieurs étapes doivent être respectées. La première d'entre-elles consiste à choisir les variables à inclure dans l'analyse (Malhotra 2004). Plutôt que d'établir la typologie sur la base de questions générales relatives à la préoccupation pour le Respect de la Vie Privée, comme l'ont fait un grand nombre d'auteurs jusqu'ici, il nous a semblé préférable de sélectionner des réponses plus précises et circonstanciées. On peut d'ailleurs souligner que d'autres auteurs avant nous ont basé leur typologie sur des intentions comportementales plutôt que sur des attitudes. Ce fut notamment le cas de Cranor et al. (1999) et de Sheehan (2002). Nous avons, pour notre part, choisi de réaliser la typologie sur la base des avantages recherchés par l'internaute en échange de ses données personnelles. Ce type de variable offre en effet l'avantage de correspondre à l'esprit de la typologie établie par Westin (1996), et donc de faciliter les comparaisons ultérieures. On peut ainsi supposer que les « pragmatiques » attendent des bénéfices en échange de leurs données, tandis que les

« non préoccupés » refuseront de se dévoiler quel que soit le type d'avantage proposé. De plus, ce choix répond aussi à un principe de base établi au sein de la littérature selon lequel le comportement de partage d'informations dépend - essentiellement - des bénéfices proposés par l'entreprise qui sollicite les données. Les recherches dans ce domaine, confirmées par les résultats de notre étude qualitative, suggèrent en effet que les individus évaluent les coûts et bénéfices de la divulgation avant de décider de transmettre ou non des données. Ce système de « compensation » est également conforme à la théorie de l'échange social (Taylor et al. 1969 ; Taylor et Altman 1975), qui a déjà été présentée au chapitre précédent et qui sera à nouveau évoquée pour établir notre cadre conceptuel.

Sur les 10 items proposés dans le questionnaire, 9 d'entre eux ont été inclus dans l'analyse, celui portant sur le gain en termes de service étant finalement exclu. Après réflexion, il est en effet apparu que la formulation de cet item portait à confusion parce qu'il contenait un double bénéfice à savoir « bénéficier d'un meilleur service » et « recevoir des informations qui m'intéressent ». Nous avons donc mené l'analyse sur les neuf items restants, représentant les avantages suivants : gain de temps, baisse de prix, confidentialité, plaisir, personnalisation, cadeaux, argent, sécurité et fait de rendre service[105]. Ce sont donc aussi bien des bénéfices utilitaires (baisse de prix par exemple) que des bénéfices hédoniques (par exemple, se faire plaisir).

3.2.2 Le choix de la méthode de classification

Avant de lancer l'analyse, nous avons procédé à la préparation des données. Nous avons tout d'abord choisi d'exclure les valeurs manquantes, de manière à ne pas créer de classes artificiellement. L'analyse univariée menée sur les 9 items montrant l'absence de points aberrants, toutes les observations sont incluses dans l'analyse. Toutefois, à ce stade du questionnaire, 347 individus ne sont plus présents ou n'ont pas répondu à l'ensemble de ces items. La typologie est donc effectuée sur un effectif valide de 1017 observations. Enfin, les variables étant toutes mesurées en 4 points, aucune standardisation n'est apparue nécessaire.

Dans toute analyse typologique, la sélection de la méthode de classification tient une place primordiale. Il s'agit en effet d'un choix à la fois délicat et fondamental. Nous avons suivi une démarche classique – recommandée dans de nombreux ouvrages[106] - consistant à réaliser une

[105] Les « bénéfices » confidentialité et sécurité sont conservés, bien qu'il s'agisse plus, à proprement parler, de garanties que de réels avantages, de par leur caractère incitatif majeur.

[106] Nous pouvons citer, entre autres, les ouvrages d'Evard et al. (2003), de Thietart et al. (2003) et de Malhotra (2004).

première classification de type hiérarchique - avec recours à la méthode de Ward, reconnue comme étant l'une des plus performantes (Miligan 1980) - suivie d'une analyse non hiérarchique, par le biais des nuées dynamiques. L'objectif majeur de l'analyse hiérarchique est de déterminer le nombre de classes pertinent. Ce dernier peut être identifié visuellement, par l'examen de l'arbre hiérarchique (critère du plus grand saut ou de la distance d'agrégation la plus grande) mais aussi, par l'examen de la chaîne d'agrégation, de laquelle on extrait la variance intra-classe à chaque étape de regroupement.

Dans le cas présent, s'agissant d'un échantillon important (1017 individus), l'arbre hiérarchique est peu lisible ; l'examen de la chaîne d'agrégation est donc privilégié. La colonne coefficient (correspondant à la variance intra-classe) suggère deux niveaux possibles de partition (correspondant à l'écart le plus important avec l'étape de regroupement suivante) : l'un à deux classes, l'autre à 4 classes (cf. tableau 6.3). De plus, l'examen de la représentation graphique du rapport de la variance intraclasse sur la variance interclasse (cf. figure 6.1) laisse bien apparaître 2 coudes : l'un au niveau de la partition en 2 groupes (courbure brutale), l'autre pour 4 groupes (après ce point, le rapport de variance se stabilise). La partition en 2 classes est peu intéressante car elle conduit à une représentation binaire du phénomène (ceux qui donnent des données, ceux qui n'en donnent pas). Nous retenons donc la partition en 4 classes qui se justifie, en outre, par l'affaiblissement du pourcentage de variabilité (Bozzo 2000). Au-delà de 4 classes, la variabilité se stabilise en effet autour de 2%, ce qui indique que la partition en 4 classes est acceptable. De plus, une répartition en 4 groupes a déjà été identifiée à plusieurs reprises dans la littérature (Spiekermann et al. 2001, Sheehan 2002), notamment s'agissant de classer des populations d'internautes.

Tableau 6.3 Chaîne d'agrégation (derniers niveaux de regroupement)

Etape d'agrégation	Coefficient	Variance intra-classe / variance interclasse	% Variabilité	Nombre de classes
1008	4046,35	45,47%		9
1009	4218,51	47,40%	1,93%	8
1010	4395,00	49,39%	1,98%	7
1011	4573,67	51,40%	2,01%	6
1012	4796,87	53,90%	2,51%	5
1013	5086,09	57,15%	3,25%	4
1014	5678,29	63,81%	6,65%	3
1015	6320,64	71,03%	7,22%	2
1016	8898,99	100,00%	28,97%	1

Figure 6.1 Variance intraclasse / variance interclasse en fonction du nombre de classes

Sur cette base, une analyse non hiérarchique, par le biais des nuées dynamiques, est menée sur l'ensemble de l'échantillon et 4 classes sont demandées[107]. Afin de conforter les résultats, les analyses sont reconduites sur deux sous-parties de l'échantillon (représentant 50% chacune), constituées aléatoirement[108]. Les résultats obtenus sont assez proches (cf. Annexe 6.4), tant du point de vue de la répartition des effectifs (pas plus de 4 points d'écart) que des centres de classes finaux (seuls 4 centres sont différents sur les 108 obtenus). Par la suite, nous présenterons donc uniquement les résultats obtenus sur la totalité de l'échantillon.

Une fois la typologie établie, nous souhaitions également décrire le profil des classes, à l'aide de l'ensemble des variables descriptives mesurées dans le questionnaire, qu'elles soient de nature sociodémographique, expérientielle, attitudinale ou comportementale. Ces réflexions nous ont conduits à proposer le cadrage suivant pour mener à bien notre approche typologique (cf. figure 6.2) :

[107] Par précaution, nous menons aussi l'analyse non hiérarchique avec 3 et 5 classes demandées. Les résultats obtenus étant systématiquement moins « bons » que ceux avec 4 classes (effectifs très déséquilibrés), nous confortons ainsi la typologie en 4 classes, lui apportant une validité supplémentaire.

[108] Une autre méthode (i.e. « Two Step Cluster ») que nous avons également utilisée pour conforter le résultat d'une typologie en 4 classes sera détaillée à la fin du paragraphe 4.1.

Figure 6.2 Cadre d'analyse en vue d'une typologie des internautes français

Nous présentons maintenant les résultats obtenus, tant du point de vue de la classification (4.1) que de la description des classes (4.2).

4 Résultats de l'étude

Nous exposerons tout d'abord la typologie obtenue à l'issue de l'analyse par les nuées dynamiques avant de décrire le profil des classes en termes de signalétique, attitudes (surtout le risque perçu) et comportements (notamment comportement habituel de partage d'informations, informations délivrées selon le contexte et le type d'interlocuteurs).

4.1 La typologie obtenue

La typologie en 4 classes obtenue apparaît satisfaisante car elle permet d'obtenir des résultats, à la fois équilibrés en termes de répartition des effectifs, et stimulants en termes d'interprétation. La répartition des individus dans chaque classe se trouve au tableau 6.4. On remarque que la classe la moins importante comporte près de 20% de l'effectif tandis que la plus importante n'en contient que 31,6%. Contrairement à certaines typologies relevées dans la littérature, on constate donc ici une répartition « homogène » dans les différentes classes. De plus, en s'appuyant sur la description des classes (cf. tableau 6.5), nous avons nommé chacune de ces classes : les désintéressés, les réticents, les négociateurs et les bienveillants.

Tableau 6.4 Répartition des internautes français en 4 classes

Classe	Effectif	Part (%)	Interprétation
1	322	31,6	Les désintéressés
2	244	24,0	Les réticents
3	249	24,5	Les négociateurs
4	202	19,9	Les bienveillants

Tableau 6.5 Tableau de bord des classes

Classe			Avant. temps	Avant. prix	Avant. confid.	Avant. plaisir	Avant. perso	Avant. cadeau	Avant. argent	Avant. sécu.	Avant. service
1	Désintéressés	Moy	2,31	1,99	3,46	1,56	1,94	1,64	1,46	3,42	2,16
		N	322	322	322	322	322	322	322	322	322
		Ecart-type	,955	,818	,674	,709	,747	,606	,626	,666	,877
2	Réticents	Moy	1,32	1,30	1,93	1,07	1,29	1,20	1,13	2,22	1,27
		N	244	244	244	244	244	244	244	244	244
		Ecart-type	,663	,533	1,002	,286	,566	,450	,415	1,058	,631
3	Négociateurs	Moy	2,15	3,08	3,34	1,62	2,97	2,86	2,28	3,40	1,60
		N	249	249	249	249	249	249	249	249	249
		Ecart-type	1,015	,640	,735	,763	,588	,666	,946	,745	,695
4	Bienveillants	Moy	2,81	3,29	3,61	2,68	3,10	3,25	3,00	3,71	3,06
		N	202	202	202	202	202	202	202	202	202
		Ecart-type	,861	,620	,654	,828	,647	,621	,941	,543	,713
Total		Moy	2,13	2,35	3,09	1,68	2,27	2,15	1,89	3,18	1,99
		N	1017	1017	1017	1017	1017	1017	1017	1017	1017
		Ecart-type	1,026	1,032	1,018	,867	,970	1,005	1,019	,953	,978

On peut donner du sens à ces quatre classes de la manière suivante.

Les Désintéressés (classe 1) sont les plus importants en terme d'effectif (31,6%). Il s'agit d'individus prêts à partager leurs données personnelles, mais uniquement si cela ne les implique pas de façon trop importante. Ils demandent avant tout sécurité et confidentialité et si, à l'occasion, ils peuvent « rendre service » et gagner du temps, cela leur convient. Ils ne sont donc pas dans une optique d'échange d'information intensive.

Les Réticents (classe 2, 24%) paraissent avoir une attitude ferme et définitive. Ils obtiennent les notes les plus faibles sur tous les avantages listés. Aucun bénéfice, parmi ceux proposés, ne les incite à dévoiler davantage de données. Leur politique de partage d'informations sera donc un comportement systématique d'évitement. Les garanties affichées par un site en

matière de sécurité et de confidentialité ont alors peu de chance de les influencer puisque leur position de principe est de ne pas délivrer d'information.

Les Négociateurs (classe 3, 24,5%) adoptent une position de donnant-donnant. Ils sont prêts à fournir des données personnelles pour des avantages ponctuels comme : un meilleur prix, un cadeau, une meilleure personnalisation. Ils ne délivrent pas systématiquement leurs données personnelles mais ont pris conscience du fait qu'elles représentent une valeur d'échange qu'ils sont enclins à faire jouer lorsque la transaction proposée les intéresse. Un ciblage pertinent de leurs attentes peut donc permettre d'obtenir de plus amples informations de leur part.

Les Bienveillants (classe 4, 19,9%) sont caractérisés par des notes élevées sur tous les critères. Tous les avantages proposés sont donc susceptibles de les faire délivrer des données personnelles, y compris le sentiment de se faire plaisir et de se sentir utiles (rendre service). Obtenir des garanties et des avantages (prix, cadeaux, etc.) serait donc un élément incitatif majeur, mais pas forcément décisif, puisqu'ils semblent délivrer assez facilement leurs données personnelles, quelles que soient les circonstances.

Cette répartition peut être comparée aux conclusions issues de la littérature. On découvre quelques similitudes avec les résultats trouvés précédemment par certains auteurs, notamment Spiekerman et al. (2001) et Sheehan (2002) qui obtenaient eux aussi 4 groupes. Ainsi, les « fondamentalistes » et les « alarmés » identifiés respectivement par ces chercheurs correspondent à nos « réticents » tandis que les « non préoccupés » peuvent être assimilés à nos « bienveillants ». Certaines de nos classes peuvent aussi se rapprocher de celles obtenues par des chercheurs n'ayant identifié que 3 groupes. Ainsi, les « convenience seekers » trouvés par Hann et al. (2003) correspondent à nos « Désintéressés », à la différence près que ces derniers sont également attentifs aux mesures de protection de la vie privée. Si les catégories extrêmes (fondamentalistes-réticents, d'une part et non préoccupés-bienveillants, d'autre part) semblent relativement stables d'un auteur à l'autre, la différence s'établit donc sur les catégories intermédiaires. La nouveauté - et l'originalité - de la typologie trouvée ici est de distinguer les individus prêts à fournir des données si les conditions de sécurité sont respectées (les désintéressés) de ceux qui sont dans une optique d'échange (les négociateurs). Cela revient alors à scinder la catégorie des « pragmatiques », identifiée par un grand nombre d'auteurs, en deux, selon l'aspect sur lequel l'individu se base pour prendre sa décision (sécuritaire ou utilitaire). Il se pourrait donc que la typologie en 3 classes, établie à l'origine par Westin (1996), trouve ses limites quand il s'agit de catégoriser des populations d'internautes. Le chiffre de 4 classes, déjà avancé par Sheehan (2002), se confirmerait ici, bien qu'il concerne une population différente (internautes américains vs français).

Afin d'apporter une validité supplémentaire à cette classification en 4 groupes, nous décidons de mener une nouvelle analyse typologique en recourant à la procédure « TwoStep Cluster » proposée par le logiciel SPSS. Cette technique a été conçue pour révéler des groupements naturels au sein d'un ensemble de données. De plus, en comparant les valeurs d'un critère de classification (au choix, le BIC ou l'AIC[109]) dans les différentes solutions de classification, elle permet de déterminer le nombre optimal de classes.

Nous menons cette analyse sur les 9 items d'avantages recherchés utilisés précédemment (cf. paragraphe 3.2.1 et tableau 6.5) et auprès des mêmes 1017 individus (ensemble des observations valides), en prenant comme mesure de distance le log-vraisemblance et comme critère de classification le BIC[110]. Le modèle à retenir est celui pour lequel les valeurs du BIC sont les plus faibles, ce qui est le cas pour les modèles à 5 et 4 classes, respectivement. Nous choisissons alors de comparer les profils de classes pour ces deux modèles (cf. Annexe 6.5).

Le modèle à 5 classes fait apparaître des groupes relativement équilibrés, représentant entre 16% et 26% de l'effectif. Toutefois, la description des classes sur les avantages recherchés indique que les groupes 3 et 5 (représentant respectivement 17,7% et 19,7% de l'effectif) sont très proches l'un de l'autre, hormis pour les avantages « cadeau » et « argent » pour lesquels le groupe 3 se montre moins favorable. La scission de ces deux groupes ne semble donc pas forcément très opportune, de par leur « proximité » sur les avantages recherchés restants.

Le modèle en 4 classes fait également apparaître des groupes équilibrés, représentant entre 21% (pour les groupes 2 et 4) et 28% (pour les groupes 1 et 3) de l'effectif. De plus, ces groupes sont bien différenciés en ce qui concerne leur attitude face aux avantages proposés (ce qui n'était pas forcément le cas pour la typologie en 5 classes). Enfin, ils peuvent aisément se rapprocher des 4 classes identifiées précédemment sur la base des nuées dynamiques. Les classes 1 et 2 correspondent ainsi respectivement aux « Réticents » et « Désintéressés », tandis que les classes 3 et 4 s'apparentent aux « Négociateurs » et « Bienveillants ».

Ce résultat nous conforte alors dans l'idée d'une typologie en 4 classes telle que celle trouvée.

La validité de la typologie en 4 classes ayant été établie, voyons à présent s'il est possible d'établir un profil des individus appartenant à ces différentes classes.

[109] BIC pour Critère Bayésien de Schwarz et AIC pour Critère d'Information Akaike.

[110] Par précaution, nous menons aussi l'analyse en ayant recours au critère AIC. Les résultats obtenus sont sensiblement identiques, quel que soit le critère (BIC ou AIC) utilisé.

4.2 La description des classes

Etablir le profil des classes consiste à distinguer les variables susceptibles de discriminer l'appartenance aux groupes identifiés. Nous avons procédé, pour ce faire, à deux catégories d'analyses, selon le type de variables étudiées : des différences de moyenne pour les variables métriques et/ou ordinales, des tests de chi-deux pour les variables catégorielles.

Au plan signalétique, les individus présentent des profils distincts en termes de nationalité, de niveau d'étude, d'expérience de surf, d'achat, et d'usage d'Internet (cf. Tableau 6.6).
Concernant les aspects purement sociodémographiques, il est remarquable de n'observer aucune différence selon le sexe et l'âge des répondants, deux variables pour lesquelles d'autres études ont pu relever des écarts significatifs, en matière de sensibilité au respect de la vie privée (Cockroft et Clutterbuck 2001 ; Dommeyer et Gross 2003). Il n'apparaît également aucune différence selon le type d'habitat (urbain vs rural) ou la profession des individus. Les deux variables significatives concernent, et la nationalité (France versus pays francophone) et, le niveau d'étude (après regroupement : inférieur et supérieur à bac+2).

Tableau 6.6 Tests de Chi² et différences de moyenne (DM) sur les variables signalétiques

Variables signalétiques	Test	Sig.
Sexe (H, F)	Chi²	ns
Age	Chi²	ns
Habitat (urbain, rural)	Chi²	ns
CSP (Profession)	Chi²	ns
Niveau d'étude (<bac+2 ; >bac+2)	Chi²	.01
Pays (France ; francophone)	Chi²	.01
Temps consacré au surf	Chi²	.05
Ancienneté d'usage d'Internet à titre professionnel	DM	.01
Ancienneté d'usage d'Internet à titre privé	DM	.01
Expérience d'achat en ligne (non, oui)	Chi²	.01

Plus précisément, les répondants de pays francophones, qui représentent 19.2% de l'effectif, sont particulièrement présents dans les classes 1 et 2 (Désintéressés et Réticents) avec respectivement 20% et 29.4%. Les Français sont au contraire surreprésentés dans les classes 3 et 4 (respectivement 85,2% et 86,8%) et seraient donc plutôt favorables à la divulgation de leurs données personnelles, à condition que des avantages leur soient concédés en échange. En termes de niveau d'étude, les répondants de niveau supérieur à bac +2 (50.2%) sont

particulièrement présents dans la classe des Désintéressés (57.4%) et sous-représentés (38.4%) dans la classe des Bienveillants, alors que les répondants dont le niveau d'études est inférieur à bac +2 (au total 49.8%) sont surreprésentés dans la classe des Bienveillants (61.6%). Les individus peu éduqués sont donc plutôt enclins à délivrer des informations, ce qui confirme les résultats établis précédemment par Sheehan (2002).

Concernant les pratiques de navigation, on constate que la majorité des répondants surfe quotidiennement sur Internet (71.8%). Les Désintéressés et les Réticents surfent légèrement moins (respectivement 69.6% et 69.6%) que les Négociateurs et les Bienveillants. Autrement dit, plus les individus consacrent une part importante de leur temps à surfer, plus ils acceptent de délivrer des informations. Sur le plan de l'ancienneté d'utilisation (à titre personnel ou professionnel), les résultats montrent que les internautes les plus anciens appartiennent aux classes 1 et 2 (Désintéressés et Réticents). Leur expérience les a-t-elle conduits à être blasés, pour les premiers, ou au contraire méfiants, pour les seconds, suite à certaines mésaventures ? Il serait intéressant de creuser de manière plus approfondie les motivations de ces positions. Enfin, concernant l'expérience d'achat en ligne, une grande majorité de l'échantillon a déjà réalisé des achats, à 85.2%. Ce sont principalement les Négociateurs (classe 3) qui, à 92%, sont les plus gros acheteurs en ligne. Ceux qui n'ont jamais acheté en ligne semblent donc plus « réticents » à délivrer des informations personnelles. Ces résultats démontrent que dans l'ensemble l'expérience et l'utilisation qui est faite d'Internet est fortement discriminante, ce qui vient contredire les résultats établis par Sheehan (2002) et Jensen et al. (2005).

En résumé, si les variables expérientielles semblent en mesure de différencier l'appartenance aux classes, tel n'est pas le cas des données sociodémographiques (hormis pour le pays et le niveau d'études). Rien ne dit cependant que les résultats seraient identiques si l'on croisait certaines de ces variables signalétiques. Car, si la répartition hommes/femmes est similaire d'une classe à l'autre, peut-être trouverions-nous que les « réticents » sont surtout des femmes âgées de plus de 50 ans et habitant une grande ville ! Pour ce faire, il faudrait mener une analyse multi-variée qui permette de croiser ces variables sociodémographiques et examiner ensuite les proximités des nouvelles variables positionnées avec la typologie obtenue (les 4 classes)[111]. Une telle analyse constitue donc une voie de recherche d'autant plus prometteuse que les résultats déduits seraient beaucoup plus riches que ceux obtenus ici. Nous reviendrons sur cette possibilité dans la discussion finale de ce document.

[111] Plus précisément, il s'agirait de réaliser une analyse des correspondances multiples (ACM), de récupérer les coordonnées factorielles des variables ainsi positionnées et de mener une analyse hiérarchique afin de détecter les éventuelles proximités avec l'appartenance aux classes.

L'examen des attitudes (risque perçu, attitude face au RVP et besoin d'information) offre des résultats plus contrastés. Seul le risque apparaît, en effet, véritablement discriminer les classes (F = 4,987 ; p = 0,002). Dans ce domaine, les classes les plus sensibles sont les Réticents (Classe 2) et les Négociateurs (Classe 3), pour des raisons naturellement différentes : les premiers parce qu'il s'agit d'un principe (ne pas trop se dévoiler) ; les seconds parce que l'information est une monnaie d'échange qui ne se dévoile pas sans une raison valable. Concernant l'attitude face au RVP, seul un item (RVP1) a été conservé suite à l'analyse factorielle exploratoire. Cet item ne montre pas de différence significative entre les classes (F = 2,080 ; p = 0,102). Enfin, pour ce qui est du besoin d'information (INFO), l'analyse de différence de moyenne est menée sur chacun des 4 items et sur le score global (par addition du score donné à chaque item). Seul le second item (INFO2) donne un résultat significatif (F = 2,630 ; p = 0,049), les individus des classes 1 et 2 (Désintéressés et Réticents) étant les plus désireux d'être informés de la durée de conservation de leurs données.

L'examen du comportement (i.e. de protection, habituel de réponse, de refus selon le type d'interlocuteur et le contexte de sollicitation) fait ressortir d'intéressantes différences.
Suite à l'élimination de 2 items à l'issue de l'analyse factorielle exploratoire, le comportement de protection (PROTEK) s'exprime par le fait que l'internaute ait déjà demandé à une entreprise de retirer son nom de son fichier (PROTEK1) et de ne pas transmettre ses données à d'autres entreprises (PROTEK2). Seul ce second item apparaît discriminer l'appartenance aux classes (F = 3,353 ; p = 0,018 contre F = 1,866 ; p = 0,134 pour PROTEK1). Dans ce cas, les internautes désireux de se protéger appartiennent aux Désintéressés (qui recherchent des garanties en termes de sécurité et confidentialité) et aux Réticents (qui n'apprécient guère de se dévoiler). Les résultats (cf. tableau 6.7) sont donc logiques.

Ces résultats sont confortés par ceux concernant les comportements habituellement adoptés par les répondants en matière de transmission de données (COMP). Les Négociateurs et les Bienveillants donnent volontiers leur véritable identité alors que les Désintéressés et les Réticents ne donnent que le nécessaire, « tapent n'importe quoi » ou ne répondent pas. Ceci suggère que le média (Internet) aurait un impact neutre sur le comportement. Il n'existe pas, en revanche, de différence significative en ce qui concerne l'utilisation d'un pseudonyme ou le fait d'utiliser l'identité de quelqu'un d'autre, signe qu'il s'agit là de comportements marginaux, d'une part et qui peuvent être adoptés par tous les individus, quelle que soit leur appartenance aux classes, d'autre part. Là encore, les résultats trouvés (cf. tableau 6.7) sont donc conformes aux attentes et confortent ceux obtenus préalablement par Sheehan (2002).

Tableau 6.7 Test de moyenne sur le comportement de protection et le comportement habituel

Classes			PROTEK 2	donne sa véritable identité	ne donne que le nécessaire	tape n'importe quoi	ne répond pas
1	Désintéressés	Moy	3,80	2,85	4,46	2,02	2,64
		N	322	314	308	299	303
		Ecart-type	1,363	1,182	,844	1,131	1,039
2	Réticents	Moy	3,69	2,54	4,39	2,24	2,76
		N	239	234	233	225	222
		Ecart-type	1,472	1,201	1,066	1,253	1,143
3	Négociateurs	Moy	3,58	2,89	4,37	1,97	2,52
		N	249	243	236	230	221
		Ecart-type	1,315	1,130	,863	1,098	1,034
4	Bienveillants	Moy	3,43	3,18	4,14	1,78	2,22
		N	202	197	193	189	184
		Ecart-type	1,3	1,203	1,064	1,032	1,054
Total		Moy	4,09	2,85	4,36	2,01	2,56
		N	749	988	970	943	930
		Ecart-type	,925	1,196	,957	1,143	1,081
Sig.			,002	,000	,003	,001	,000

On peut voir également des différences entre les classes, en ce qui concerne le nombre d'informations délivrées aux différents interlocuteurs possibles (INTERLOC, cf. tableau 6.8). Quels que soient le type d'interlocuteur et le niveau de familiarité avec celui-ci (organisme public, entreprise réputée non familière / familière / peu connue ou association), la hiérarchie est toujours la même. Les Bienveillants sont ceux qui fournissent le plus de données, suivis par les Négociateurs et les Désintéressés. Les Réticents sont, de loin, ceux qui en délivrent le moins, quel que soit l'interlocuteur. Il est intéressant de remarquer, au-delà de cette hiérarchie, que les replis les plus importants concernent les requêtes émises par les entreprises peu connues (5,73) mais aussi par celles qui, bien que connues, n'ont pas de lien antérieur avec l'internaute (entreprise réputée mais non familière) (4,35). Par contre, être sollicité par une firme avec laquelle on a des contacts réguliers réduit ce comportement de repli (2,82). On peut donc conclure, qu'en matière de respect de la vie privée, la familiarité avec l'entreprise et/ou le site pourrait être un facteur plus incitatif que la seule réputation ou notoriété.

Tableau 6.8 Test de moyenne sur le comportement de refus selon le type d'interlocuteur (nombre de données refusées par interlocuteur parmi les 10 proposées)

Classes			organisme public	entreprise réputée non familière	entreprise familière	entreprise peu connue	association
1	Désintéressés	Moy	2,53	4,52	3,07	6,02	4,63
		N	322	322	322	322	322
		Ecart-type	2,210	2,855	2,401	2,944	2,927
2	Réticents	Moy	3,14	5,25	3,75	6,70	5,37
		N	244	244	244	244	244
		Ecart-type	2,509	2,911	2,654	2,911	3,027
3	Négociateurs	Moy	2,36	4,02	2,44	5,44	4,43
		N	249	249	249	249	249
		Ecart-type	2,035	2,817	1,919	2,941	2,812
4	Bienveillants	Moy	2,19	3,38	1,78	4,44	3,45
		N	202	202	202	202	202
		Ecart-type	2,081	2,821	1,702	2,950	2,761
Total		Moy	2,57	4,35	2,82	5,73	4,53
		N	1017	1017	1017	1017	1017
		Ecart-type	2,244	2,920	2,336	3,033	2,957
Sig.			,000	,000	,000	,000	,000

Ce type de résultat se retrouve s'agissant du comportement de refus induit par le contexte dans lequel les données sont sollicitées (CONTEXTE, cf. tableau 6.9). Là encore, les individus des différentes classes ont des comportements de refus distincts (tests de différences de moyennes significatifs pour tous les contextes) et la hiérarchie est identique à celle établie pour le type d'interlocuteurs (Bienveillants, Négociateurs, Désintéressés et Réticents). On remarque également que les refus sont plus systématiques s'il faut fournir des données pour entrer sur un site (6,38) et discuter sur un forum (5,68). Les replis sont par contre moins marqués quand il s'agit d'acheter sur un site (3,14) ou d'obtenir un document ou une information (4,67). Les contextes d'abonnement à une newsletter et de participation à un jeu-concours offrent une situation intermédiaire, dans laquelle les individus ne donneraient en moyenne que la moitié des informations personnelles demandées. Le contexte de sollicitation s'avère donc discriminant des comportements de transmission de données en ligne[112].

[112] Pour aller plus loin et, à la fois, conforter les résultats, et accroître leur pouvoir explicatif, ces trois dernières analyses (i.e. tableaux 6.7, 6.8 et 6.9) auraient pu faire l'objet d'une méthodologie plus « sophistiquée », en recourant à la méthode de régression par classes latentes. L'objectif aurait alors été de construire les groupes de manière à maximiser, au sein de chaque classe, le pouvoir explicatif des variables comportementales (i.e. comportement de protection et de refus notamment). Le recours à une typologie explicative de ce type constitue une voie de recherche prometteuse.

Tableau 6.9 Test de moyenne sur le comportement de refus selon le type de contexte (nombre de données refusées par contexte parmi les 10 proposées)

Classes			entrer sur un site	obtenir document / information	discuter sur un forum	s'abonner à une news	acheter en ligne	participer à un jeu-concours
1	Désintéressés	N	322	322	322	322	322	322
		Moyenne	6,59	4,85	5,90	5,21	3,31	5,39
		Ecart-type	2,71	2,47	2,73	2,53	2,51	3,05
2	Réticents	N	244	244	244	244	244	244
		Moyenne	6,91	5,30	6,05	5,78	3,89	5,87
		Ecart-type	2,80	2,71	3,02	2,84	2,57	3,35
3	Négociateurs	N	249	249	249	249	249	249
		Moyenne	6,22	4,48	5,62	5,00	2,86	4,52
		Ecart-type	2,72	2,46	2,69	2,47	2,43	2,65
4	Bienveillants	N	202	202	202	202	202	202
		Moyenne	5,58	3,88	4,95	4,30	2,30	3,95
		Ecart-type	2,66	2,36	2,65	2,58	2,52	2,73
Total		N	1017	1017	1017	1017	1017	1017
		Moyenne	6,38	4,67	5,68	5,11	3,14	5,01
		Ecart-type	2,76	2,55	2,80	2,65	2,56	3,05
Sig.			,000	,000	,000	,000	,000	,000

Nous synthétisons à présent l'ensemble de ces résultats, tout en les mettant en perspective.

5 Mise en perspective des résultats

Nous abordons, dans cette dernière partie, les limites liées à cette étude (5.1), avant d'évoquer les principaux enseignements de celle-ci (5.2).

5.1 Les limites de l'étude

Plusieurs limites inhérentes à ce second travail exploratoire doivent être soulignées. Elles représentent cependant autant de pistes d'amélioration et de recherches futures.

La première limite concerne la représentativité de l'échantillon. Comme nous l'avons déjà indiqué, nos répondants sont surtout des hommes, plus expérimentés que l'internaute français « moyen ». Ceci vient limiter le caractère généralisable des résultats à l'ensemble de la population française utilisant Internet, mais leur confère une valeur prospective.

Une autre limite tient à la possible sous-représentation des « réticents » au sein de notre échantillon. En effet, des individus très préoccupés par les questions de vie privée pourraient avoir décliné l'invitation à participer à l'étude. Il faut néanmoins préciser que ce type de biais est présent dans la majorité des recherches sur ce thème. Afin de limiter au maximum son impact sur les résultats, plusieurs mesures ont été prises. Tout d'abord, la caution apportée par les deux sites « recruteurs », l'éthique sur laquelle ils sont fondés ainsi que les sujets qu'ils abordent (i.e. « l'innovation et la recherche sur l'Internet et ses usages » pour la FING et la sécurité pour secuser.com) réduisent le risque qu'un internaute préoccupé par la protection de son intimité ne s'exprime pas. Ensuite, nous avons, dès la page d'accueil, garanti la confidentialité des données fournies par les répondants. Enfin, nous nous sommes engagés à publier les résultats de l'enquête afin que les personnes sollicitées visualisent l'apport de leur participation et soient encouragées à répondre, malgré leurs réticences premières.

Enfin, des limites liées au type d'analyse choisie doivent aussi être mentionnées, même si les contraintes auxquelles nous devions faire face expliquent en grande partie les choix opérés. Ainsi, l'un des inconvénients des méthodes de classification est de reposer sur la subjectivité. De plus, la plupart des algorithmes proposés fonctionnant de manière heuristique ; ils permettent de trouver une solution, sans que l'on ne puisse toujours démontrer qu'il s'agit de la meilleure. Les fondements théoriques - le recours aux études antérieures sur le sujet - et empiriques - par le biais d'indices et de représentations graphiques - utilisés pour justifier le nombre de classes retenues, les nommer et les interpréter, viennent toutefois limiter ce biais. De plus, la concordance des résultats sur les deux parties de l'échantillon et la qualité de la classification obtenue (des groupes homogènes, équilibrés et de taille suffisante) accroissent la validité de nos analyses et confortent l'idée selon laquelle il est possible de segmenter la population des internautes français en 4 catégories majeures. Nous avons, de surcroît, vérifié que l'analyse non hiérarchique en 4 classes était meilleure que celle obtenue respectivement pour 3 et 5 groupes.

5.2 Les enseignements de l'étude

Trois enseignements majeurs peuvent être signalés.

Le premier apport de cette étude est d'avoir une vision plus claire de la gestion de données personnelles opérée par les internautes français. Bien qu'à visée descriptive, cette étude offre en effet un premier aperçu des arbitrages opérés en diverses occasions (contexte et interlocuteur différents notamment).

En réponse aux deux questions que nous avions posées dans l'introduction de ce chapitre (i.e. *les internautes français diffèrent-ils quant aux avantages recherchés en échange de la divulgation de leurs données personnelles en ligne ? Peut-on expliquer ces différences ?),* il est possible de répondre par l'affirmative. Cette étude montre en effet qu'il est possible de segmenter la population des internautes français sur la base des avantages recherchés, en échange de leurs données. La classification réalisée aboutit à la formation de 4 classes d'individus (i.e. « désintéressés, réticents, négociateurs et bienveillants ») qui se différencient par leur attrait pour les avantages susceptibles d'être obtenus contre leurs données. Ces quatre groupes se distinguent aussi par leurs profils contrastés d'un point de vue sociodémographique, expérientiel, attitudinal et comportemental. Parmi les variables discriminantes majeures, on trouve ainsi la nationalité, le niveau d'étude, l'expérience d'Internet (y compris l'expérience d'achat en ligne), le risque perçu à fournir des données en ligne et enfin différents éléments comportementaux comme le comportement habituel de réponse et le comportement de partage selon le type de contexte et/ou d'interlocuteur.
Ces résultats confortent donc l'idée selon laquelle le type de bénéfices proposés en échange des données serait un élément incitatif majeur, pour certaines catégories d'individus au moins, les autres se décidant aussi sur la base d'autres éléments (comme les garanties en termes de confidentialité pour les « désintéressés »). Seuls les « réticents » seraient réellement peu enclins à délivrer des données personnelles, cette réticence pouvant cependant se réduire en cas de sollicitation par une entreprise avec laquelle ils ont déjà établi un contact au préalable.

Le troisième et dernier apport de cette étude est de donner des indications quant à l'impact de certains facteurs sur la décision de fournir ou non des données personnelles en ligne. Ainsi, on apprend que la familiarité avec l'entreprise serait un élément plus décisif que sa seule réputation ou notoriété. On voit aussi tout l'intérêt de limiter la sollicitation à des informations qui seront jugées pertinentes selon le contexte dans lequel elles sont demandées et l'objectif sous-jacent. Ainsi, si la grande majorité des internautes interrogés sont d'accord pour fournir des renseignements précis quand ils passent commande ou quand ils souhaitent obtenir un document ou une information particulière, ils le sont beaucoup moins quand il s'agit simplement d'entrer sur le site. Le contexte de sollicitation semble donc engendrer des comportements de réponse différents et pourrait s'avérer être un facteur important quand il s'agit d'évaluer la demande et de décider de fournir ou non les renseignements demandés.

Au-delà des résultats obtenus, la mise en évidence de différentes classes suggère que la question de l'échange de données personnelles doit être abordée avec nuance.
A l'évidence, toute solution générale pour contrer les craintes relatives au respect de la vie privée en ligne aura peu de chance de fonctionner. Les sites souhaitant mieux connaître leurs clients, de manière à développer une politique relationnelle efficace, devront donc réfléchir aux moyens d'inciter les internautes à se dévoiler, sans pour autant dépasser le seuil de tolérance admis. Si peu de problèmes se posent pour les individus peu préoccupés par la question de l'exploitation de leurs données (les Bienveillants), l'essentiel de la réflexion devra être consacré à la manière de convaincre les internautes appartenant aux trois autres segments. Les solutions devront alors être différenciées selon que l'individu se montre concerné par la protection de sa vie privée ou envisage une approche plus pragmatique de l'échange.

Tous ces éléments, auxquels s'ajoutent les résultats de l'étude qualitative, nous seront très utiles lors de l'élaboration du cadre conceptuel. Ils permettent en effet, en les confrontant aux conclusions de la revue de littérature, de formuler certaines hypothèses de recherche, que nos études ultérieures viendront (ou non) confirmer. Ainsi, l'effet de l'interlocuteur à l'origine de la sollicitation (et plus particulièrement du niveau de familiarité avec celui-ci) et du contexte dans lequel les données sont demandées semblent très intéressants à tester, d'autant que la littérature est quasi inexistante sur le sujet. De même, l'impact de solutions visant à inciter certaines catégories d'internautes à se dévoiler pourrait aussi être vérifié. Nous pensons, plus particulièrement, aux garanties de confidentialité (cas des désintéressés) et au processus de remplissage (nombre de champs à remplir, par exemple) (cas des réticents).

Conclusion du chapitre 6

Le travail empirique, présenté ici, avait pour objectif de décrire les pratiques de partage d'informations des internautes français et de mettre en évidence une typologie, en s'appuyant sur les avantages recherchés lors de la divulgation de données personnelles. L'échantillon de 1364 individus peut ainsi être scindé en quatre classes regroupant des internautes aux comportements et attitudes bien différenciés.
Il est intéressant de constater que, bien qu'une majorité de répondants se disent concernés par les questions liées au RVP, leurs réactions apparaissent extrêmement variées quand il s'agit d'évaluer leur comportement de partage d'informations. Si les uns divulguent rarement leurs données (les réticents), d'autres y sont au contraire plutôt favorables (les bienveillants). Enfin, si certains cherchent à obtenir le maximum de bénéfices (les négociateurs), d'autres rejettent toute compensation monétaire qui pourrait leur être proposée en échange (les désintéressés).
Cette étude a, en outre, été menée auprès d'un échantillon d'internautes aguerris. Or, si des craintes apparaissent chez certains d'entre eux, il est logique de penser que cela sera *a fortiori* le cas pour des néophytes. Analyser le comportement de l'ensemble de la population semble donc indispensable, ce à quoi nous nous attacherons dans la suite de ce travail doctoral.

Bien qu'exploratoire, il s'agit de la première étude française de ce type, sur ce sujet. Celle-ci souligne que la question du respect de la vie privée en ligne ne peut être envisagée de manière uniforme. Cet aspect est essentiel du point de vue des pratiques managériales car les sites, concernés par le recueil de données clients et par la validité de ces informations, doivent pouvoir se doter de stratégies adaptatives. Elle prouve aussi que la question des avantages proposés en échange des données est d'une importance stratégique. Il s'agit, en effet, d'un élément décisif, susceptible de convaincre une grande partie des consommateurs. D'autres facteurs, comme la familiarité avec l'organisme à l'origine de la requête ou le contexte dans lequel les informations sont demandées, pourraient également être des éléments sur lesquels les individus s'appuient, afin de décider s'ils fourniront ou non les renseignements requis.

Ayant indiqué l'ensemble des constats issus de la revue de littérature (Chapitres 2 à 4) et les résultats de nos deux études exploratoires (Chapitres 5 et 6), nous présentons maintenant le modèle conceptuel sur lequel nous nous baserons dans le cadre de ce travail doctoral et qui servira de base à la formulation de nos hypothèses de recherche.

Partie II – EXPLORER ET BATIR

Sollicitation de données personnelles
et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

CHAPITRE 7
FORMULATION DU CADRE CONCEPTUEL GÉNÉRAL DE LA RECHERCHE

« The process by which a sample member views and answers questions has been, for the most part, a black box ».

Bosnjak and Tuten (2001)

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

Chapitre 1 : Les enjeux liés à la collecte de données clients

Chapitre 2 : Préambule sur la notion de vie privée

Chapitre 3 : Dévoilement de soi et divulgation de données personnelles

Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)

Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)

Chapitre 7 : Formulation du cadre conceptuel général de la recherche

Chapitre 8 : Méthodologie de la recherche

Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)

Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)

Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 7

Webb (1981) suggère qu'il est difficile d'étudier les comportements adoptés par les individus face aux problèmes liés à la vie privée parce que ceux-ci sont susceptibles de varier selon la situation rencontrée. Le fait qu'un individu estime qu'il y a violation ou non dépendra alors des caractéristiques de la situation, mais aussi de son propre jugement de celles-ci. De plus, chaque situation sera jugée à travers un ensemble préexistant de valeurs (préoccupation pour le respect de la vie privée, notamment) mais surtout d'après les aspects saillants de celle-ci.

Ce constat issu de la littérature pourrait s'avérer d'une grande utilité au moment de théoriser nos intuitions en hypothèses testables. Dans ce chapitre, après avoir exposé les résultats inhérents à nos deux études préalables, de nature qualitative et descriptive (chapitres 5 et 6 respectivement), nous abordons en effet la formulation du cadre conceptuel général de la recherche, visant à expliquer les réponses attitudinales et comportementales d'un individu, face à une sollicitation de ses données personnelles, en particulier par le biais d'un formulaire à remplir en ligne.

Nous commençons par justifier et commenter les différents choix auxquels nous avons pu être confrontés. Ceux-ci concernent, notamment, le contexte (1.1) et la phase d'analyse (1.2), mais aussi le type de processus considéré (1.3 et 1.4). Nous abordons ensuite l'ancrage théorique majeur de ce cadre conceptuel, en exposant les trois théories utilisées pour formuler nos hypothèses de recherche. Il s'agit de la théorie de l'échange social (2.1), de la théorie de la motivation (ou de l'espérance attendue) (2.2) et de la théorie de l'action raisonnée (2.3).

Le point suivant vise à présenter le modèle conceptuel général de la recherche (3) et à justifier son articulation avec la littérature existante et les résultats issus des études préalables. Nous terminons par la formulation des propositions hypothèses de recherche (4), qui sont finalement résumées à la fin de ce chapitre, dans le tableau 7.1.

1 Justifications et choix préalables

Plusieurs choix préalables à la constitution du cadre conceptuel doivent être exposés de prime abord. Ils expliquent et justifient en effet (au moins en partie) le choix des théories mobilisées, les concepts retenus dans le cadre de ce travail et leur imbrication au sein du modèle. Ces choix concernent le contexte d'analyse (1.1), la phase d'analyse (1.2)[113] et le type de processus considéré (1.3 et 1.4).

1.1 Le contexte d'analyse

Notre travail de recherche se concentre sur la réponse du consommateur face à une sollicitation de données personnelles faite de manière ouverte par l'entreprise (mode de collecte directe). En ce sens, nous nous intéressons uniquement au cas où le consommateur est conscient que des informations le concernant sont collectées. De plus, celui-ci a le choix de fournir ou non les données demandées. Il prend donc sa décision sur la base d'éléments objectifs et subjectifs lui permettant d'évaluer la situation et de décider s'il est préférable ou non de répondre.

Nous étudions, plus particulièrement, le cas où la sollicitation s'opère en ligne, lorsque le consommateur est en train de surfer et qu'il se retrouve face à un formulaire (contenant des données personnelles) à compléter. De plus, nous faisons l'hypothèse que c'est l'individu (et non l'entreprise) qui est à l'origine du contact, dans le sens où celui-ci se retrouve, de son propre fait (y compris s'il est guidé par un moteur de recherche), sur le site en question et sur la page contenant le formulaire. Ainsi, même si le site utilise des procédés visant à amener l'internaute sur cette page (comme des pop-up), ce dernier a le choix de ne pas s'y rendre.

Enfin, nous nous intéressons à la situation où le consommateur doit décider de remplir ou non le formulaire proposé, quel que soit le contexte dans lequel cette décision s'opère (inscription sur le site, participation à un jeu-concours en ligne, demande de renseignements, …). Nous nous interdisons cependant d'étudier le cas où le remplissage du formulaire concerne un achat en ligne (bien qu'il s'agisse là d'un cas courant). Nous considérons en effet que ce cas implique des mécanismes de décision spécifiques, qui semblent par ailleurs davantage liés à l'achat lui-même qu'au dévoilement d'informations personnelles (ce qui nous intéresse dans le cadre de cette étude). Les concepts afférents à l'achat en ligne (risques liés au paiement notamment) sont donc omis de cette recherche.

[113] La différence entre contexte et phase d'analyse sera clarifiée à la lecture des paragraphes correspondants.

1.2 La phase d'analyse

Par phase d'analyse, nous entendons l'étape plus particulièrement étudiée au sein d'un processus de choix. L'objectif majeur de notre recherche est d'expliquer le processus de décision du consommateur et les critères mobilisés pour ce faire, une fois que celui-ci se retrouve face au formulaire à compléter. Certains concepts clés, y compris parmi ceux identifiés à travers la littérature ou les études préalables, ne seront donc pas inclus dans notre modèle puisque nous considérons qu'ils interviennent avant cette « ultime » phase de décision (remplir ou non le formulaire proposé). Ainsi en est-il, par exemple, de la perspective temporelle (définie comme le temps dont dispose l'individu pour accomplir la tâche)[114]. Selon nous en effet, l'impact de cette variable serait essentiellement antérieur à la décision de compléter le formulaire : un individu pressé ne se serait pas rendu sur le site ou n'aurait pas cliqué sur le lien conduisant au formulaire sachant qu'il n'aurait pas le temps d'y répondre.

A l'instar de Furse et Stewart (1982), plutôt que d'étudier la réponse globale à un phénomène complexe, nous proposons de découper l'ensemble du processus en une série de décisions séquentielles (cf. Annexe 7.1). Dans le cadre qui nous intéresse ici, ce processus se décompose en 4 phases majeures, au cours desquelles l'individu a, à tout moment, la possibilité de s'arrêter. Nous choisissons de nous intéresser plus particulièrement à la dernière phase (i.e. arrivée sur la page contenant le formulaire et réponse à la sollicitation).
Les deux questions auxquelles nous souhaitons répondre sont alors les suivantes :

- arrivé sur une page Internet contenant un formulaire avec des données personnelles à fournir, quelle est la décision prise par le consommateur parmi les trois possibilités qui s'offrent à lui (compléter le formulaire et le valider, abandonner, quitter le site)[115] ?
- quels éléments influencent cette décision ? sur quelle base se décide t-il ?

1.3 Le type de processus considéré : heuristique et/ou systématique

Dans le cadre de la réponse à un sondage, la distinction entre un processus heuristique et/ou systématique de prise en compte de l'information semble pertinente pour mieux comprendre la décision de participer (ou non). Dans ce contexte, le processus utilisé serait essentiellement heuristique : les gens basent leur décision sur quelques considérations majeures (comme la longueur du questionnaire ou la crédibilité du prestataire). Groves, Cialdini et Couper (1992)

[114] L'impact de cette variable avait notamment été souligné dans les chapitres 4 et 5.
[115] En réalité, nous étudierons surtout la première possibilité : compléter (ou non) le formulaire.

estiment que la justification d'une telle approche est liée au fait que les gens ont souvent peu d'intérêt personnel à répondre et/ou ne sont pas prêts à dépenser leur temps et leur énergie pour prendre cette décision. A l'opposé, dans les sondages de populations fortement affectées par les résultats de l'étude (par exemple de nature intra-organisationnelle) ou sur des sujets de saillance élevée, les principes d'un choix rationnel peuvent s'appliquer.

Le modèle ELM (Elaboration Likelihood Model) de persuasion (Petty et Cacioppo 1986) semble être un cadre théorique intéressant pour expliquer le processus de décision du consommateur, dans le contexte qui nous intéresse ici. Le modèle explique que la motivation (volonté et intention de la personne de traiter l'information) et la capacité à répondre (compétences nécessaires pour décoder l'information) sont deux éléments importants lors de l'évaluation du message (ici, du formulaire sollicitant les données). MacInnis et Jaworski (1989) y ajoutent un troisième élément : l'opportunité, c'est-à-dire les conditions qui facilitent ou inhibent le décodage de l'information (comme les distractions ou la pression du temps). Deux possibilités existent alors. Ainsi, un niveau élevé d'élaboration intervient quand l'information est véhiculée par la route centrale de la persuasion. Dans ce cas, les répondants sont très attentifs au message, analysent son contenu et arrivent ensuite à une conclusion. Ce processus, de nature systématique, requiert de grandes ressources cognitives (Meyers-levy et Sternthal 1991). A l'opposé, un faible niveau d'élaboration (faibles motivation et capacité) conduit à ce que l'information soit véhiculée par une route périphérique ou heuristique. Dans ce cas, l'allocation de ressources cognitives est limitée. L'individu n'utilise que quelques éléments clés pour se décider - comme la sympathie avec la source (Chaiken 1980, Petty et al. 1986) ou l'attractivité du questionnaire (Miniard et al. 1991) - dans le cadre d'un sondage. Ces deux routes sont en fait employées tour à tour, selon les circonstances : quand le niveau d'élaboration est élevé, la composante dominante de la réaction est cognitive tandis que, lorsqu'il est faible, elle est essentiellement affective.

Dans le contexte d'une sollicitation de données personnelles, par le biais d'un formulaire à remplir, il est difficile de savoir si le processus de décision est systématique ou heuristique. D'un côté, les individus soumis à cette situation se sentent très concernés puisqu'il s'agit de fournir des données personnelles (parfois sensibles). Ils ne prennent donc pas cette décision à la légère et auront tendance à être attentif à tout signe susceptible de les aiguiller dans leur choix. Ceci plaide en faveur d'une vision systématique du phénomène. D'un autre côté, plusieurs éléments (temps disponible, expérience antérieure, asymétrie d'informations..) peuvent les conduire à une allocation de ressources cognitives plus limitée. Dès lors, la

décision se prendra sur la base de quelques considérations majeures. Le choix de l'une ou l'autre optique (systématique ou heuristique) est délicat. Il est très dépendant de la situation étudiée et de la personnalité de l'individu soumis à ce choix. Son niveau d'implication par rapport aux produits/services offerts par l'entreprise concernée pourrait ainsi être un élément clé. Au final, nous pensons que le processus est rarement systématique puisque les individus ne sont pas prêts à allouer plus de temps et d'énergie que nécessaire pour prendre leur décision. Le processus de choix (conduisant à remplir ou non le formulaire) serait donc essentiellement heuristique, fondé sur quelques éléments clés, tels que ceux trouvés au travers de l'étude qualitative préalable (confidentialité perçue, sensibilité et pertinence perçues des données sollicitées, gain net à répondre).

1.4 Un processus guidé par les composantes cognitives et/ou affectives

Woodman et al. (1982) expliquent que la perception de respect de la vie privée (RVP) est un phénomène psychologique ayant des composantes à la fois cognitives (par exemple, croyances sur le type de données considérées comme privées) et affectives (sentiments et émotions liées aux pratiques de sollicitations). Une personne peut ainsi reconnaître la légitimité d'une telle pratique (composante cognitive) mais aussi être anxieuse à la pensée de devoir se dévoiler (composante affective). La reconnaissance de la dualité du phénomène suppose que toute tentative de mesure de ce processus inclut à la fois les croyances des individus sur la sollicitation et leurs sentiments à cet égard. Selon Fromkin et al. (1979), la mesure d'une seule de ces composantes donnera une image incomplète de la réponse du consommateur à ce sujet. Nous veillerons donc, dans le cadre de la formulation du cadre conceptuel, à intégrer ces deux niveaux d'analyse afin d'avoir une vision complète, la plus proche possible de la réalité.

2 Ancrage théorique

Trois principales théories[116] seront mobilisées pour établir notre cadre conceptuel : il s'agit de la théorie de l'échange social (Blau 1964, Emerson 1972), de la théorie de la motivation ou de l'espérance attendue (*Expectancy Theory)* (Vroom 1964) et, enfin, de la théorie de l'action raisonnée (*Theory of Reasoned Action*) (Fishbein 1967, Fishbein et Ajzen 1975).

[116] Nous en utiliserons en réalité une quatrième, la théorie de l'engagement (exposée à la fin de ce chapitre, paragraphe 4.4.1), qui nous permettra de développer des hypothèses spécifiques à une variable de notre modèle (i.e. l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données). Nous ne l'exposons donc pas dans cette section étant donné son caractère plus « accessoire », au sens où elle n'a pas (réellement) contribué à l'élaboration du cadre conceptuel.

Ces trois théories seront exposées brièvement ci-dessous, en insistant sur leurs apports pour notre réflexion et en expliquant la manière dont elles ont contribué à l'élaboration du modèle. La théorie de l'action raisonnée (dite TRA) constitue le « ciment » de ce cadre conceptuel. Comme nous le verrons plus loin, elle est, notamment, à l'origine de la nature processuelle du modèle que nous serons amenés à tester et de la place des différentes variables qui y figurent. Elle aurait donc dû logiquement être exposée en premier. Nous la présenterons toutefois en dernier (i.e. après celle de l'échange social et de l'espérance attendue) afin de montrer le lien qui existe entre les principes issus de cette théorie (TRA) et notre propre modèle conceptuel.
Nous débutons donc cette présentation par deux théories qui, loin d'être anecdotiques dans le cadre de notre réflexion, sont toutefois moins fondamentales, car elles sont spécifiques à certaines variables et ne constituent pas le cœur du modèle. Elles seront néanmoins fort utiles dans la formulation de certaines de nos hypothèses de recherche.

2.1 Les théories de l'échange et de l'échange social

Parmi les fondements théoriques majeurs de cette recherche, on trouve la théorie de l'échange et de l'échange social. Cette théorie est en effet fréquemment mobilisée, dans le cadre de la réponse à un sondage ou dans les recherches sur la sollicitation de données personnelles.
Voyons en quoi les principes issus de cette théorie s'appliquent à notre cadre d'analyse[117].

2.1.1 Les principes de la théorie de l'échange et de l'échange social

Le paradigme de l'échange - et la théorie qui en découle - est au cœur du concept de marketing (Bagozzi 1975) et constitue donc une base théorique utile pour étudier la question de la fourniture de données personnelles, dans le domaine commercial. Cette théorie permet en effet de décrire les interactions entre les entreprises et les consommateurs.
L'échange est le résultat d'un comportement axé sur l'atteinte d'un but (Houston et Gassenheimer 1987). Il implique le transfert d'éléments tangibles ou intangibles, utilitaires ou symboliques entre deux ou plusieurs acteurs (Bagozzi 1975). Chaque partie à l'échange donne et reçoit une chose de valeur, qu'elle soit monétaire, matérielle, informationnelle ou symbolique. Selon Blalock et Wilken (1979), pour qu'il y ait échange, il faut que les gens soient capables d'anticiper les conséquences de leurs actions. Dès lors, ils adopteront le comportement qui engendrera leurs conséquences anticipées préférées. Kotler (1984) identifie

[117] Cette théorie a déjà été abordée dans le chapitre 3. Nous revenons brièvement sur ses principes avant d'indiquer en quoi elle s'applique à notre cadre conceptuel.

d'autres modalités préalables à la réalisation de l'échange : 1) au moins deux parties autonomes (capables de prendre une décision seules) ; 2) chaque partie dispose d'éléments ayant une valeur pour l'autre (tangible ou intangible) ; 3) chaque partie est capable de communiquer ou de livrer ces éléments ; 4) chaque partie est capable de rejeter ou d'accepter l'offre ; 5) chaque partie pense qu'il est approprié de faire affaire avec l'autre. L'échange sera donc réussi si chaque partie estime avoir donné et reçu au moins ce qui était prévu.

Lorsqu'un prospect ou un client répond favorablement à une sollicitation « directe » de données personnelles de la part d'une entreprise, il s'agit bien d'un échange. Tout d'abord, chacun donne et reçoit un élément ayant de la valeur pour l'autre partie (les informations personnelles du consommateur contre l'octroi d'avantages spécifiques ou la possibilité de se voir proposer des offres mieux adaptées à son profil). De plus, le consommateur a la possibilité de s'y soustraire, ce qui ne serait pas le cas si la collecte se faisait « discrètement ». Cet échange est dit « mixte » - au sens de Bagozzi (1975) - puisqu'il combine à la fois des éléments utilitaires et symboliques. Ce type d'échange se base sur les principes suivants : l'individu est parfois rationnel, parfois irrationnel ; il est motivé par des récompenses tangibles et intangibles et par des forces extérieures. De plus, surtout si l'information dont il dispose pour se décider est incomplète, il fait ce qu'il pense être le mieux pour lui, sans vouloir forcément maximiser ses profits, mais en cherchant à limiter au maximum les risques. Cette idée fait référence à l'analyse coûts/bénéfices prônée par la théorie de l'échange social (Blau 1964). Selon cette théorie en effet, utilisée à l'origine pour expliquer les relations et le comportement humain, l'individu choisit le comportement générant les résultats les plus favorables (Blau 1964, Emerson 1972).

2.1.2 La théorie de l'échange social appliquée au dévoilement de soi

Les recherches menées dans le domaine de la sollicitation d'informations personnelles auprès des consommateurs ont montré que ces derniers pouvaient accepter de fournir leurs données en échange d'un bénéfice économique et/ou social. Plus précisément, les individus compareraient ce type de bénéfices avec les coûts engendrés par la divulgation. Ils se décideraient alors, notamment, en fonction des bénéfices à rendre leurs données publiques et des risques encourus à ne pas les garder privées. Cette détermination par laquelle le consommateur accepte de livrer ses informations personnelles à une entreprise si les bénéfices sont supérieurs aux risques est appelée « *privacy calculus* », en référence au « *calculus of behavior* » de Laufer et Wolfe (1977). Il s'agit du jugement par lequel passe l'individu pour

savoir si ses données valent la peine d'être dévoilées et ne risquent pas d'être utilisées à son insu par la suite. Plus précisément, le consommateur acceptera de dévoiler des données personnelles s'il estime : 1) que ces dernières seront utilisées de manière équitable et 2) qu'il ne souffrira pas de conséquence négative à l'avenir (Milne et Gordon 1993 ; Stone et Stone 1990). D'autres travaux montrent que les consommateurs sont souvent d'accord pour donner des informations en échange de récompenses comme des cadeaux ou des catalogues (Schwartz 2000). Ce système de compensation est conforme aux théories de l'échange social et des relations interpersonnelles (Taylor et al. 1969 ; Taylor et Altman 1975). Un des prémices de cette théorie est en effet de dire que l'acteur est motivé à se dévoiler par la présence d'intérêts et/ou de récompenses, un de ces intérêts étant satisfait à travers l'interaction sociale (Cook et Whitemeyer 1992). Klopfer et Rubenstein (1977) soulignent, pour leur part, que la protection des données personnelles (par exemple, le fait de ne pas s'inscrire sur un site) peut aussi s'interpréter en termes de rapport coûts/bénéfices : elle a lieu et persiste tant qu'un tel comportement s'avère bénéfique.

Dans le cadre de ce travail doctoral, le comportement de réponse à une sollicitation de données personnelles sera donc examiné dans le contexte d'un échange dyadique, impliquant le consommateur d'une part, et l'entreprise (ou le site) qui le sollicite, d'autre part. Cette dernière cherche à obtenir des informations de son prospect/client, en contrepartie de services rendus ou de gains divers (personnalisation, cadeaux, …). Chaque acteur doit donc fournir des ressources à l'échange et peut espérer en retirer un gain substantiel. Dès lors, conformément à la littérature et aux études exploratoires que nous avons menées, l'individu évaluera les coûts et bénéfices liés à la communication de ses données et n'acceptera de fournir ces dernières que si le rapport entre les deux lui est favorable. Dans ce cas, la réponse du consommateur sera considérée comme un échange coopératif d'informations.
Ceci peut se résumer à travers le schéma suivant :

Figure 7.1 Analyse coûts/bénéfices dans le cadre d'une sollicitation de données personnelles

Toutefois, ce calcul du rapport coûts/bénéfices, bien que primordial lors de l'évaluation de la sollicitation, ne serait pas le seul élément sur lequel se fonde l'individu pour décider de fournir ou non les données personnelles demandées. D'autres éléments, identifiés notamment à travers l'étude qualitative préalable, devront aussi être pris en compte et intégrés dans le modèle conceptuel final.

2.2 La théorie de la motivation ou de l'espérance attendue

La théorie de la motivation ou de l'espérance attendue (*Expectancy Theory*) est fondée sur l'idée selon laquelle l'individu qui doit prendre une décision décode les informations qui lui parviennent, afin d'évaluer le lien entre son comportement et les résultats qui en découleront. Il formerait alors des espérances qui guideraient son choix quant à l'alternative à privilégier.

Comme toutes les théories cognitives, la théorie de la motivation se focalise sur le processus cognitif qui se forme avant que le choix ne soit fait. Formulée à l'origine par Vroom (1964), ce cadre explique comment un individu choisit entre différentes alternatives possibles. L'idée de base est que les résultats (attendus) mènent aux comportements. L'individu prend donc sa décision après avoir évalué les résultats prévus, c'est-à-dire les conséquences futures anticipées de chaque alternative. La théorie indique que la personne considère les alternatives comme associées à différents niveaux de performance et différentes probabilités d'obtenir ces résultats. Parmi les alternatives, l'individu choisit alors celle pour laquelle son score de motivation est le plus élevé. La motivation est donc la force qui induit le comportement et est fonction de 3 perceptions distinctes : l'espérance, l'instrumentalité et la valence.

L'espérance (*expectancy*) reflète la croyance selon laquelle un effort donné conduira à un niveau de performance donné. Elle concerne donc le lien entre l'effort et la performance attendue. Dans notre cadre d'analyse, l'effort concerne le fait de remplir le formulaire et la performance, ce que l'individu peut espérer obtenir en échange. Répondre à un formulaire peut ainsi le conduire à plus de confort (plus besoin de s'identifier à chaque fois qu'il revient sur le site), à une meilleure reconnaissance en tant que client, à des bénéfices financiers (grâce à des promotions ciblées) ou encore à des cadeaux, s'il s'agit d'un jeu-concours.

L'instrumentalité (*instrumentality*) reflète la croyance selon laquelle le niveau de performance conduira aux résultats attendus. Elle repose sur l'idée que, si la performance est atteinte, on recevra un meilleur résultat et correspond à la probabilité d'obtenir le résultat espéré. L'expérience passée de l'individu, notamment en matière de traitement de données,

correspond à cette partie instrumentale de la théorie. Les gens ayant vécu une invasion de leur vie privée pourront ainsi ne pas croire que le fait de se dévoiler mènera au résultat attendu.
La valence (*valence*) fait référence à la valeur que l'individu place dans le résultat. Elle correspond à l'attractivité du résultat pour l'individu et est fonction de ses propres besoins, buts et valeurs. La valence peut être positive ou négative. Une valence positive peut, par exemple, provenir d'un sentiment de sécurité lorsqu'on complète un formulaire. A l'opposé, un site avec une politique de RVP peu protectrice va générer une valence négative.

Cette théorie peut se matérialiser à travers le schéma suivant :

Figure 7.2 Processus d'évaluation dans le cadre de la théorie de la motivation
(d'après Farag et Krishnan, 2003)

Dans le cadre d'une sollicitation de données, une firme peut donc influencer la décision du consommateur en agissant sur l'un ou l'autre des éléments de motivation. Elle peut, par exemple, focaliser son attention sur un résultat associé à une valence positive, afin d'accroître le score de motivation et donc la probabilité qu'il accepte de fournir les données demandées. Le résultat peut ainsi prendre la forme d'une utilisation éthique des données fournies. L'objectif est alors de faire en sorte que l'individu associe l'effort (fait de fournir les informations) à ce résultat.

Bien que reconnue et utilisée dans de nombreuses recherches, y compris dans le domaine de la sollicitation de données personnelles, cette théorie souffre cependant d'un manque de validation empirique (Farag et Krishnan 2003). De plus, Connolly (1976) indique que son utilisation est délicate, du fait de la difficulté à conceptualiser chaque élément du processus d'évaluation. Elle paraît toutefois fort adaptée au cadre d'analyse qui nous intéresse ici et sera donc mobilisée pour expliquer et justifier le cadre conceptuel, mais aussi la formulation des hypothèses de recherche. Nous n'aurons pas recours à la modélisation telle que présentée dans le schéma ci-dessus, du fait des limites évoquées précédemment. De nombreux auteurs, avant nous, se sont ainsi contentés de l'utiliser en tant que base théorique.

2.3 La théorie de l'action raisonnée

Contrairement au modèle précédent qui relie directement le processus d'évaluation au comportement, le modèle de structure cognitive de Olson et Dover (1978) met l'attitude au centre du processus qui mène l'individu, de l'exposition au comportement (cf. figure 7.3).

Figure 7.3 Modèle de structure cognitive de Olson et Dover (1978)

Ce modèle n'est pas sans rappeler celui de la théorie de l'action raisonnée (TRA) de Fishbein (1967) et Fishbein et Ajzen (1975), qui considère également l'attitude comme un élément central du processus conduisant au comportement du consommateur (cf. figure 7.4).

Figure 7.4 Théorie de l'action raisonnée de Fishbein (1967) et Fishbein et Ajzen (1975)

Avec : b_i = croyances liées au fait d'accomplir ce comportement (conséquences possibles)
e_i = évaluation de ces conséquences
nb_j = croyances normatives
mc_j = motivations comportementales

Le modèle TRA de Fishbein et Ajzen (1975) vise à étudier les déterminants des comportements consciemment réalisés. Il a d'ailleurs montré sa capacité à prédire les comportements dans de nombreux domaines. Comme l'expliquent les auteurs eux-mêmes, ce modèle est « capable de prédire tout comportement humain » (Ajzen et Fishbein, 1980, p4). Ce modèle semble donc particulièrement adapté pour étudier les facteurs expliquant la réponse du consommateur face à une sollicitation directe de données personnelles en ligne.

Selon le modèle TRA, le comportement est déterminé par l'intention comportementale, elle-même déterminée par l'attitude de la personne envers le comportement en question et les normes subjectives liées à ce comportement. L'intention comportementale (IC) correspond à une mesure de la force de l'intention de se conduire de cette façon. L'attitude est définie comme l'ensemble des sentiments positifs ou négatifs associés au fait d'agir de cette manière ou comme l'évaluation favorable ou non de ce comportement. Dès lors, plus l'individu a une attitude favorable quant au fait d'adopter ce comportement, plus l'intention de s'y conformer sera élevée (Ajzen 1987). Les normes subjectives font, pour leur part, référence à ce que l'individu perçoit de ce que ses proches voudraient qu'il fasse.

Contrairement au modèle de Olson et Dover (1978) qui distinguait une seule forme de croyances, Fishbein et Ajzen (1975) scindent ces croyances en deux composantes (comportementale et normative), la première influençant l'attitude de l'individu quant au fait de réagir de cette manière tandis que la seconde influence les normes subjectives. Plus précisément, l'attitude de la personne est déterminée par la probabilité subjective selon laquelle le fait de réagir de cette manière aura certaines conséquences (croyances comportementales) multipliée par l'évaluation (estimation implicite) de ces conséquences. Les normes subjectives correspondent, quant à elles, à une fonction multiplicative des croyances normatives (attentes perçues de la part de référents spécifiques) et de la motivation à répondre à ces attentes. Comme Ajzen et Fishbein (1980) le reconnaissent eux-mêmes, les normes subjectives constituent la partie la plus floue du modèle, principalement du fait de l'incertitude quant au statut théorique et psychométrique de cette variable au sein du modèle.
Plusieurs études réalisées par la suite montrent d'ailleurs que ces normes n'ont pas toujours d'influence, ce qui a amené certains auteurs à supprimer cette variable du modèle. Cette suggestion serait d'autant plus censée lorsqu'on étudie le comportement des individus dans une situation hautement personnelle et/ou pour laquelle l'influence des autres est faible, ce qui serait notamment le cas sur Internet. En effet, plusieurs recherches montrent que lorsque l'individu navigue sur le web, les informations sur le contexte social sont souvent limitées (hormis le cas particulier des forums de discussion et de tous les sites qui favorisent le « rapport aux autres »), ce qui rend la situation relativement impersonnelle. Cela oblige alors celui-ci à être centré sur lui-même et le rend peu concerné par les normes sociales et l'impression qu'il laisse aux autres (Kiesler et al. 1984). Matheson et Zanna (1989) trouvent ainsi que les internautes sont plutôt tournés vers eux-mêmes et ont une faible conscience des autres (ou préoccupation pour ce que pense autrui). Ils en concluent qu'ils ont une plus grande

tendance à exprimer leurs sentiments et qu'ils ont moins tendance à adopter des comportements inhibés. Ces résultats sont confirmés par la recherche menée par Lwin et Williams (2003) qui montre l'absence d'effet des normes subjectives sur l'intention des internautes de déguiser leur identité en ligne (de mentir lorsqu'ils remplissent un formulaire). Toutes ces conclusions confortent l'idée selon laquelle, dans certains cas, les normes subjectives n'auraient qu'une influence relative (voire nulle) sur l'intention comportementale de réponse. Dans la mesure où nous étudions le comportement des individus face à une sollicitation de données personnelles en ligne, nous supposons que ceux-ci seront peu influencés par leur entourage pour prendre la décision de délivrer ou non les renseignements demandés. Nous ne tiendrons donc pas compte de cette variable dans notre modèle.

Les chercheurs dans le domaine des systèmes d'information suggèrent que les modèles d'intentions tirés de la psychologie sociale (dont le modèle TRA) pourraient constituer un fondement théorique majeur, pour les recherches sur les déterminants du comportement, dans le domaine des technologies de l'information, et donc de l'Internet. Le modèle TAM (*Technology Acceptance Model*), introduit par Davis (1989), est une adaptation du TRA, spécialement élaborée pour mesurer l'acceptation des systèmes d'information. L'objectif de ce modèle est d'expliquer l'impact de facteurs extérieurs sur les croyances, l'attitude et l'intention des individus d'adopter un comportement lié aux NTIC. Il montre que deux perceptions majeures (l'utilité et la facilité d'usage perçues) sont particulièrement pertinentes pour expliquer l'acceptation d'utilisation des systèmes d'informations, par le biais de l'attitude et de l'intention comportementale. Ce modèle remplace donc les croyances par des perceptions relatives à l'utilisation du système. De plus, il n'intègre pas les normes subjectives, dont on a démontré la faiblesse dans le paragraphe précédent.

En nous basant sur le modèle TRA, complété par les apports du modèle TAM, il est possible d'élaborer un processus de réponse de l'individu face à une sollicitation de ses données personnelles en 4 phases. Une fois sollicité pour fournir ses données personnelles (arrivée sur la page contenant le formulaire à remplir), l'internaute procèderait à l'évaluation de la sollicitation (formation des croyances), sur la base des perceptions relatives à l'ensemble des éléments caractérisant la situation. Suite à cette évaluation, l'individu développerait une attitude (favorable ou défavorable) face à la communication de ses données, qui déterminerait son intention comportementale de réponse (probabilité de fournir les données), celle-ci déterminant à son tour le comportement réel, en cas de sollicitation effective (cf. figure 7.5).

Figure 7.5 Processus de réponse du consommateur face à une sollicitation de données

Inspiré de la Théorie de l'Action Raisonnée (Fishbein et Ajzen 1975)

Nous avons fait le choix de baser notre cadre conceptuel sur le modèle tiré de la théorie de l'action raisonnée (*Theory of Reasonned Action* ou TRA) de Fishbein (1967) et Fishbein et Ajzen (1975). Nous aurions pu, tout aussi bien, utiliser la théorie du comportement planifié (en anglais, *Theory of Planned Behavior*, dite TPB).

Cette dernière représente en effet une version améliorée du modèle TRA, avec l'addition du contrôle perçu comme nouvelle variable (Ajzen 1991). Cet ajout est le résultat de la découverte que le comportement ne peut pas être entièrement volontaire et sous contrôle. Selon cette nouvelle théorie (i.e. TPB), l'intention d'agir est alors prédite par 3 antécédents : l'attitude, les normes subjectives et le contrôle perçu. L'impact de l'attitude est similaire à celui prédit par le modèle TRA ; nous ne reviendrons donc pas dessus. Nous avons déjà expliqué pourquoi, dans le cas présent, les normes subjectives ne seraient pas étudiées. Concernant le contrôle perçu, celui-ci serait fonction des croyances en termes de contrôle et de la facilité perçue, qui se réfèrent à la présence (ou l'absence) de ressources ou d'opportunités permettant de mettre en œuvre le comportement en question (dans notre cas, remplir le formulaire) (Ajzen et Madden 1986). Cette perception de contrôle volontaire, qui peut donc être mesurée par la facilité perçue à accomplir l'action, est censée affecter l'intention d'agir ainsi que la réussite de l'action (Chang 1998), en contrôlant l'effet de l'attitude et des normes subjectives (Ajzen et Madden 1986). Ainsi, dans notre cas, plus l'individu estime qu'il est facile de remplir le formulaire, plus son intention de le faire serait élevée. Selon la théorie, le contrôle perçu aurait également un effet direct sur le comportement (sans passer par l'intention comportementale), mais uniquement quand les perceptions de contrôle s'avèrent justes, ce qui ne serait pas toujours le cas (Madden et al. 1992).

Le principal apport du modèle TPB (par rapport au modèle TRA) est donc de souligner que, si les individus estiment avoir peu de contrôle sur le fait d'adopter le comportement en question, alors leur intention de se comporter comme tel pourrait être faible, et ce, même si l'attitude y est favorable. Bandura et al. (1980) ont ainsi montré empiriquement que le comportement des

individus était influencé par la confiance qu'ils avaient dans leur capacité à accomplir cette tâche. Le chemin direct entre l'intention comportementale et le comportement réel refléterait donc le contrôle dont l'individu dispose pour se comporter de cette manière.

Bien que plus aboutie et disposant d'un pouvoir explicatif plus important (comparé au modèle TRA), nous ne retiendrons pas la théorie TPB dans le cadre de notre modèle conceptuel[118] (dans le sens où nous n'intégrerons pas le contrôle perçu en tant que tel), et ce, pour plusieurs raisons. Tout d'abord, comme nous le justifierons ci-après, nous n'étudierons pas (au sens de mesurer) le lien entre l'intention comportementale de réponse et la réponse. L'impact du contrôle perçu pouvant se faire à ces deux niveaux (tout en contrôlant les autres variables impliquées), nous aurions alors des difficultés à le démontrer. De plus, en cas d'intégration du contrôle perçu (mesuré par la facilité perçue à remplir le formulaire), nous supposons l'existence possible de colinéarités avec une autre variable du modèle, le « gain net perçu à répondre », vu comme la différence entre les bénéfices et les coûts à remplir le formulaire. La facilité perçue (ou au contraire la difficulté) peut en effet être considérée comme un des coûts perçus à répondre. En ce sens, contrairement à ce que nous venons d'indiquer, nous intégrerons partiellement le contrôle perçu (et donc la théorie TPB), à travers cette variable de « gain net perçu à répondre ».

3 Présentation du modèle conceptuel général de la recherche

Dans leur ouvrage, Ajzen et Fishbein (1980) indiquent que le modèle TRA est un modèle général qui ne précise pas les croyances spécifiques à un comportement en particulier. Autrement dit, il incombe aux chercheurs souhaitant utiliser ce modèle, de déterminer les croyances pertinentes correspondant au type de comportement étudié. Ajzen et Fishbein (1980, p 8) recommandent ainsi d'utiliser des croyances spécifiques au sujet d'étude, déduites d'interviews auprès de membres représentatifs de la population étudiée et représentant les éléments les plus fréquemment cités.

En nous appuyant sur les résultats de notre étude qualitative préalable (issus d'entretiens menés auprès d'internautes français et d'allers-retours constants avec la littérature), nous pouvons en déduire les croyances pertinentes dans le cadre de notre recherche. Quatre types de perceptions servant à évaluer la sollicitation ressortent de nos analyses : la confidentialité

[118] Du fait de son intérêt potentiel, notamment dans le cadre de notre sujet de recherche, nous nous réservons cependant le droit d'y avoir recours dans un second temps, notamment pour améliorer le pouvoir explicatif de notre modèle. Nous reviendrons sur cette possibilité dans la discussion finale, à travers les voies de recherche.

perçue de la sollicitation, la pertinence et la sensibilité perçues des données demandées et, enfin, le gain net perçu à répondre (différence entre les coûts et bénéfices perçus).

En référence au modèle présenté précédemment (cf. figure 7.5), le cadre conceptuel général de la recherche présente donc l'impact de l'exposition à la sollicitation (matérialisée par les caractéristiques situationnelles) sur le processus de réponse, par le biais des croyances, représentées par les quatre variables perceptuelles identifiées au cours de l'étude qualitative. Ce cadre matérialise ensuite l'influence de ces perceptions sur l'attitude face à la communication des données, qui détermine à son tour les intentions de comportements. Trois types d'intentions comportementales sont distinguées au sein du modèle : l'intention de répondre (de remplir le formulaire comprenant les données sollicitées), l'intention de mentir (de fournir des données erronées) et l'intention de poursuivre la relation avec l'entreprise sollicitante (qui est liée à l'intention de répondre). Ces trois variables ont déjà été étudiées dans la littérature, même si la majorité des recherches se contentent d'étudier plus particulièrement l'une d'entre elles. Enfin, la réponse à la sollicitation (fait de remplir le formulaire) constitue la variable dépendante majeure du modèle, étant donné son intérêt managérial évident. A noter que nous n'étudierons pas le lien entre les intentions comportementales de réponse et le comportement effectif. Nous avons en effet préféré nous concentrer sur la première partie du cadre conceptuel (qui constitue la partie innovante) et laisser de côté ce qui ne constituait pas (selon nous) la partie fondamentale du modèle. De plus, ce lien a déjà été démontré, notamment dans le cadre du modèle TRA ou du modèle TAM. Enfin, comme l'expliquent Ajzen et Fishbein (1980), l'intention comportementale (IC) est souvent utilisée pour prévoir les comportements réels. Ils précisent toutefois que trois conditions peuvent affecter la magnitude de la relation entre l'IC et le comportement : 1) le degré de correspondance entre la mesure de l'IC et celle mesurant le comportement réel, 2) la stabilité dans le temps des IC et 3) le degré de contrôle volontaire dont l'individu dispose. Nous pouvons donc supposer, à condition de respecter ces trois exigences, que le lien entre les intentions comportementales et le comportement de réponse (fait de fournir les données demandées) sera vérifié et nous nous contenterons alors de vérifier la pertinence du modèle (enchaînement des différentes phases) jusqu'aux intentions comportementales.

Parmi les déterminants du processus de réponse face à une sollicitation de données, identifiés dans la littérature et/ou à travers l'étude qualitative préalable, nous conservons, au sein du modèle, les facteurs situationnels (matérialisés par la phase « exposition à la sollicitation »), puisqu'ils sont directement liés à la sollicitation et aux caractéristiques de celle-ci.

Deux autres déterminants ressortaient des analyses : il s'agit des facteurs individuels (sociodémographiques, expérientiels, ..) d'une part et des facteurs environnementaux (régulation, culture, bouche à oreille, ..) d'autre part. Toutefois, nous avons déduit de nos entretiens qu'ils n'agissaient qu'indirectement sur le processus de réponse, à travers l'attitude face au traitement des données. Nous avons également indiqué qu'il était possible de mesurer ce construit par celui de préoccupation envers le respect de la vie privée, qui est couramment utilisé dans la littérature. Les recherches passées soulignent d'ailleurs l'impact des variables individuelles et de l'environnement sur la préoccupation pour le RVP, laissant ainsi supposer que cette dernière serait en mesure de capter, au moins en partie, l'effet de ces deux types de facteurs. De plus, selon Ajzen et Fishbein (1980), les variables extérieures au modèle ne sont supposées influencer les IC que si elles influencent soit les attitudes soit les normes subjectives. Comme nous n'étudions pas les normes subjectives dans le cadre de ce travail, nous pouvons supposer que la préoccupation pour le RVP aura une influence sur l'attitude face à la communication des données. Cet effet est d'ailleurs très largement documenté dans la littérature sur la vie privée et a déjà été abordé dans les chapitres précédents.

Le modèle conceptuel général de notre recherche se présente donc comme suit (cf. figure 7.6).

Ce modèle fonctionne de la manière suivante : la nature du stimulus, autrement dit les caractéristiques de la sollicitation (i.e. facteurs situationnels : politique de RVP proposée par le site, nature de la relation avec l'entreprise sollicitante, caractéristiques du dispositif de collecte, circonstances dans lesquelles intervient la sollicitation) sont appréciées (évaluées) par le récepteur (l'internaute soumis au formulaire). Ensuite, pour que l'individu accepte de répondre, il faut qu'il développe une attitude favorable face à la communication de ses données. L'établissement de cette attitude favorable est liée à l'existence de quatre conditions majeures : que les données divulguées ne soient pas transférées à un tiers sans accord préalable (confidentialité perçue) et ne le rendent pas potentiellement vulnérable (faible sensibilité perçue), que la sollicitation de telles données soit considérée comme légitime (pertinence perçue) et enfin que la transaction soit perçue comme juste et équitable, autrement dit qu'il puisse en retirer un gain substantiel (valeur perçue[119]).

[119] Le concept de valeur perçue a finalement été préféré à celui de gain net à répondre. Nous reviendrons sur cette décision dans le paragraphe 4.2.1

Figure 7.6 Modèle conceptuel général de la recherche

= = = = ▶ Hypothèse non testée dans la thèse

Variables de contrôle

- Niveau d'implication vis-à-vis de la catégorie de produits (IMPL)
- Variables sociodémographiques (âge, sexe, CSP, niveau d'éducation)
- Expérience (EXP) et utilisation (UTIL) d'Internet
- Comportement habituel de réponse (COMP_HAB)
- qualité perçue de la relation (QUALITE)

Nous précisons que, dans ce modèle, chaque variable (ou bloc de variables) constitutive du processus de réponse est identifiée par un chiffre et/ou une lettre, de manière à faciliter la numérotation des hypothèses sous-jacentes. Ainsi, les variables « indépendantes » sont repérées par un chiffre de 1 à 6, qui identifie la phase correspondante du processus de réponse et qui se retrouve dans la numérotation de l'hypothèse. Le bloc « exposition à la sollicitation » correspond alors au (1), l'évaluation de la sollicitation au (2), la préoccupation pour le RVP au (3), l'attitude au (4), les IC au (5), et enfin, la réponse à la sollicitation (fait de fournir ou non les données) au (6). De même, concernant les variables « dépendantes »[120], les perceptions sont numérotées de (a) à (d), l'attitude (e), les intentions comportementales de (f) à (h) et enfin la réponse (i). A titre d'exemple, l'hypothèse d'un lien direct entre l'exposition à la sollicitation et la réponse est ainsi nommée « P1 (i) » car elle correspond à l'effet supposé des variables indépendantes du bloc 1 sur la variable dépendante notée i. Une exception notable à ce principe sera toutefois opérée au niveau de la proposition P2 (effet de l'évaluation de la sollicitation sur l'attitude). Cette proposition sera en effet déclinée en 4 hypothèses (notées H2a, H2b, H2c et H2d) correspondant à l'impact des différentes perceptions (identifiées respectivement par les lettres a à d) sur l'attitude.

Concernant ce cadre conceptuel général, deux points principaux appellent à commentaire.

Le premier concerne l'absence du concept de confiance au sein du modèle, alors qu'elle est considérée comme un élément clé, présent dans la plupart des modèles existants[121]. L'importance de la confiance dans le cadre d'un échange, n'est plus à démontrer, et ce, d'autant plus que l'on s'intéresse aux comportements sur Internet. Plusieurs auteurs mentionnent également son influence dans le contexte d'une sollicitation de données

[120] Nous employons les termes « indépendante » et « dépendante » pour identifier une variable ayant respectivement le statut d'antécédent ou de conséquence d'une autre. Certaines variables du modèle sont donc à la fois indépendantes et dépendantes selon le lien considéré. Ainsi, l'attitude est la conséquence de l'évaluation de la sollicitation (variable dépendante) et l'antécédent des intentions comportementales (variable indépendante).

[121] L'intégration de la confiance au sein du modèle aurait notamment pu aisément se concevoir si nous avions retenu la théorie du comportement planifié (TPB), présentée au point 2.3 de ce chapitre. En effet, les travaux de Hoffman et al. (1999) soulignent l'importance du contrôle perçu (variable supplémentaire par rapport au modèle TRA) pour la construction de la confiance dans les environnements électroniques. De plus, il a également été montré, qu'en ligne, la capacité de l'individu à maîtriser les tâches qu'il effectue (proxi du contrôle perçu) est une variable déterminante dans la confiance dont il fait preuve à l'égard des sollicitations qu'il reçoit. Un individu qui pense maîtriser la situation s'estimera alors plus en confiance et se sentira donc moins menacé (au sens de trompé ou d'agressé) par ce type de sollicitations. Dès lors, l'intégration de la variable « contrôle perçu » au sein de notre modèle aurait été un moyen indirect d'inclure le concept de confiance et donc de résoudre le paradoxe de son absence alors même qu'il s'agit d'un élément déterminant dans le cadre d'Internet. La logique aurait alors voulu que nous trouvions un contrôle perçu élevé en situation de confiance. Nous avons toutefois déjà expliqué les raisons de ne pas inclure (nommément) le contrôle perçu au sein de notre modèle. Ce dernier est toutefois présent de façon indirecte, à travers la variable « valeur perçue » (« gain net perçu à répondre »).

personnelles. Hoffman, Novak et Peralta (1999) expliquent ainsi que les consommateurs peuvent décider de ne pas s'engager dans l'échange s'ils n'ont pas confiance dans la capacité des entreprises à honorer leurs engagements, en matière de RVP. En se basant sur le modèle KMV de Morgan et Hunt (1994), ces auteurs font l'hypothèse que les consommateurs ne font pas toujours confiance aux sites parce qu'ils craignent que ces derniers adoptent un comportement opportuniste. Selon ce modèle, le résultat d'une telle situation est une plus faible implication dans la relation, qui génère une plus grande incertitude lors de la prise de décision, une plus faible coopération et une plus grande probabilité de quitter l'entreprise.

Plusieurs raisons expliquent cependant l'absence de la confiance au sein de notre modèle. Tout d'abord, la majorité des auteurs incluant ce concept, le considèrent comme une variable médiatrice entre le stimulus (variables situationnelles) et la réponse. Nous cherchions pour notre part, sans renier les recherches précédentes, à nous démarquer de ce schéma, en travaillant davantage sur le processus conduisant à la réponse (évaluation – attitude – intention comportementale – réponse). De plus, l'impact de la confiance ayant déjà été démontré dans le cadre d'une sollicitation de données personnelles, nous souhaitions davantage nous intéresser aux mécanismes susceptibles de la créer et/ou de la renforcer (i.e. confidentialité perçue des données, faible sensibilité perçue des données sollicitées, etc …). En ce sens, on pourrait considérer la confiance comme un construit de second ordre, les perceptions relatives à la sollicitation constituant les construits de premier ordre. Pour ne pas alourdir notre modèle, déjà compliqué de par sa nature processuelle, nous nous sommes donc contentés de mesurer les quatre variables perceptuelles.
Enfin, nous précisions que la confiance est incluse de manière implicite au sein du modèle, à travers la nature de la relation avec l'entreprise sollicitante. Par nature de la relation, nous entendons en effet, tout élément nous renseignant sur les relations passées entre l'individu sollicité et l'entreprise (ou le site). Ces éléments concernent ainsi l'ancienneté de la relation, le niveau de confiance établi entre les parties, la satisfaction vis-à-vis des expériences passées, etc … Le fait d'inclure la variable confiance à un autre niveau du modèle, ferait alors courir le risque de créer des colinéarités dommageables pour le test des hypothèses.

Le second point appelant à commentaire concerne l'absence de variable modératrice. Là encore, plusieurs raisons expliquent ce choix. Tout d'abord, la littérature, tout comme nos analyses préalables, laissent penser que l'influence éventuelle de variables individuelles et environnementales serait en partie « absorbée » par le concept de préoccupation pour le RVP. De plus, le concept d'attitude est lui-même reconnu comme absorbant les différences

individuelles (Ajzen et Fishbein 1980). L'influence de variables modératrices de type individuel et/ou environnemental est donc, si ce n'est improbable, au moins limitée. De plus, la littérature fait peu état de variables modératrices susceptibles d'influencer la réponse du consommateur face à une sollicitation de ses données personnelles. Aucune des recherches que nous avons analysées ne présente en effet de modèle incluant ce type de variables. Même si nous ne pouvons en conclure qu'elles sont totalement absentes (bien au contraire), force est de constater que nous manquions de matériau solide pour établir des hypothèses dans ce sens. Enfin, étant donné le caractère extrêmement novateur de notre modèle (à notre connaissance, jusqu'à présent, aucun schéma de ce type n'a été testé de manière empirique dans le domaine de la sollicitation de données personnelles), il nous semblait préférable de vérifier sa pertinence et son applicabilité, avant de tester d'éventuels effets modérateurs.
Les éléments de nature individuelle ou environnementale susceptibles d'avoir une influence sur certains liens du modèle ne seront donc pas considérés comme modérateurs mais seront inclus dans les analyses en tant que variables de contrôle. Nous contrôlerons donc l'effet éventuel de ces variables, afin de nous assurer qu'elles n'exercent pas d'influence significative, susceptible de perturber les résultats[122].

Après avoir présenté le modèle conceptuel général de cette recherche, nous indiquons maintenant les propositions et hypothèses de recherche y afférant.

4 Formulation des propositions et hypothèses de recherche

Nous formulons à présent les propositions et hypothèses de recherche qui ont conduit à élaborer ce modèle conceptuel général. Nous distinguons en effet l'hypothèse de recherche (qui porte sur un lien bivarié simple) et la proposition qui permet, dans un premier temps, de ne pas alourdir le modèle en détaillant sous forme d'hypothèses l'ensemble des relations qui la sous-tendent (Debenedetti 2001). Plusieurs propositions (notamment P1 et P6) seront ainsi exposées ci-après et ne seront développées, en hypothèses testables, que dans la partie empirique de notre travail (i.e. chapitres 10, 11 et 12). Les raisons d'un tel choix tiennent principalement au fait que les variables correspondantes seront exposées à cet endroit. La proposition P2 sera, quant à elle, directement traduite en hypothèses, les variables y afférant étant exposées au sein du présent chapitre. Enfin, dans certains cas, nous formulerons immédiatement des hypothèses puisqu'elles portent sur des liens bivariés simples.

[122] Ces variables dites « de contrôle » seront exposées en détail au chapitre suivant.

Au risque de perturber le lecteur, notre modèle conceptuel comprend donc à la fois des hypothèses et des propositions de recherche[123], ces dernières étant amenées à être déclinées par la suite, dans la partie empirique.

Nous serons guidés dans cette démarche à la fois par la revue de littérature et par les résultats de nos études exploratoires préalables, notamment ceux liés à l'étude qualitative. Nous commençons par décliner les propositions concernant l'influence de l'exposition à la sollicitation (au formulaire de données) sur les croyances relatives à la sollicitation, le comportement de réponse et l'intention de poursuivre la relation (4.1). Viennent ensuite les propositions et hypothèses concernant les antécédents de l'attitude face à la communication des données (4.2) suivies de celles relatives à ses conséquences (4.3).

4.1 L'influence de l'exposition à la sollicitation

Le modèle de Olson et Dover (1978) montre clairement le lien entre l'exposition à un stimulus (dans notre cas, l'exposition à la sollicitation de données, par le biais du formulaire à remplir) et les croyances formées par les individus à son sujet. Notre étude qualitative préalable confirme cette idée et dénombre quatre dimensions majeures sur lesquelles les consommateurs évaluent la sollicitation, à savoir la confidentialité, la sensibilité, la pertinence perçue des données sollicitées et l'évaluation du gain net à répondre (valeur perçue).

De plus, la revue de la littérature et l'étude qualitative nous permettent de distinguer quatre catégories d'éléments relatifs au stimulus (facteurs situationnels), parmi lesquelles :

- la politique de RVP mise en place par l'entreprise (le site) et la manière dont elle est valorisée sur le site et/ou sur la page contenant le formulaire à remplir
- les relations existantes entre l'entreprise sollicitant les données et l'individu sollicité (ancienneté de la relation, satisfaction vis-à-vis des expériences antérieures, …)
- le dispositif de collecte (c'est-à-dire tout ce qui a trait au formulaire, dans son fond et dans sa forme) : nombre et type de données demandées, formulation des questions, …
- les circonstances de collecte et notamment le contexte dans lequel les informations sont sollicitées

[123] Ce choix répond, à la fois, à la volonté d'employer un vocabulaire « scientifique » adapté (i.e. appeler proposition un ensemble de relations qui seront amenées à être déclinées par la suite) et de « simplifier » le processus en formulant directement sous forme d'hypothèse ce qui correspond à un lien bivarié simple.

Les résultats de l'étude qualitative montrent que chaque facteur situationnel (i.e. bloc 1) est susceptible d'influencer une ou plusieurs variables perceptuelles à l'origine de l'évaluation de la sollicitation. Ainsi, par exemple, la politique de RVP mise en place par le site aurait une incidence majeure sur le niveau de confidentialité perçu par l'internaute sollicité.
Nous formulons alors la proposition suivante[124] :

Proposition 1 (a à d) : *Les facteurs situationnels influencent l'évaluation par le consommateur de la sollicitation de ses données personnelles en termes de confidentialité (a), de sensibilité (b), de pertinence (c) et de valeur perçues (d).*

Plusieurs résultats issus de la littérature indiquent que les facteurs situationnels pourraient aussi influencer directement la réponse à la sollicitation. Ainsi, Dinev et Hart (2002) montrent que la politique de respect de la vie privée mise en place par l'entreprise (et notamment le fait d'annoncer clairement ses pratiques et de laisser la possibilité au consommateur de contrôler l'utilisation qui sera faite des données) serait bénéfique aussi bien pour développer la confiance que pour encourager ce dernier à donner des informations à caractère personnel. Ce résultat est confirmé par Farag et Krishnan (2003) qui trouvent que les consommateurs sont encouragés à partager leurs données personnelles quand les entreprises les informent des procédures utilisées. L'existence d'un effet direct de la relation entre le consommateur et l'entreprise sollicitant les données sur la réponse à la sollicitation est également suggérée dans la littérature. Ainsi, Schoenbachler et Gordon (2002) montrent que le fait que la relation soit déjà établie encouragerait le consommateur à fournir des renseignements. De même, Hoffman, Novak et Peralta (1999) indiquent que, parmi les premiers éléments empêchant le consommateur de dévoiler des informations personnelles à une entreprise, on trouve la nature de la relation qu'ils entretiennent. Dans le domaine des enquêtes, l'effet du design du questionnaire et de sa longueur (nombre de questions posées) sur le taux de participation a été largement démontré, même si les résultats obtenus sont parfois contradictoires (Heberlein et Baumgartner 1978). Ceci nous laisse supposer un effet identique dans le domaine de la sollicitation de données personnelles, ce qui sera d'ailleurs démontré par Andrade et al. (2002). De plus, ces derniers indiquent qu'au-delà du nombre et du type d'informations, le contexte dans lequel celles-ci sont demandées (achat en ligne vs. accès au site) affecterait également la décision, pour 7% des répondants. Tous ces résultats confortent donc l'idée d'un

[124] Il s'agit d'une proposition car elle ne porte pas sur un lien bivarié simple mais sur une relation entre deux blocs de variables. Pour rappel, sa numérotation (i.e. P1 (a à d)) est liée au fait qu'elle concerne le lien entre les facteurs situationnels (bloc 1) et des variables perceptuelles (identifiées par les lettres a à d sur le modèle).

lien direct entre l'exposition à la sollicitation (facteurs situationnels, bloc 1) et la réponse du consommateur (identifiée par la lettre i), que nous formulons de la manière suivante :

Proposition 1 (i) : *Les facteurs situationnels influencent de manière directe la réponse du consommateur face à une sollicitation de ses données personnelles (i).*

Enfin, nous supposons également l'existence d'un effet direct entre l'exposition à la sollicitation (les facteurs situationnels) et l'intention de poursuivre la relation avec l'entreprise sollicitante. L'étude qualitative révèle en effet que le fait d'être sollicité pour fournir des données personnelles peut être considéré, par certains individus, comme une invasion de leur vie privée et peut alors les conduire à suspendre la relation avec l'entreprise demandeuse. A l'opposé, certains peuvent voir cette sollicitation comme un moyen d'accroître les bénéfices de la relation, ce qui les encourage à continuer de commercer avec la firme. Nous formulons donc la proposition suivante relative à l'effet direct de l'exposition à la sollicitation (bloc 1) sur l'intention de poursuivre la relation (identifiée par la lettre h dans notre modèle) :

Proposition 1 (h) : *Les facteurs situationnels influencent de manière directe l'intention du consommateur de poursuivre la relation avec l'entreprise ayant sollicité les données (h).*

Ces trois propositions relatives à l'effet de l'exposition à la sollicitation sur certaines variables du processus de réponse ne seront traduites en hypothèses que dans les chapitres suivants (i.e. 10, 11 et 12), quand nous aurons identifié les facteurs situationnels réellement étudiés.

4.2 Les antécédents de l'attitude

Nous distinguons, ici, l'influence des variables perceptuelles correspondant à l'évaluation de la sollicitation (4.2.1) de celle liée à la préoccupation pour le respect de la vie privée (4.2.2).

4.2.1 L'influence des variables perceptuelles

Comme le suggèrent le modèle de structure cognitive (Olson et Dover 1978) et la théorie de l'action raisonnée (Fishbein et Ajzen 1975), les croyances (représentées dans notre modèle par les quatre variables perceptuelles), influencent de manière directe l'attitude de l'individu face au comportement étudié. Ceci nous amène à formuler une proposition relative à l'influence de l'évaluation de la sollicitation sur l'attitude face à la communication des données, comme suit :

Proposition 2 : *Plus l'évaluation de la sollicitation (en termes de confidentialité (a), de sensibilité (b), de pertinence (c) et de valeur perçues (d)) est positive (vs négative), plus l'attitude face à la communication des données personnelles est favorable (vs défavorable).*

Cette proposition peut être immédiatement traduite sous la forme d'hypothèses concernant l'effet de chaque perception sur l'attitude. Nous présenterons brièvement ci-après chacune des variables perceptuelles donnant lieu à l'évaluation de la sollicitation, avant d'élaborer les hypothèses relatives à leur influence sur l'attitude face à la communication des données.

L'influence de la confidentialité perçue des données sollicitées

Comme précisé au chapitre 4, le concept de confidentialité a été préféré à celui de contrôle (qu'on trouve souvent dans la littérature), notamment pour des raisons de clarté et de lisibilité du modèle. Ce terme dispose en effet d'une définition précise et est compréhensible par tout un chacun, alors que celui de contrôle est beaucoup plus flou. De plus, le concept de contrôle est relativement large (de nature multidimensionnelle) alors que celui de confidentialité est plus étroit (*a priori* unidimensionnel). Nous préférons donc, pour des raisons de parcimonie, nous en tenir à l'idée de confidentialité. Enfin, et c'est là la raison principale de notre choix, si l'idée de contrôle a été évoquée par les répondants dans l'étude qualitative, ces derniers sont surtout concernés par « la manière dont les informations divulguées seront transmises et utilisées par la suite », qui correspond parfaitement au principe de confidentialité.

Une fois les données fournies, l'individu n'a plus de pouvoir quant à leur utilisation. Dès lors, s'il ne croit pas que l'organisation qui le sollicite en assurera la confidentialité, il pourra estimer qu'il ne faut pas prendre le risque de coopérer avec elle (Mayer 2002). Moore et McDonald (1987) montrent d'ailleurs que l'ignorance ou une croyance négative quant à la confidentialité des données divulguées contribue à la formation d'une opinion défavorable.
Les études passées indiquent que les consommateurs ont moins de chance de percevoir un risque et sont donc plus enclins à se dévoiler quand ils perçoivent une possibilité de contrôler l'utilisation de leurs données (Culnan et Armstrong 1999, Bies 1993, Stone et Stone 1990). De plus, la théorie de la motivation (espérance attendue) postule que le contrôle décroît le risque perçu d'une action, en augmentant la probabilité que cette action soit suivie du résultat attendu. La confidentialité perçue correspond donc à la partie valeur attendue de la théorie, dans la mesure où un niveau élevé de confidentialité conduit l'individu à penser que son

action conduira au résultat espéré. Ceci accroît alors la motivation de l'individu à s'engager dans cette action (fournir ses données). Ceci nous permet de formuler l'hypothèse suivante :

Hypothèse 2 (a) : *Plus la confidentialité des données sollicitées (a) est perçue comme élevée (vs faible), plus l'attitude face à leur communication est favorable (vs défavorable).*

L'influence de la sensibilité et de la pertinence perçues des données sollicitées

L'étude qualitative fait ressortir l'importance de la sensibilité et de la pertinence perçues des données sollicitées dans l'évaluation que l'individu fait de la sollicitation, et donc dans sa décision finale de fournir (ou non) les informations. Bien que l'influence de ces deux concepts soit clairement évoquée au sein de la littérature, elle n'a jamais été vraiment testée, si ce n'est indirectement. La sensibilité perçue est ainsi régulièrement étudiée à travers la « manipulation » du type de données demandées mais n'a jamais été mesurée en tant que telle. De même, la pertinence perçue ressort comme un concept important pour les chercheurs étudiant ce phénomène, mais son influence n'a jamais été mesurée. Seule la pertinence de l'utilisation des données a été testée jusqu'ici. Ainsi, la pertinence des relances de marketing direct affecterait la manière dont le consommateur va se sentir préoccupé par l'utilisation de ses données. Comme l'indique Sherman (1991), si un consommateur est intéressé par un sujet et reçoit un coup de téléphone ou un courrier non sollicité à ce propos, il ne le verra pas comme une invasion de sa vie privée, dans la mesure où l'offre est pertinente.

Parmi les facteurs importants lors de la prise de décision de dévoiler ou non les informations demandées, on trouve donc le fait que les données ne soient pas jugées comme trop sensibles (Cranor, Reagle et Ackerman 1999). Ainsi, dans le domaine des sondages, Singer (1984) montre que le fait de considérer que certaines questions ne concernent pas l'entreprise est significativement lié à un comportement de non-réponse ou d'évitement.

La pertinence perçue est un autre critère important. Plusieurs recherches antérieures identifient clairement son impact sur le processus de réponse. Ainsi, Hine et Eve (1998) notent que toute demande d'informations non considérées comme nécessaires en vue de conduire une action apportant un bénéfice au consommateur est vécue comme intrusive. Culnan et Armstrong (1999) confirment cette idée selon laquelle la pertinence des données sollicitées est une des conditions (nécessaires) pour que la collecte ne soit pas perçue comme une invasion (de la vie privée). Hine et Eve (1998) concluent que les sollicitations jugées non pertinentes s'accompagnent généralement d'hostilité. Ce résultat est confirmé par Wang et Petrison (1993) qui montrent que les consommateurs rejettent les actions non pertinentes. Une

des explications à cela est le fait que l'individu a tendance à se poser la question des raisons amenant l'entreprise à collecter ces données (Patterson, O'Malley et Evans 1997). Or, sans information précise à ce sujet, il a tendance à imaginer des raisons peu « louables » (i.e. constitution et revente de fichiers, relances ultérieures non sollicitées). Ceci l'amène alors à développer une opinion défavorable à ce sujet.

Sensibilité et pertinence perçues correspondent à la partie instrumentale de la théorie de la motivation (ou de l'espérance attendue). En effet, les personnes estimant que les données demandées sont (trop) sensibles et/ou ne paraissent pas cohérentes par rapport à l'objectif affiché (absence de pertinence) pourront ne pas croire que le fait de se dévoiler mènera au résultat attendu. Ceci est de nature à réduire leur motivation à répondre et nous conduit à formuler les hypothèses suivantes :

Hypothèse 2 (b) : *Plus la sensibilité des données sollicitées (b) est perçue comme élevée (vs faible), moins l'attitude face à leur communication est favorable (vs défavorable).*

Hypothèse 2 (c) : *Plus la pertinence des données sollicitées (c) est perçue comme élevée (vs faible), plus l'attitude face à leur communication est favorable (vs défavorable).*

L'influence des coûts et bénéfices perçus à répondre (notion de valeur perçue)

Le gain net à répondre (différence entre les coûts et bénéfices perçus) est l'une des (si ce n'est la) variables les plus importantes dans l'évaluation de la sollicitation. Ce construit se rapproche du concept d'équité, défini comme une attente d'échange équilibré entre les parties (Childers et Ruekert 1982). Ce dernier est régulièrement utilisé dans la littérature pour représenter le rapport coût/bénéfice. Il correspond en effet à l'idée selon laquelle les individus analysent ce qu'ils sont susceptibles de perdre et de gagner avant d'adopter un comportement spécifique (dans notre cas avant de fournir leurs données). Développée à l'origine dans le domaine psychologique, la théorie de l'équité (Adams 1963) est utilisée pour expliquer les comportements, dans une variété de contextes. Cette théorie explique que lorsque deux acteurs sont parties prenantes d'un échange, ils comparent leurs contributions et gains respectifs. Si un acteur perçoit que l'échange est déséquilibré, cela crée chez lui une tension qu'il cherchera à éliminer en n'adoptant pas le comportement prévu. De plus, c'est le gain relatif (ou la perte) comparé aux contributions qui détermine l'équité. Ce concept est également fréquemment utilisé dans les études sur la participation à un sondage. Les recherches dans ce domaine suggèrent ainsi que l'équité doit être préservée pour faciliter un échange réussi et encourager la participation. De plus, il a été démontré que si le répondant a

une idée de compensation adéquate en échange de sa participation, alors toute violation de celle-ci conduira à un jugement d'inéquité, pouvant l'amener à refuser de répondre.
Nos entretiens semi-directifs menés auprès d'internautes français montrent que, dans le cadre d'une sollicitation de données personnelles en ligne, l'individu procède à une évaluation essentiellement « interne » de la sollicitation, plutôt que de comparer ses contributions et gains respectifs à ceux de l'entreprise. Ce phénomène peut probablement se justifier par le fait que les internautes sont centrés sur eux quand ils surfent sur le web. De plus, évaluer les gains et les coûts pour l'entreprise est d'une part difficile et d'autre part de faible intérêt (ce qui intéresse l'individu est de savoir ce qu'il peut gagner). L'évaluation porte donc principalement sur le gain espéré, comparé aux coûts liés au remplissage du formulaire.

Le compromis entre les bénéfices et les coûts liés à la fourniture d'informations personnelles a été étudié sous le terme de « *privacy calculus* », qui mesure les bénéfices du dévoilement contre les possibles conséquences négatives d'un tel choix, notamment une utilisation ultérieure non autorisée des données divulguées (Laufer et Wolfe 1977, Milne et Gordon 1993, Stone et Stone 1990). Récemment, Awad et Krishnan (2006) ont conceptualisé ce compromis comme la différence entre les coûts et bénéfices liés au dévoilement.

Dans une approche économique et traditionnelle, la valeur perçue est une notion subjective qui correspond à un surplus perçu de bénéfices, déduction faite des coûts monétaires et psychologiques liés à la réalisation d'une transaction. Ce concept intervient généralement comme une variable explicative du comportement d'achat et repose uniquement sur une confrontation cognitive (ratio ou différence) entre ce qui est reçu (utilité, qualité, bénéfices ou outputs) et ce qui est donné (prix, coûts, investissement, sacrifices ou inputs) (Zeithaml 1988, Bolton et Drew 1991, Wooddruff 1997). Zeithaml (1988) juge que le consommateur conçoit la valeur comme une fonction positive d'éléments tangibles donnés et comme une fonction négative des sacrifices financiers et/ou non financiers consentis ; il s'agit du jugement global de l'offre basé sur la perception de ce qui est reçu en comparaison de ce qui est donné.
D'un point de vue théorique, la valeur est donc un arbitrage entre des coûts et des bénéfices. C'est la raison pour laquelle nous utiliserons ce concept de valeur perçue pour mesurer l'arbitrage réalisé par les individus confrontés à une sollicitation de données personnelles entre les coûts et bénéfices liés au dévoilement.

Selon Phelps, Nowak et Ferrell (2000), une des motivations importantes dans la décision de fournir des données personnelles est l'existence d'un intérêt personnel à répondre. Or, le fait de percevoir plus de bénéfices que de coûts en fournissant les données demandées indique que l'individu estime qu'il est plus intéressant de répondre que de se taire. De même, dans le domaine de la participation à une enquête, Sharp et Frankel (1983) déduisent que les gens qui estiment retirer un bénéfice direct de la participation sont moins nombreux à percevoir la participation comme un fardeau et sont donc plus enclins à répondre.
L'impact de la valeur perçue sur l'attitude, bien que non démontré empiriquement, est sous-entendu dans de nombreuses recherches. Ainsi, dans le secteur de la téléphonie, Bolton et Drew (1991) trouvent que la valeur (bénéfices moins coûts) est un déterminant significatif des intentions comportementales des consommateurs à rester fidèle à leur opérateur. De même, Phelps, Nowak et Ferrell (2000) montrent que les conséquences futures du dévoilement (en termes de coûts et bénéfices) ont un effet sur les croyances relatives aux pratiques des entreprises en matière de traitement d'informations.

La valeur perçue à répondre (différence entre les coûts et bénéfices liés au dévoilement) correspond à la partie valence (attractivité du résultat) de la théorie de la motivation (espérance attendue). Hann et al. (2005) prouvent ainsi que les résultats attendus du dévoilement d'informations, sous la forme de récompenses monétaires ou de gain de temps, sont associés à des valences positives qui accroissent le score de motivation. Un surplus perçu de bénéfice pourrait donc encourager le consommateur à répondre, tandis qu'un surcroît de coût pourrait au contraire le décourager. Ceci nous amène à formuler l'hypothèse suivante :

Hypothèse 2 (d) : *Plus la valeur perçue à répondre (d) est élevée (vs faible), plus l'attitude face à la communication des données sollicitées est favorable (vs défavorable).*

4.2.2 L'influence de la préoccupation pour le RVP

Par préoccupation pour le Respect de la Vie Privée (RVP), nous entendons le niveau d'inquiétude ressenti par l'individu concernant les menaces qui pèsent sur sa vie privée lorsque des entreprises sollicitent et/ou utilisent ses données personnelles. De nombreux auteurs rapprochent ce concept de l'attitude du consommateur face aux pratiques de traitement des données (limitées à la collecte et à l'utilisation des données).

Cette préoccupation devrait affecter à la fois l'attitude de l'individu face à la communication de ses données (hypothèse H3 e) et son comportement de réponse (proposition P3 i)[125].

Les « valeurs » personnelles, comme la préoccupation pour le RVP, affecte la partie valence de la théorie de la motivation c'est-à-dire l'attractivité du résultat attendu pour l'individu (la valeur que le consommateur associe au résultat de son dévoilement). Les gens très préoccupés percevront alors le résultat attendu comme étant de moindre valeur que les gens peu préoccupés et leur motivation à se dévoiler sera donc plus faible. De plus, Hann et al. (2005) indiquent que les méthodes visant à réduire les préoccupations pour le RVP sont associées à des valences positives, qui se traduisent par une motivation plus élevée à répondre.
Nous en déduisons l'hypothèse suivante relative à l'effet de la préoccupation pour le RVP sur l'attitude face à la communication des données (symbolisée par la lettre e) :

Hypothèse 3 (e) : *Plus la préoccupation pour le RVP est élevée (vs faible), moins l'attitude face à la communication des données sollicitées (e) est favorable (vs défavorable).*

Au sein de la littérature, la préoccupation pour le RVP est aussi associée aux variables aval du processus de réponse, notamment aux intentions comportementales et aux comportements. Ainsi, Smith, Milberg et Burke (1996) notent qu'un haut niveau de préoccupation est associé à de fortes intentions comportementales de dévoiler les informations demandées. Stewart et Segars (2002) trouvent, pour leur part, que les individus à la préoccupation élevée ont une plus forte tendance à retirer leur nom de fichiers et à refuser de divulguer des données. Enfin, les résultats obtenus par Farag et Krishnan (2003) confirment que la préoccupation pour le RVP est associée négativement à l'acceptation de fournir des informations personnelles. Ils montrent ainsi que les individus les plus préoccupés sont les moins enclins à accepter de partager leurs données dans un but de « profilage » (en anglais, *profiling*).
Nous en déduisons la proposition suivante relative à l'effet direct de la préoccupation pour le RVP sur la réponse à la sollicitation (notée i dans notre modèle) :

Proposition 3 (i) : *Plus la préoccupation pour le RVP est élevée (vs faible), moins (vs plus) l'individu acceptera de communiquer ses données personnelles (i).*

[125] Dans le second cas, il s'agit bien d'une proposition puisque la réponse à la sollicitation (i) sera mesurée de différentes manières et correspondra donc à plusieurs variables, présentées au chapitre 12.

4.3 Les conséquences de l'attitude

Les théoriciens de l'attitude soulignent que ce concept mesure la réaction ou l'évaluation de l'individu à l'égard d'un objet (ici à l'égard de la communication de données personnelles) sur un continuum favorable-défavorable ou aime-déteste (Fishbein et Ajzen 1975). Suivant Triandis (1971), on peut voir l'attitude comme un concept ayant 3 composantes différentes : cognitive (connaissance et perception de l'objet), affective (sentiments ou réaction émotionnelle à l'égard de l'objet), comportementale (prédisposition à agir vis-à-vis de l'objet). Dans cette recherche, les perceptions (correspondant à une évaluation cognitive de la situation) se situent en amont et les intentions comportementales en aval. Nous considérons donc l'attitude comme un concept unidimensionnel, essentiellement affectif, une évaluation globale de l'acte qui consiste à communiquer des données personnelles[126].
Le modèle TRA démontre clairement l'effet de l'attitude sur les intentions comportementales. Deux intentions comportementales (IC) de réponse sont distinguées dans cette recherche : il s'agit de l'IC de répondre (fournir les données personnelles demandées en complétant le formulaire) et de l'IC de mentir (fournir des données erronées). Si la première est évidente, la présence de la seconde au sein du modèle amène quelques commentaires.

Tout d'abord, le fait que le consommateur accepte de répondre aux questions ne signifie pas que les réponses qu'il donne sont justes. Fry (1996) montre en effet que, lorsqu'on les oblige à s'enregistrer, certains individus peuvent être amenés à fournir de fausses informations (fausse adresse, faux numéro de téléphone, …). Quand on les interroge à ce sujet, les répondants expliquent qu'ils falsifient les données dans plusieurs cas, et notamment quand : 1) les conditions d'utilisation des données ne sont pas (ou mal) spécifiées (absence d'information) ; 2) le fait de révéler l'information n'apparaît pas nécessaire (faible pertinence) ; 3) ils ne croient pas que le site saura garder leurs données (faible confidentialité perçue) ; 4) les données sont considérées comme trop sensibles (sensibilité perçue).
Les conséquences d'un tel comportement peuvent être dommageables pour les entreprises pour 2 raisons : 1) elles peuvent prendre de mauvaises décisions du fait des informations erronées ; 2) cela peut les empêcher de bâtir une relation saine avec les consommateurs. Il semblait donc intéressant, au-delà de l'intention comportementale de répondre, d'évaluer

[126] Sans entrer dans le débat qui divise les chercheurs concernant la vision uni ou tri (voire bi) dimensionnelle de l'attitude, nous faisons donc clairement le choix de la première au détriment des deux autres. Nous sommes conscients qu'en cela, nous privilégions la simplification. Ce choix est cependant justifié par le fait que plusieurs recherches concluent au caractère plus robuste de l'approche unidimensionnelle, notamment du fait des inter-corrélations qui existent lorsqu'on mesure simultanément les trois composantes (Breckler 1984).

également l'intention de mentir. Outre le fait qu'il s'agit d'intention et non de comportement réel, les résultats devront cependant être analysés avec prudence : comment être sûr en effet que quelqu'un qui avoue avoir l'intention de mentir dise la vérité ?

Justifions et formulons à présent les propositions concernant l'effet de l'attitude envers ces deux types d'intentions comportementales, l'intention de répondre (4.3.1) et de mentir (4.3.2).

4.3.1 L'effet de l'attitude sur l'intention de répondre

Dans le domaine de la participation à un sondage, Goyder (1986) note qu'une attitude défavorable au fait de compléter le questionnaire correspond à une forte probabilité de refus de participer. Hornback (1971) indique que plus l'attitude envers les éléments du sondage est favorable, plus le répondant se trouve engagé vis-à-vis de tout ce qui concerne ce sondage et plus il aura tendance à accepter de participer.
En appliquant ces résultats au dévoilement de données personnelles, on peut formuler l'hypothèse suivante relative à l'effet de l'attitude sur l'intention de répondre (notée f) :

Hypothèse 4 (f) : *Plus l'attitude face à la communication des données sollicitées est favorable (vs défavorable), plus (vs moins) l'individu a l'intention de répondre (de communiquer les données personnelles demandées) (f).*

4.3.2 L'effet de l'attitude sur l'intention de mentir

Dans la littérature, la confiance est souvent considérée comme un moyen d'éviter le mensonge, en réduisant les risques liés à la divulgation d'informations personnelles (Culnan et Armstrong 1999). Dès lors, comme le confirment Lwin et Williams (2003), dans un environnement où la confiance est faible, mentir est un moyen pour le consommateur d'obtenir les bénéfices associés au dévoilement, sans s'exposer aux risques. Un consommateur ayant une attitude défavorable face à la communication de ses données peut donc accepter de répondre mais fournir des réponses erronées, afin de limiter les risques associés au dévoilement.
On peut alors formuler l'hypothèse suivante concernant l'effet de l'attitude sur l'intention de mentir (symbolisée par la lettre g dans notre modèle) :

Hypothèse 4 (g) : *Plus l'attitude face à la communication des données sollicitées est favorable (vs défavorable), moins (vs plus) l'individu a l'intention de mentir (de fournir des données erronées) (g).*

4.4 Les conséquences sur l'intention de poursuivre la relation

Au-delà de l'intention de répondre et de l'intention de mentir, nous avons choisi de mesurer une autre intention comportementale : l'intention de poursuivre la relation avec l'entreprise sollicitante. Nous souhaitons en effet mesurer l'impact de la sollicitation sur les liens unissant l'entreprise à l'individu sollicité. Nous suivons en cela le conseil de Zeithalm, Berry et Parasuraman (1996) qui insistent sur l'importance de mesurer les intentions comportementales futures des consommateurs pour évaluer leur potentiel à rester ou à quitter l'entreprise (la marque). Nous pensons que le fait de solliciter une personne pour qu'elle accepte de fournir des données personnelles (dans le but de maintenir ou de développer une relation forte avec elle) n'est pas forcément neutre. Elle peut en effet estimer que le fait de fournir ces informations est bénéfique car cela lui permettra d'être davantage considérée et donc de bénéficier des avantages de la politique de marketing relationnel développée par l'entreprise. Elle peut aussi voir la sollicitation comme une invasion de sa vie privée ce qui peut avoir une incidence sur son niveau de confiance vis-à-vis de l'entreprise.

La volonté du consommateur de poursuivre la relation avec l'entreprise correspond au concept d'engagement. Bennett et Gabriel (2001) définissent ainsi l'engagement comme la croyance selon laquelle la relation (avec l'entreprise) continuera à l'avenir. Ils signalent que ce concept peut être mesuré à la fois par l'intention de continuer la relation et par la volonté d'entreprendre des investissements spécifiques à celle-ci (Pels 1992). De même, le comportement qui vise à rester dans l'échange est l'indicateur traditionnel de l'engagement dans le domaine organisationnel (Halaby et Weakliem 1989, Mueller et Price 1990).
Nous verrons, dans un premier temps, les principes de la théorie de l'engagement (4.4.1) avant de formuler des hypothèses relatives à l'effet de la sollicitation sur l'intention du consommateur de poursuivre la relation avec l'entreprise qui l'a sollicité (4.4.2 et 4.4.3).

4.4.1 La théorie de l'engagement

Nous commencerons par présenter l'origine du concept d'engagement avant d'aborder les définitions et les différentes formes d'engagement décrites dans la littérature. Nous évoquerons ensuite les principes relatifs à la théorie de l'engagement.

L'origine du concept d'engagement

Le concept d'engagement apparaît au cœur des recherches traitant des relations interpersonnelles (Thibaut et Kelley 1959), des relations entre les employés et les organisations (Allen et Meyer 1990, Meyer et Allen 1991), des relations entre les firmes (Dwyer et al. 1987, Anderson et Weitz 1992, Moorman et al. 1992, Morgan et hunt 1994, Gundlach et al. 1995) et, plus récemment, des relations entre les firmes et les consommateurs (Storbacka et al. 1994, Gruen 1995). Ce terme vient à l'origine des ressources humaines et représente la volonté d'un individu de maintenir sa relation avec un objet, sur la base de son attachement envers celui-ci. La notion d'engagement du client envers une marque (ou une entreprise) a été introduite dans l'étude du comportement du consommateur par Day (1969) qui lui donne plusieurs définitions possibles : une décision rationnelle de fidélité à une marque, une habitude d'achat, une orientation affective forte en faveur de cette marque.

La notion d'engagement est également centrale dans la littérature sur l'échange social (Blau 1964, Thibaut et Kelley 1959), ce qui justifie d'autant plus son utilisation, de par l'importance de cet ancrage théorique dans le cadre de notre travail. Cook et Emerson (1978) caractérisent ainsi l'engagement comme une variable centrale permettant de distinguer l'échange social de l'échange économique. Ainsi, parce que l'engagement entraîne une certaine vulnérabilité, les parties vont seulement chercher des partenaires dignes de confiance. La théorie de l'échange social explique cette relation causale à travers le principe de réciprocité (évoqué au chapitre 3), qui soutient que « la méfiance engendre la méfiance et sert aussi à diminuer l'engagement » (McDonald 1981, cité par Farissi Sainz 1997).

Les définitions du concept d'engagement

En psychologie, le terme d'engagement fait référence à deux concepts différents. D'une part, certains psychologues font l'hypothèse que c'est l'individu lui-même qui s'engage, de sa propre volonté, en fonction de ses représentations, attitudes et motivations. L'engagement est alors qualifié d'interne puisqu'il dépend uniquement des caractéristiques individuelles. Il est à noter que cette perspective se rapproche de l'acception dans le langage commun selon laquelle « être engagé » signifie « être impliqué ». D'autre part, les psychologues sociaux affirment que ce qui engage un individu est son comportement et que le niveau d'engagement dépend de la situation dans laquelle l'acte se produit. L'engagement est alors dit « externe ».

Dans le domaine du marketing, les chercheurs ont d'abord vu dans l'engagement un acte prenant la forme d'une contribution (en anglais, *input*) ou d'un gage (en anglais, *pledge*) qui témoigne de l'intérêt personnel que le client prend dans la relation (Gundlach, Achrol et Mentzer 1995). Aujourd'hui, la majorité (si ce n'est la totalité) des chercheurs en comportement du consommateur sont d'accord pour dire que l'engagement peut se définir comme l'intention du consommateur de poursuivre une relation suivie avec la marque. De nombreux auteurs considèrent ainsi ce construit comme une des conditions nécessaires à l'apparition de la fidélité. De plus, l'engagement implique une volonté de faire des concessions à court terme pour réaliser des bénéfices sur le long terme. Il s'agit donc d'un état psychologique qui reflète l'intensité du lien entre l'acteur et son partenaire.

Nous ne détaillerons pas l'ensemble des définitions données par les chercheurs ayant travaillé sur ce concept. Il existe en effet plus d'une vingtaine de significations, très proches les unes des autres mais qui diffèrent souvent sur un ou deux points. Nous reprendrons plutôt la typologie proposée par Frisou (2000) qui invite à retenir deux définitions principales de l'engagement. La première est une définition cognitiviste, à laquelle se réfère Morgan et Hunt (1994) : il s'agit de la volonté du client de maintenir, au prix d'un effort personnel, une relation qu'il juge importante et dont la valeur croît en fonction de la durée, ce que Frisou (2000) qualifie d'engagement explicite. La seconde est plutôt d'inspiration behavioriste et renvoie aux travaux de Kiesler (1971) : il s'agit de l'engagement implicite (un concept proche de l'engagement externe, défini plus haut) qui peut être défini comme la connaissance que le client a de ses comportements d'achat envers le partenaire. Il est question ici des comportements adoptés par l'individu pour « prouver » son engagement. Cette seconde définition rend bien compte de l'aspect aussi bien économique que psychologique de l'engagement. Gurviez, en 1998, faisait déjà cette distinction entre engagement implicite et explicite en définissant l'engagement comme une intention implicite ou explicite de maintenir une relation durable avec la marque.

Les différentes formes d'engagement

Il existe, au sein de la littérature, différentes typologies de formes d'engagement.
Plusieurs chercheurs, dont Lacoeuille (2000), Henning-Thurau et al. (2002) et bien d'autres, différencient ainsi l'engagement calculé (cognitif ou calculatoire) qui traduit une disposition à adopter un comportement, fondé sur une décision économique rationnelle (calcul des profits

associés au fait de rester et des coûts associés au fait de partir) et l'engagement affectif qui correspond plus à un désir d'appartenance et à une volonté d'adopter les valeurs et les objectifs du partenaire. D'autres chercheurs comme Allen et Meyer (1990) ou Gundlach, Achrol et Mentzer (1995) penchent plutôt pour un modèle à trois composantes, les deux premières correspondant aux catégories données ci-dessus, la dernière évoluant d'un auteur à l'autre. Ainsi, quand Allen et Meyer (1990) relèvent 3 natures différentes d'engagement suivant que la relation constitue pour l'individu : une nécessité (calculé), un désir (affectif) ou une obligation (normatif), Gundlach et al. (1995) distinguent, quant à eux, une composante instrumentale, calculée ou comportementale (basée sur les investissements passés), une composante affective ou attitudinale (lien affectif, attachement) et une composante temporelle (fondée sur l'idée que la relation durera toujours). Cette typologie se rapproche de celle proposée par Frisou entre l'engagement implicite (proche de la composante comportementale) et explicite (correspondant à la composante temporelle).

Les principes de la théorie de l'engagement

Le fondateur de la théorie psychosociale de l'engagement est Kiesler (1971), qui définit l'engagement comme le lien qui unit l'individu à ses actes comportementaux. Cette définition implique que seuls les actes engagent et ce, à des degrés divers. Comme le précisent Beauvois et Joule (1981) : « l'engagement correspond, dans une situation donnée, aux conditions dans lesquelles la réalisation d'un acte ne peut être imputable qu'à celui qui l'a réalisé ».

Selon Kiesler (1971), les individus sont engagés parce qu'ils trouvent *a posteriori*, dans les actes qu'ils ont commis, des certitudes qui les justifient (principe de rationalisation). Deux modèles distincts peuvent alors expliquer l'engagement du client dans une relation d'échange. Dans le premier, le client s'engage parce que sa confiance à l'égard du partenaire dissipe l'incertitude qu'il ressent. Dans le second, le client assume seul l'incertitude et se résout à maintenir la relation parce qu'il se sent déjà engagé par ses actes antérieurs.

Nous avons vu que le concept d'engagement correspondait à celui d'intention de poursuivre la relation que nous souhaitons mesurer au cours de cette recherche. Les principes de la théorie de l'engagement ainsi que les recherches menées sur le sujet vont nous permettre de formuler les hypothèses sous-jacentes. Nous commencerons par celle liée à l'effet de l'intention de répondre (H5) avant d'évoquer celle liée au comportement de réponse (P6).

4.4.2 L'effet de l'intention de répondre

Dans la littérature sur la participation à un sondage, l'engagement est représenté par l'intention comportementale de répondre de la manière prédite (Childers et Skinner 1979). Ces auteurs précisent même que l'obtention de l'engagement du répondant est une condition nécessaire à sa coopération. Il s'agit d'un engagement « mental » qui est censé se manifester à travers une intention de participer. On peut dès lors considérer l'intention de participer (dans notre cas de fournir ses données personnelles) comme une forme d'engagement. Or, si la théorie prédit que seuls les actes engagent, Helme-Guizon et Amato (2004) précisent que « même un acte aussi simple qu'exprimer une attitude peut conduire à de l'engagement ». L'intention de répondre constituant la phase postérieure à la formation de l'attitude, il est possible d'imaginer qu'elle aura une incidence sur l'engagement, vu comme l'intention de poursuivre la relation avec l'entreprise sollicitante. Nous en déduisons l'hypothèse suivante :

Hypothèse 5 : *Plus (vs moins) l'individu a l'intention de communiquer les données demandées, plus (vs moins) il a l'intention de poursuivre la relation avec l'entreprise qui l'a sollicité.*

Bien que nous soupçonnions l'existence d'une relation entre l'intention de mentir (IM) et l'intention de poursuivre la relation (IPR) avec l'entreprise ayant sollicité les données, nous ne formulerons pas d'hypothèse à ce sujet, et ce, pour plusieurs raisons. Tout d'abord, nous manquons de matériau théorique solide pour justifier une telle hypothèse. La mesure du mensonge (en l'occurrence ici, de l'intention de mentir) est en effet très peu étudiée au sein de la littérature. De plus, nous ne savons pas s'il est possible de qualifier l'intention de mentir (ou à l'opposé de dire la vérité) comme une forme d'engagement. Comment en effet être certain qu'un individu prêt à mentir dit bien la vérité ? L'intention de mentir peut, en outre, être considérée, par certains individus, comme une stratégie « moyenne » permettant de participer à l'échange, sans réellement s'engager. Comment dans ce cas justifier le fait de remplir un formulaire et d'y indiquer de fausses informations si on n'a pas l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données ?

Le flou quand au lien exact entre ces deux variables[127], nous amène donc à ne pas formuler d'hypothèse à ce sujet, tout en reconnaissant l'intérêt qu'il y aurait à le faire. Des recherches spécifiques sur ce thème seraient donc particulièrement bienvenues.

[127] Le modèle à tester en figure 7.6 contient donc « abusivement » un lien entre les intentions comportementales de réponse (IR et IM) et l'intention de poursuivre la relation alors qu'il devrait concerner uniquement la première relation (i.e. lien entre l'intention de répondre et celle de poursuivre la relation).

4.4.3 L'effet du comportement de réponse

Dans son article, Frisou (2000) cherche à démontrer que la volonté du client de s'engager dans une relation dépend autant de ses facultés cognitives à élaborer une relation de confiance que des comportements qu'il a eus précédemment (effet d'engagement). Il montre ainsi que les comportements d'achat du client (engagement implicite) ont une influence positive sur le désir de ce dernier de maintenir une relation durable (engagement explicite). De même, Gundlach, Achrol et Mentzer (1995) montrent que l'engagement calculé (inputs engagés dans la relation) a un effet en spirale : il renforce l'attitude et l'engagement de long terme (engagement temporel). Ces auteurs mesurent d'ailleurs l'engagement calculé sous la forme de divulgation d'informations considérées comme confidentielles.

Au cours de sa navigation, l'internaute est amené à réaliser de nombreux actes différents, du plus simple (par exemple, cliquer sur un lien hypertexte) au plus élaboré (comme remplir un formulaire). Selon Helme-Guizon et Amato (2004), certains de ces actes devraient induire un effet d'engagement. Proposer à l'internaute de participer à un jeu, de recevoir une lettre d'information, d'être tenu au courant des nouveautés sur le site ou d'avoir accès à un contenu valorisé sont autant de moyens d'amorcer la relation. L'engagement sera ainsi initié si le visiteur accepte de remplir un formulaire, de préférence long (et donc coûteux en temps et en charge cognitive). Les étapes préalables à l'achat (par exemple : cliquer à plusieurs reprises ou donner des informations personnelles) constituent donc des actes préparatoires, producteurs d'engagement. Ceci nous permet de postuler que le fait pour l'individu d'accepter de fournir ses données personnelles aura une incidence sur sa volonté de s'engager dans une relation durable avec l'entreprise l'ayant sollicité.
Nous formulons la proposition[128] correspondante de la façon suivante :

Proposition 6 : *Plus (vs moins) l'individu accepte de communiquer les données demandées, plus (vs moins) il a l'intention de poursuivre la relation avec l'entreprise qui l'a sollicité.*

[128] Nous précisons, là encore, qu'il s'agit bien d'une proposition puisque l'acceptation de communiquer les données demandées sera mesurée de différentes façons et correspondra donc à différentes variables exposées dans le chapitre 12. Cette proposition sera donc déclinée en hypothèses à ce niveau.

L'ensemble des propositions et hypothèses de recherche qui viennent d'être exposées sont récapitulées dans le tableau ci-après (cf. tableau 7.1).

Nous précisons à nouveau que les propositions P1, P3 et P6 seront traduites en hypothèses dans les chapitres suivants. Les hypothèses relatives à P1, correspondant à l'impact des variables situationnelles, seront ainsi développées dans les trois derniers chapitres de ce document (10, 11 et 12). Les propositions P3(i) et P6 seront, pour leur part, déclinées en hypothèses dans le chapitre 12, où nous exposerons différentes mesures de la « réponse », correspondant à la variable dépendante majeure de notre modèle (identifiée par la lettre i).

Tableau 7.1 Récapitulatif des propositions et hypothèses de recherche

N°	Propositions et Hypothèses
P1	Les facteurs situationnels influencent l'évaluation par le consommateur de la sollicitation de ses données personnelles en termes de confidentialité (a), de sensibilité (b), de pertinence (c) et de valeur perçues (d).
P1 (i)	Les facteurs situationnels influencent de manière directe la réponse du consommateur face à une sollicitation de ses données personnelles (i).
P1 (h)	Les facteurs situationnels influencent de manière directe l'intention du consommateur de poursuivre la relation avec l'entreprise ayant sollicité les données (h).
P2	Plus l'évaluation de la sollicitation (en termes de confidentialité (a), de sensibilité (b), de pertinence (c) et de valeur perçues (d)) est positive (vs négative), plus l'attitude face à la communication des données personnelles est favorable (vs défavorable).
H2[129] (a, b, c et d)	**2a** : Plus la confidentialité des données sollicitées (a) est perçue comme élevée (vs faible), plus l'attitude face à leur communication est favorable (vs défavorable). **2b** : Plus la sensibilité des données sollicitées (b) est perçue comme élevée (vs faible), moins l'attitude face à leur communication est favorable (vs défavorable). **2c** : Plus la pertinence des données sollicitées (c) est perçue comme élevée (vs faible), plus l'attitude face à leur communication est favorable (vs défavorable). **2d** : Plus la valeur perçue à répondre (d) est élevée (vs faible), plus l'attitude face à la communication des données sollicitées est favorable (vs défavorable).
H3 (e)	Plus la préoccupation pour le RVP est élevée (vs faible), moins l'attitude face à la communication des données sollicitées (e) est favorable (vs défavorable).
P3 (i)	Plus la préoccupation pour le RVP est élevée (vs faible), moins (vs plus) l'individu acceptera de communiquer ses données personnelles (i).
H4 (f, g)	Plus l'attitude face à la communication des données sollicitées est favorable (vs défavorable), plus (vs moins) l'individu a l'intention de répondre (de communiquer les données demandées) (f) et moins (vs plus) il a l'intention de mentir (g)
H5	Plus (vs moins) l'individu a l'intention de communiquer les données demandées, plus (vs moins) il a l'intention de poursuivre la relation avec l'entreprise qui l'a sollicité.
P6	Plus (vs moins) l'individu accepte de communiquer les données demandées, plus (vs moins) il a l'intention de poursuivre la relation avec l'entreprise qui l'a sollicité.

[129] Pour rappel, H2 (a, b, c et d) correspond à la déclinaison de la proposition P2 (présentée au-dessus) en hypothèses testables (i.e. un ensemble de liens directs entre deux variables).

Conclusion du chapitre 7

Ce chapitre visait à présenter le cadre conceptuel de cette recherche, à travers un modèle « général » expliquant le processus adopté par l'individu lorsqu'il se trouve confronté à une sollicitation de ses données personnelles.

Après avoir justifié certains choix faits au préalable, notamment concernant le contexte d'analyse (nous excluons toute sollicitation de données personnelles relative à un achat en ligne), la phase d'analyse (décision de fournir ou non les données demandées, une fois arrivé sur la page contenant le formulaire à remplir) et le type de processus considéré (essentiellement heuristique), nous avons exposé l'ancrage théorique de cette recherche.

Trois théories majeures ont ainsi été exposées et ont permis de justifier la formulation de ce modèle conceptuel général. Nous avons alors démontré en quoi la théorie de l'échange social, la théorie de la motivation (ou de l'espérance attendue) et la théorie de l'action raisonnée étaient pleinement utiles à notre cadre d'analyse.

Le modèle conceptuel général auquel nous avons abouti est un modèle processuel qui identifie l'impact des facteurs situationnels et de la préoccupation pour le respect de la vie privée (valeur individuelle) sur les réponses attitudinales et comportementales des individus, face à une sollicitation de leurs données personnelles. Il répond en cela à la remarque de Webb (1981), présentée en introduction de ce chapitre, qui rappelle le caractère éminemment subjectif de l'évaluation de la sollicitation et son incidence sur le comportement observé.

La dernière section de ce chapitre, a consisté à formuler les propositions et hypothèses de recherche, qui seront ensuite testées empiriquement dans les chapitres suivants. Ces propositions et hypothèses se scindent en quatre blocs : celles relatives à l'influence de l'exposition à la sollicitation (c'est-à-dire aux facteurs situationnels), les antécédents de l'attitude face à la communication des données et ses conséquences, enfin, les conséquences sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données.

Le modèle général présenté dans ce chapitre (cf. figure 7.6) n'est pas celui qui sera réellement testé dans le cadre de ce travail doctoral. Ce dernier sera présenté dans le chapitre qui suit (chapitre 8), où nous justifierons, notamment, le choix des facteurs situationnels étudiés.

Partie II – EXPLORER ET BATIR

Sollicitation de données personnelles
et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

CHAPITRE 8
MÉTHODOLOGIE DE LA RECHERCHE

« Test of a theory should not be so complicated or expensive
as to make refutation impossible ».
Zaltman, Pinson and Angelmar (1973)

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 8

Pour rappel, l'objectif de ce travail doctoral est d'évaluer les réactions attitudinales et comportementales des consommateurs, face à une sollicitation de leurs données personnelles, plus précisément, quand celle-ci s'opère par le biais d'un formulaire à remplir sur Internet.
Après avoir présenté, dans le chapitre précédent, le cadre conceptuel de cette recherche ainsi qu'un modèle conceptuel général (cf. figure 7.6), nous abordons, dans le présent chapitre, la méthodologie mise en œuvre pour tester ce modèle.

Puisque nous cherchons essentiellement à mesurer des relations de cause à effet entre nos variables, nous avons opté pour un design expérimental. La justification de ce choix (1.1) et des facteurs manipulés (1.2) est présentée dans la première section de ce chapitre. Quatre facteurs principaux ont ainsi été choisis, correspondant chacun à une des catégories de facteurs situationnels recensées au chapitre 4 et reprises dans le modèle conceptuel général.

Dans un second temps (2), nous soumettons les principales réflexions auxquelles nous avons été confrontés avant de constituer et de mettre en œuvre le plan expérimental. Nous évoquons ainsi, après avoir rappelé les objectifs de l'expérimentation (2.1), les principaux types de biais qu'il sera nécessaire de contrôler (2.2). Nous terminons par le choix des individus interrogés (2.3). Au final, le plan expérimental choisi comprend trois expérimentations successives.

Nous exposons, dans une troisième partie (3), le modèle qui sera finalement testé dans le cadre de ce travail doctoral (3.1), avant d'indiquer quels facteurs seront manipulés dans chacune des expériences prévues (3.2).

La quatrième partie concerne le contrôle des variables externes (4). Nous justifions la nécessité de ce contrôle (4.1), avant d'exposer les variables qui seront effectivement contrôlées (4.2). Nous terminons en notifiant le type de contrôle qui sera mis en œuvre (4.3).

Dans la cinquième et dernière partie (5), nous abordons les questions de validité. Nous y exposons les menaces à la validité interne (5.1) puis à la validité externe (5.2) qui devront être prises en compte dans le cadre de cette recherche.

1 Le choix d'un design expérimental

Nous justifions, dans un premier temps, le choix d'un design expérimental (1.1), avant d'exposer les facteurs situationnels qui seront effectivement manipulés (1.2).

1.1 La justification du choix d'un design expérimental

Pour tester le modèle conceptuel proposé au chapitre précédent (chapitre 7[130]) et identifier, notamment, l'impact de l'exposition à la sollicitation sur le processus de réponse, nous avons choisi d'adopter une logique expérimentale, et ce, pour deux raisons principales : d'une part, pour nous démarquer des recherches existantes et, d'autre part, parce qu'elle nous a semblé être la méthode la plus adaptée, compte tenu de notre cadre de recherche.

La majorité des recherches menées jusqu'ici, dans le domaine de la sollicitation de données personnelles, adoptent une logique de sondage. Or, il n'est pas certain qu'il s'agisse de la méthode la plus appropriée pour mesurer un tel phénomène. Le problème majeur des enquêtes d'opinion est que le sujet même de l'étude peut augmenter la préoccupation des individus, telle qu'ils l'expriment à travers leurs réponses. De plus, l'opinion des individus les plus préoccupés n'est pas forcément représentée car ils refusent souvent de participer à ce genre d'enquêtes. Ensuite, les résultats de ces sondages ne sont valables que dans la mesure où les répondants connaissent les pratiques à propos desquelles ils sont interrogés, ce qui n'est pas toujours le cas. Ainsi, comme le signale Schultz (1990), certains sondages n'examinent pas les connaissances des individus interrogés alors que la compréhension de leurs préoccupations rend nécessaire de connaître leurs croyances dans ce domaine. Enfin, comme le soulignent Cranor et al. (1999), une approche par sondage est souvent peu adaptée car l'attitude déclarée est rarement conforme aux décisions prises dans la réalité. La variance en matière de dévoilement de soi pourrait ainsi mieux se révéler sur la base de questions liées à des scénarios spécifiques, ce qui est le cas dans une logique expérimentale.

L'expérimentation est, à côté de l'observation et de l'enquête, un des outils majeurs de collecte de données primaires en marketing. C'est une forme d'investigation qui cherche à vérifier des relations de cause à effet entre des variables indépendantes et dépendantes en

[130] Pour rappel, nous avons déjà souligné, à la fin du chapitre 7, que le modèle proposé à la figure 7.6 était un modèle général et ne serait pas celui qui serait réellement testé. Celui-ci sera exposé dans le présent chapitre.

manipulant et contrôlant les premières et en mesurant les variations concomitantes des dernières (Calciu 2003) ; elle facilite donc les études causales qui sont la forme la plus évoluée de la recherche en marketing (Lambin 1990, p307). La première vertu de l'approche expérimentale tient en effet à sa capacité à démontrer de manière solide la relation causale entre des variables (Bagozzi 1977). Comme l'indiquent Emory et Cooper (1991) cités par Liu et al. (2005), bien que la causalité ne puisse jamais être établie avec certitude, la méthode expérimentale est celle qui se rapproche le plus de cet objectif, comparativement à d'autres méthodes. De plus, elle permet d'avoir un haut niveau de validité interne quand on teste le modèle, du fait du contrôle exercé sur certaines variables. En effet, l'expérience permet de reproduire des situations en milieu fermé, de contrôler et manipuler certains paramètres et de mesurer leur incidence sur les variables étudiées (Herbert, 2005 a).

Dans notre cas, puisque nous cherchons à mesurer des relations de cause à effet entre l'exposition à la sollicitation et les variables du processus de réponse, une logique expérimentale semble tout à fait adéquate. De plus, du fait de la multitude de facteurs explicatifs, seule une expérience sous contrôle peut nous permettre de déterminer quels facteurs jouent un rôle dominant. Pour l'ensemble de ces raisons, nous aurons donc recours à une logique expérimentale pour tester notre modèle.

1.2 Le choix des facteurs situationnels étudiés

Sur l'ensemble des facteurs situationnels identifiés comme ayant une influence sur le processus de réponse face à une sollicitation de données personnelles et présents dans le modèle conceptuel général (cf. figure 7.6 du chapitre 7), il semblait intéressant, théoriquement et managérialement parlant, d'étudier plus particulièrement certains d'entre eux.

Tout d'abord, seuls certains facteurs situationnels ont déjà fait l'objet de recherches par le passé, et ce, exclusivement dans le cadre américain. De plus, la majorité des travaux adoptent une logique de sondage. Les résultats étant basés sur du « déclaratif », il est alors difficile d'évaluer, de façon certaine, l'effet réel des facteurs étudiés. Analyser l'impact de nouveaux facteurs et/ou de facteurs déjà étudiés, mais dans un autre contexte, serait donc bienvenu.
Tester l'impact de facteurs situationnels présente un intérêt managérial certain. Dans ce domaine, comme dans beaucoup d'autres, les entreprises sont en effet à la recherche de techniques permettant d'accroître la probabilité de réponse des clients. Etudier l'influence de

variables sur lesquelles elles sont susceptibles d'agir doit permettre de répondre à cette attente. Il s'agira donc de choisir des facteurs susceptibles d'être « utilisés » par les managers.

Comme quatre types de facteurs situationnels ont été identifiés dans la littérature et repris dans notre cadre conceptuel général (politique de RVP, liens avec l'entreprise sollicitante, dispositif de collecte et circonstances de collecte), nous avons choisi d'étudier plus particulièrement un élément de chaque et de mesurer son effet sur le processus de réponse.
Ces quatre facteurs ont été choisis pour des raisons à la fois théoriques, managériales et opérationnelles (cf. tableau 8.1). Il s'agit en effet de variables dont l'influence a été très peu (voire jamais) étudiée dans la littérature (en tout cas, sous la forme dont nous les opérationnalisons ici), dont la mesure de l'impact présente un intérêt managérial certain et qui concernent des éléments sur lesquels une entreprise peut facilement agir (aspect opérationnel).

Concernant l'aspect théorique, nous soulignons l'appel de certains chercheurs pour étudier davantage l'impact des facteurs liés au demandeur (entreprise sollicitant les données) ou aux circonstances de collecte, les deux autres types de facteurs étant beaucoup plus étudiés. Ainsi, Ward, Bridges et Chitty (2005) étudient l'impact du type de données personnelles demandées et de bénéfices offerts sur les préoccupations et l'acceptation de fournir les renseignements sollicités. Ils suggèrent que leur modèle soit amélioré en y ajoutant d'autres facteurs manipulables tels que, par exemple, la familiarité avec l'entreprise collectant les données. Il est vrai que l'expérience précédente avec l'entreprise sollicitante est considérée comme un facteur discriminant dans l'acceptation de stocker des données dans un fichier (Culnan et Armstrong 1999). L'impact de ce facteur sur l'acceptation de fournir des informations personnelles n'a cependant, à notre connaissance, jamais été validé de manière empirique. De même, Ward et al. (2005) soulignent que les entreprises peuvent collecter des données clients à de multiples occasions (achat, accès au site, participation à un jeu-concours, …) et que, outre le type de bénéfices proposés, le contexte dans lequel les informations sont demandées pourrait également affecter la réponse du consommateur face à la sollicitation. Ce type de variable a été très peu (voire jamais) étudié jusqu'ici, invitant ainsi à combler ce manque. De plus, notre étude exploratoire descriptive (présentée au chapitre 6) révèle l'effet potentiel de ces deux facteurs (familiarité et contexte) sur l'acceptation de fournir des données personnelles en ligne. Nous avions notamment montré que le refus de fournir les données demandées était plus marqué quand les requêtes provenaient d'entreprises peu connues ou qui n'avaient pas de lien antérieur avec l'internaute (entreprises réputées mais non familières). A l'opposé, le fait d'être sollicité par une firme avec laquelle on a eu des contacts préalables

réduit ce comportement de repli. Il se pourrait donc qu'en matière de respect de la vie privée, la familiarité de l'internaute avec l'entreprise et/ou le site soit un facteur plus incitatif que la seule réputation, ce que nous souhaitons vérifier dans la suite de ce travail doctoral. En ce qui concerne le contexte, les résultats de notre étude montrent que les refus sont plus systématiques s'agissant de fournir des données pour discuter sur un forum, tandis qu'ils sont moins marqués quand il s'agit d'obtenir un document et/ou une information. Le contexte de sollicitation serait donc discriminant des comportements de transmission de données personnelles en ligne, ce que nous souhaitons aussi vérifier. Ces deux variables (familiarité et contexte) présentent, en outre, un intérêt managérial certain. Dans une optique de management de la relation client, une entreprise doit en effet savoir quelles informations demander à un client/prospect selon le niveau (moment) de la relation et selon le contexte dans lequel elle le sollicite. L'ensemble de ces éléments nous ont donc conduit à choisir d'étudier l'impact de ces deux facteurs sur les variables du processus de réponse.

Les facteurs liés à la politique de RVP et au dispositif de collecte ont, pour leur part, été beaucoup plus étudiés dans les recherches précédentes. Ainsi, en ce qui concerne le premier, de nombreuses études montrent l'impact des mentions de RVP sur l'acceptation de fournir des données personnelles (cf. chapitre 4). Cependant, l'ensemble de ces travaux ont été menés outre-Atlantique, où la mise en place de politiques de RVP dépend davantage du bon vouloir des entreprises (principe d'autorégulation) que de la loi. Il semblait donc intéressant d'étudier l'impact de ce facteur dans le contexte français, régi par la Loi Informatique et Libertés et par la Directive Européenne de 1995. De plus, même si les entreprises françaises sont censées se conformer aux principes réglementaires, elles disposent cependant de marges de manœuvre, notamment quant à la manière d'informer les clients de leurs pratiques en matière de traitement de données personnelles. Connaître la réaction des consommateurs face à ces dispositions paraissait donc tout à fait intéressant, notamment pour évaluer l'écart éventuel entre la vision réglementaire du phénomène et la perception qu'en ont les individus.

Le dernier facteur étudié, lié au dispositif de collecte, concerne la quantité de données demandées, autrement dit le nombre de champs figurant sur le formulaire à compléter. Si l'impact de ce facteur a été largement étudié dans le contexte des enquêtes par questionnaire, il l'est beaucoup moins en ce qui concerne la sollicitation de données personnelles. La plupart des chercheurs ont en effet focalisé leur attention sur le type de données demandées, sans s'intéresser à leur nombre. Or, même si les études sur l'impact de la longueur du questionnaire, dans le cadre de la participation à une enquête, ont du mal à établir de manière certaine un tel effet (certaines trouvent un effet positif, d'autres un effet négatif et d'autres

encore aucun effet), il semble probable qu'il existe un seuil de tolérance au-delà duquel cet effet serait démontré. De la même manière, il paraît intéressant d'étudier l'effet de la longueur du formulaire, dans le cadre d'une sollicitation de données personnelles.

Les facteurs dont nous testerons l'impact dans le cadre de cette thèse sont donc les suivants :

- les mentions de RVP figurant sur le site (plus précisément sur la page contenant le formulaire de sollicitation)
- la familiarité avec l'entreprise sollicitant les données
- la quantité de données demandées (correspondant au nombre de champs à compléter, figurant sur le formulaire)
- le contexte de sollicitation et les avantages liés à l'échange

Tableau 8.1 Justification du choix des facteurs à manipuler

Types de facteurs situationnels	Facteur choisi	Justification		
		Intérêt théorique	Intérêt managérial	Intérêt opérationnel
Liés à la politique de RVP (Respect de la Vie Privée)	Mentions de RVP sur la page du formulaire	Leur influence a été démontrée dans un contexte américain. Qu'en est-il dans le contexte français (très réglementé) ?	Connaître le type de mentions susceptibles de réduire les craintes des internautes	Consiste à réfléchir aux mentions RVP à faire figurer sur le site
Liés au demandeur (entreprise sollicitant les données)	Familiarité avec l'entreprise	L'influence de la familiarité n'a jamais été démontrée de manière empirique	Dans une optique de management de la relation client, savoir quelles données demander selon le niveau de la relation	Consiste à réfléchir aux types d'informations à demander selon l'ancienneté de la relation
Liés au dispositif de collecte	Quantité de données demandées	Son influence est très étudiée dans le cadre des enquêtes mais très peu dans celui d'une sollicitation de données personnelles	Mesure d'un éventuel seuil de tolérance	Consiste à réfléchir à la longueur optimale du formulaire
Liés aux « circonstances de collecte »	Contexte de sollicitation et avantages liés à l'échange	Son influence a été peu étudiée, alors qu'il s'agirait d'un facteur prépondérant (un individu peut accepter de fournir des données pour un jeu-concours mais refuser dans un autre contexte)	Dans une optique de management de la relation client, savoir quelles informations demander dans tel ou tel contexte	Consiste à réfléchir aux occasions permettant de solliciter des données ainsi qu'aux avantages à proposer en échange

Ces quatre facteurs seront détaillés dans la suite de ce chapitre et dans les chapitres suivants. Le modèle conceptuel que nous testerons finalement dans la thèse (cf. figure 8.2) revient donc à analyser l'impact de ces 4 variables situationnelles sur l'évaluation de la sollicitation par le consommateur (correspondant aux différentes perceptions), sur le reste du processus de réponse (notamment sur l'attitude face à la communication des données et la réponse à la sollicitation) et, enfin, sur l'intention de poursuivre la relation avec l'entreprise sollicitante.

Après avoir justifié le choix de la logique expérimentale et des facteurs manipulés, nous exposons, à présent, les conclusions issues des réflexions préalables à la constitution du plan expérimental. Nous présentons, tout d'abord, les objectifs de ce plan expérimental (2.1), avant d'indiquer les principaux types de biais que nous devrons contrôler (2.2). Nous terminerons cette section en évoquant le choix des échantillons interrogés (2.3).

2 Les réflexions préalables à la constitution du plan expérimental

2.1 Les objectifs du plan expérimental

Afin de mesurer l'effet de nos 4 facteurs situationnels sur le processus de réponse, un plan expérimental a donc été mis en place. Parce que nous souhaitions également mesurer leurs éventuels effets d'interaction, nous avons opté pour un plan factoriel, qui offre en outre l'avantage de nécessiter moins d'unités expérimentales qu'il n'en faudrait si l'effet de chacun des facteurs était étudié au cours d'expériences séparées (Evrard et al. 2003, p223). Toutefois, comme il est difficile d'interpréter les résultats issus de la manipulation simultanée de plus de 3 facteurs (surtout concernant les effets d'interaction), nous avons fait le choix d'un plan expérimental en plusieurs phases. De plus, nous voulions aussi mesurer un maximum d'effets d'interaction possibles entre les différents facteurs. Ce choix nous a conduits à retenir un plan expérimental comprenant un ensemble de 3 expériences successives menées auprès d'échantillons variés (étudiants, panélistes et clients réels d'une entreprise) et dans des contextes d'interrogation différents (questionnaire papier-crayon puis en ligne). L'objectif était d'assurer une plus grande validité des résultats obtenus, par la diversité des publics et des situations testés. Ainsi, la troisième expérience, a été menée en partenariat avec une entreprise et s'inscrivait dans une situation réelle, sans que les participants aient conscience de faire l'objet d'une expérimentation. Elle se rapprochait ainsi d'une logique de test visant à mesurer l'impact de différents facteurs sur une variable dépendante unique (ici, la réponse). Nous nous

inscrivions alors dans une logique relativiste (vs. universelle) qui tend à préférer étudier le comportement du consommateur dans la situation dans laquelle il a lieu (Belk 1975).

Notre cadre conceptuel étant relativement complexe (comme nous le verrons par la suite, certaines propositions correspondent à un grand nombre d'hypothèses), ce plan d'expérience offrait en outre la possibilité de le tester en plusieurs fois, en le découpant en 3 parties (cf. figure 8.1) Ainsi, la partie gauche du modèle (jusqu'à l'attitude) sera testée dans l'expérience 1 (EXP1), tandis que l'expérience 2 (EXP2) analysera l'effet des facteurs situationnels depuis l'évaluation de la sollicitation jusqu'aux intentions comportementales. Enfin, l'expérience 3 (EXP3) se focalise surtout sur la partie droite du modèle, des intentions comportementales à la réponse. Chaque expérience vise alors à tester une partie du modèle, indépendamment des autres, tout en permettant des recoupements. Ainsi, la variable dépendante d'un modèle constitue la variable médiatrice du modèle suivant, ce qui permet de consolider les résultats trouvés précédemment. Il s'agit là, en outre, d'un moyen de répliquer l'expérience précédente dans d'autres conditions. Or, l'importance de la réplication pour assurer la validité de l'étude est défendue par de nombreux chercheurs, dont Calder, Philips et Tybout (1981).

Figure 8.1 Modèle conceptuel simplifié (*)

(*) Nous omettons dans ce modèle la variable Préoccupation pour le Respect de la Vie privée

Après avoir exposé les objectifs du plan expérimental, évoquons à présent les principaux types de biais qu'il nous faudra contrôler.

2.2 Les principaux biais à contrôler

Le type d'individus interrogés, la méthode d'échantillonnage retenue et la maîtrise des différents biais sous-jacents constituent, dans les recherches, un axe de réflexion majeur, plus encore pour celles qui adoptent une logique expérimentale d'une part et qui s'intéressent à la Vie Privée d'autre part. Plusieurs types de biais ont été répertoriés dans les recherches en général et dans celles portant sur le RVP en particulier, qu'il convient de limiter.

3 principaux types de biais ont ainsi été identifiés comme potentiellement dommageables ici :
- le biais de non-réponse : ne mesurer que la réponse des individus qui acceptent de participer
- biais de couverture : interroger un échantillon non représentatif de la population de base
- le biais du répondant : lié à sa manière de se comporter face à l'expérimentateur

Avant de détailler ces trois types de biais, plusieurs remarques doivent être formulées à ce propos. Tout d'abord, nous ne prétendons pas avoir identifié, à travers ces trois types de biais, l'ensemble des biais susceptibles d'avoir une incidence sur la validité de nos résultats. D'autres types de biais peuvent donc exister. Les biais de désirabilité sociale et d'auto-sélection font, notamment, partie de ceux-là. Nous considérons cependant qu'ils peuvent se rapprocher de ceux évoqués précédemment. Ainsi, le biais de désirabilité sociale nous semble proche du biais du répondant ; nous les aborderons donc conjointement. De même, le biais d'auto-sélection (fait que seuls les individus intéressés par le sujet et ayant suffisamment de temps acceptent de participer) est très proche du biais de non réponse.
Nous ne parlons pas ici des biais inhérents aux facteurs externes (variables susceptibles de masquer l'effet des traitements sur la variable dépendante étudiée) non contrôlés. Nous reviendrons ultérieurement sur le contrôle exercé sur ces facteurs externes (cf. section 4). Enfin, d'autres biais spécifiques au design expérimental et conditionnant sa validité existent. Ils seront abordés dans la section 5, relative aux questions de validité.

Voyons à présent les caractéristiques de ces trois principaux types de biais avant d'évoquer les solutions adoptées pour les limiter.

2.2.1 Les biais d'auto-sélection et de non réponse

Nous revenons sur le biais d'auto-sélection avant de détailler celui dû aux non-réponses.

Le problème d'auto sélection, évoqué plus haut, implique que seuls les individus intéressés par le sujet et ayant suffisamment de temps pour répondre ont participé. Il est surtout présent lors de questionnaires administrés par courrier et/ou par Internet car les répondants ont alors la possibilité de s'y soustraire. Son effet est donc dommageable si les individus ayant refusé de répondre ont une opinion différente de ceux qui ont accepté. Il convient alors de s'assurer qu'il n'est pas lié au sujet même de l'étude car, dans ce cas, l'effet négatif est démultiplié.

Le biais d'auto-sélection s'avère proche du biais de non réponse, dont il constitue une forme particulière. Ce dernier concerne en effet, soit l'impossibilité de contacter un individu sélectionné pour appartenir à l'échantillon, soit le refus de la part d'un individu contacté de participer à l'étude[131] (Evrard et al. 2003). Ce biais de non réponse doit être évalué puis maitrisé car il a une incidence majeure sur les résultats. L'évaluer consiste à comptabiliser le nombre d'individus n'ayant pas répondu (taux de non réponse) puis à identifier l'étendue avec laquelle les répondants diffèrent des non répondants sur des variables importantes (Groves et Couper 1998). Si l'existence d'un tel écart est dommageable quel que soit le sujet de la recherche, il l'est encore plus dans le cadre des études sur le RVP. On a en effet déjà souligné le fait que l'opinion des individus les plus préoccupés était généralement sous-représentée car ils refusent souvent de participer à des sondages. Dans notre cas, outre les individus qui, d'une manière générale, sont peu enclins à participer à des enquêtes, nous risquons également d'exclure ceux peu enclins à se prononcer sur un sujet (le RVP) qu'ils estiment sensible. Or, il y a fort à parier que ces individus refusant de participer à nos expériences seront très différents de ceux qui acceptent de le faire, sur les variables que nous souhaitons mesurer (la préoccupation pour le RVP, l'attitude face à la communication des données, …). Si nous n'y prenons garde, nous aurons alors créé un biais substantiel, nous empêchant de généraliser nos résultats à l'ensemble de la population étudiée (celle des internautes français, dans notre cas).

Plusieurs méthodes sont préconisées par les chercheurs pour tenter de réduire et/ou limiter ce biais. Certaines reposent sur la possibilité de comparer la répartition de l'échantillon à celle de la population de base et consistent à procéder à un redressement sur la base des sous - ou sur - représentations de certaines classes de la population. La stratification *a posteriori* serait ainsi la méthode la plus utilisée pour le redressement des non-réponses (Droesbeke et al. 1987, cité dans Thiétart et al. 2003). Une autre solution consiste à interroger un sous-échantillon de non répondants, désigné par tirage aléatoire. Le chercheur doit alors s'efforcer d'obtenir une

[131] Nous excluons, dans cette définition, les non réponses partielles à certaines questions du questionnaire, qui nous semblent refléter un autre phénomène.

réponse de l'ensemble des éléments sélectionnés (Droesbeke et al. 1987, cité dans Thietart et al. 2003), ce qui est très difficile. On peut aussi, si on ne dispose pas d'informations sur l'ensemble de la population, remplacer les individus défaillants, soit en pondérant le poids de répondants aux caractéristiques similaires, soit en incluant des répondants supplémentaires aussi semblables que possible aux non répondants. D'autres méthodes ont été proposées[132], notamment par Daniel (1975), pour estimer les caractéristiques des non répondants et ainsi évaluer et limiter le biais de non réponse. Une d'entre elles consiste ainsi à considérer les derniers répondants comme des « clones » des non-répondants. Bien que régulièrement utilisée, elle n'a toutefois jamais réellement démontré son efficacité (Smith, 1983). Ce dernier souligne d'ailleurs que toutes les méthodes utilisées jusqu'ici sont d'une efficacité limitée et qu'il n'existerait pas de moyen sûr de mesurer et de limiter ce biais.

2.2.2 Le biais de couverture

Un échantillon présente un biais de couverture lorsque la population interrogée ne correspond pas à la population de référence, cette dernière étant l'univers de généralisation des résultats de l'étude (Thiétart et al. 2003). Ainsi, dans le cadre de cette recherche, la population de référence correspond à l'ensemble des internautes français. Pour pouvoir généraliser nos résultats à cette population, il faut donc interroger un échantillon qui soit le plus représentatif possible de celle-ci, sachant qu'il n'existe pas (encore) de base de sondage disponible sur le marché (liste indiquant l'ensemble des individus appartenant à cette population). Plusieurs chercheurs, dans le domaine du e-commerce et du e-marketing, soulignent d'ailleurs ce problème dont l'incidence majeure est d'empêcher de recourir à une méthode d'échantillonnage probabiliste.

Un des (seuls) moyens de réduire ce biais consiste alors à interroger un panel d'internautes, dans la mesure où les sociétés qui les développent cherchent à bâtir des « échantillons » qui soient les plus représentatifs possibles de la population de laquelle ils sont tirés. Une autre solution consiste à réduire la population de base (les internautes français, dans notre cas) à un ensemble plus réduit, par exemple les étudiants internautes ou les internautes expérimentés et à réduire l'univers de généralisation à cet ensemble.

[132] Nous ne détaillerons pas ici l'ensemble de ces méthodes et renvoyons les lecteurs intéressés aux principales recherches dans ce domaine.

2.2.3 Le biais de désirabilité sociale et le biais du répondant

La qualité d'une expérimentation repose, avant tout, sur la mise en condition des participants (comportement, volonté, conditions d'environnement, …). En aucun cas, ceux-ci ne doivent adopter un comportement induit par la situation d'expérimentation (de façon plus large par la situation d'interrogation). Cependant, malgré toutes les précautions qu'il est possible de prendre, il reste que le sujet humain est une unité expérimentale complexe, une part de cette complexité provenant du fait qu'il sait pertinemment qu'il est le sujet de recherche (Rosenthal et Rosnow 1969). Le problème du répondant qui donne des réponses socialement désirables (biais de désirabilité sociale) ou qui cherche à agir comme un bon répondant (biais du répondant) doit donc être considéré avec précaution (Fernandes et Randall 1992).

La question de la désirabilité sociale peut se définir comme « la tendance d'un individu à admettre et à adopter les traits et les comportements qui sont socialement désirables ou acceptables, et à rejeter, à l'inverse, ceux qui ne le sont pas » (Zerbe et Paulhus 1987). Dans le cas de la réponse à un questionnaire, ce problème se manifeste « lorsqu'un thème ou un évènement est perçu comme étant sensible par nature, embarrassant, angoissant ou divergent par rapport à l'image que l'individu a de lui-même ». Dans ce cas, « [..] quand on demande à cet individu de reporter ses opinions et ses attitudes par rapport à ces thèmes, il est probable soit qu'il s'abstienne, soit qu'il distorde et falsifie son discours dans un sens socialement désirable » (Cannell, Oksenberg et Converse 1977). Cette situation peut être soit consciente (tendance de l'individu à falsifier de manière volontaire la description qu'il donne de lui-même afin de se valoriser) soit inconsciente (tendance à se décrire positivement de façon involontaire). A l'instar de nombreux chercheurs (dont Gur et Sackeim 1979, Paulhus et Reid 1991), nous nous intéressons ici à la dimension inconsciente du phénomène, puisqu'elle est la principale concernée dans le cadre de la réponse à un questionnaire. Si certains auteurs considèrent la désirabilité sociale comme une source de contamination (elle ne reflète pas directement le comportement de l'interviewé mais est induite par la formulation des questions), la majorité la voit comme un trait de personnalité correspondant à la recherche de l'approbation des autres. Dès lors, les différences, en termes de désirabilité sociale, trouvées dans les réponses reflèteraient, au moins en partie, des différences de personnalité entre les individus. Or, ce type de différences individuelles est en partie capté par les variables attitudinales et comportementales (Ajzen et Fishbein 1980). Dans notre modèle, elles devraient donc se retrouver dans les différences d'attitude et de comportement face à la communication des données personnelles demandées. De plus, une manière de limiter ce biais

est de recourir à la randomisation des répondants (Dalton et Metzger 1992), technique que nous emploierons dans nos trois expériences. Bien que potentiellement dommageable dans le cadre de notre étude, du fait du caractère éminemment sensible du sujet (le RVP), ce biais de désirabilité sociale devrait donc être, au moins limité, si ce n'est supprimé[133].

Comme nous l'avons déjà signalé, le problème du biais du répondant provient du fait que, quelle que soit la situation de questionnement, le questionné est conscient d'être objet de recherche. Parce qu'il est un être humain (donc doué de pensée), il cherche naturellement à comprendre pourquoi il est interrogé. Or, la situation de questionnement se caractérise par une asymétrie d'information où le questionné est dans la situation inconfortable de celui qui en sait le moins. Deviner la finalité de l'exercice et ajuster son rôle en fonction de celle-ci lui permet alors d'avoir prise sur une situation qui lui échappe en partie. Il faut aussi prendre en compte les capacités d'interprétation des sujets étudiés (Herbert 2005 b). Ainsi, le contrôle et la manipulation de certains paramètres peuvent ne pas être homogènes car ils sont l'objet de représentations différentes chez les différents individus interrogés[134].

Dans le cadre de la méthode expérimentale, la totalité des signes susceptibles de communiquer les hypothèses au sujet expérimental sont donc des déterminants significatifs de son comportement de réponse. Or, les conséquences de l'adoption d'un rôle expérimental sont considérées comme biaisant systématiquement les résultats (Orne 1969 ; Sawyer 1975). Cette considération est cependant liée aux recherches initiales de Orne (1962) qui estimait qu'un sujet expérimental ne pouvait adopter qu'un seul et unique rôle, celui du « sujet complaisant ». Depuis, l'ensemble des recherches concernant les rôles du sujet expérimental (Sigall et al. 1970, Weber et Cook 1972, Berkowitz et Donnerstein 1982), a montré que ce dernier pouvait adopter des rôles différents, et pas seulement complaisant, susceptible de biaiser les résultats. Toutefois, l'idée de masquer la véritable finalité de l'expérience pour réduire ce risque n'est pas forcément une solution efficace et peut même créer plus de problèmes qu'elle n'en évite (Papageorgis 1968).

[133] Même si nous pensons que le biais de désirabilité sociale devrait être limité, il aurait été intéressant de pouvoir l'évaluer. Or, pour une question de limitation de la longueur du questionnaire, nous n'avons pas pu inclure d'échelle nous permettant de le mesurer et donc de mieux contrôler, le cas échéant, son effet. Si ce choix ne nous semble pas de nature à compromettre nos résultats, l'insertion d'une telle mesure dans le cadre d'une nouvelle étude serait tout à fait souhaitable. Nous reviendrons sur ce sujet dans la discussion finale.

[134] Pour mieux connaître l'ensemble des caractéristiques et des effets du biais du répondant, nous renvoyons les lecteurs intéressés à la thèse de Maud Herbert (2005).

Parmi les solutions visant à limiter le biais du répondant, les plus répandues consistent donc à procéder à une affectation aléatoire des sujets aux différents traitements, à soigner la mise en condition des répondants pour ne pas les encourager à adopter un comportement spécifique et à éviter tout signal susceptible de leur faire deviner les hypothèses de recherche, même si, dans ce dernier cas, tous les auteurs ne sont pas d'accord sur l'efficacité d'une telle pratique.

En résumé, nous donnons, dans le tableau suivant, une définition de chacun de ces biais, tout en suggérant les principales solutions envisagées pour les réduire. Nous indiquons en italique, les solutions employées dans le cadre de notre dispositif expérimental.

Tableau 8.2 Principaux biais à maîtriser et moyens de limiter ces biais

Type de biais	Définition et caractéristiques	Moyens de limiter ce biais
Biais de non réponse	Biais lié soit à l'impossibilité de contacter un individu sélectionné soit au refus de la part d'un individu contacté de participer à l'étude	- *Réduire les problèmes d'accessibilité aux répondants (EXP1)* - *Encourager les personnes à répondre (relances, présence d'incentives) (EXP2)* - Limiter l'auto-sélection liée au sujet de l'étude (qui aboutit à ne pas interroger les individus très préoccupés par le RVP) - *Interroger des non-répondants pour comparer leur attitude avec celle des répondants sur certaines variables clés du modèle (EXP3)* - *Redresser l'échantillon (EXP3)*
Biais de couverture	Fait de ne pas interroger certaines « parties » de la population de base (manque de représentativité de l'échantillon retenu)	- Disposer d'une liste (base de sondage) actualisée et recouvrant le plus exactement possible la population étudiée - *Eviter de recourir à des échantillons de convenance qui, par définition, ne sont pas représentatifs de la population de base (EXP2 et EXP3)*
Biais du répondant	Provient du fait que le sujet expérimental est conscient d'être l'objet de recherche. Cela peut le conduire à adopter un rôle visant à maîtriser une situation qui lui échappe en partie. Ceci se traduit par une manière de se comporter vis-à-vis de l'expérimentateur qui peut biaiser les résultats obtenus	- *Adopter une logique de randomisation (affectation aléatoire des unités expérimentales aux différents traitements) (EXP 1, EXP2 et EXP3)* - *Eviter tout signal susceptible de communiquer les hypothèses cherchant à être validées (EXP2)* - *Soigner la mise en condition des répondants pour ne pas les encourager à adopter un comportement spécifique (EXP1)* - *Faire en sorte que le répondant n'ait pas conscience de participer à une expérimentation (EXP3)*

Voyons à présent le choix des échantillons interrogés.

2.3 Le choix des échantillons interrogés

Il est difficile de réduire conjointement tous les biais (présentés dans la partie précédente), connaissant les contraintes qui sont les nôtres, en termes de logistique, timing et finances notamment. Nous avons donc décidé de limiter tour à tour chacun de ces biais, en choisissant de manière opportune les individus à interroger. Pour chacune des expériences menées (trois au total), l'objectif est donc de maîtriser un des biais évoqués, tout en limitant au maximum l'effet possible des 2 autres. Outre la randomisation des individus aux différents traitements, la limitation de ces trois principaux types de biais sera donc le fait d'une sélection appropriée des unités expérimentales interrogées. Nous expliquons ci-dessous le type d'individus interrogés dans chacune des expériences, tout en précisant le type de biais maîtrisé.

Notre première expérience (EXP1) a ainsi été menée auprès d'étudiants d'université (de niveau Master), pendant des TD ou des cours magistraux. Afin d'atteindre la taille voulue (soit 240 étudiants environ), les questionnaires ont été soumis à 5 « classes » différentes (de niveau Master 1ère ou 2ème année). L'administration s'est déroulée sur une semaine et demie et a mobilisé un enquêteur unique. Dans la mesure où il s'agit d'un sujet sensible, une démarche éthique recommande d'obtenir un consentement volontaire et éclairé des personnes interrogées. Sieber (1982) explique que « volontaire » signifie de manière libre (sans contrainte ni promesse), tandis qu'« éclairé » signifie que le sujet sait tout ce qu'une personne raisonnable voudrait savoir dans une telle situation. Dans notre cas, tous les étudiants présents ont été soumis aléatoirement à l'un des questionnaires, chacun étant toutefois informé et libre de ne pas y répondre[135]. La population interrogée était donc en grande partie « captive » puisque, malgré nos précautions, on peut penser que certains étudiants n'ont pas osé se manifester, malgré leur désir éventuel de ne pas participer. Cette volonté était notamment réduite par le fait qu'aucun de leurs camarades n'avait refusé avant eux. Au final, tous les étudiants sollicités ont ainsi répondu. Cette solution limite donc fortement (voire annule) le biais de non-réponse, tout en réduisant parallèlement la validité externe des résultats, de par l'homogénéité et l'absence de représentativité de l'échantillon interrogé (biais de couverture). De plus, malgré toutes les précautions que nous pouvons prendre, le biais du répondant est probablement élevé.
La seconde expérience vise à limiter, au moins en partie, ce problème de validité externe.

[135] Cette possibilité leur était signalée verbalement au début de la séance puis lors de la remise du questionnaire.

Dans cette seconde expérience, le questionnaire est administré en ligne auprès d'internautes. Interroger ce type d'individus pose un problème majeur, déjà évoqué précédemment. Il est en effet très difficile (voire impossible) d'utiliser une méthode probabiliste pour sélectionner les personnes à questionner, ce qui empêche d'avoir un échantillon entièrement représentatif de la population de base. En outre, les méthodes basées sur l'auto-sélection (notamment par le biais d'annonces sur des sites fréquemment visités) ont l'inconvénient de réduire la proportion d'individus très préoccupés par le RVP au sein de l'échantillon. Une alternative consiste à faire appel à une société de panel afin de sélectionner, au sein de leur base de données, un échantillon qui soit le plus représentatif possible de la population des internautes français et de soumettre les individus sélectionnés au questionnaire, sans qu'ils soient pour autant informés du sujet précis de l'étude[136]. Le premier avantage de cette solution est d'interroger un échantillon assez représentatif de la population des internautes français, notamment d'un point de vue sociodémographique. L'autre avantage est de limiter les biais de non réponse et d'auto-sélection (les individus étant panélistes, il y a de grandes chances pour qu'ils acceptent de participer). Elle pose cependant un inconvénient majeur, lié au fait d'interroger des panélistes. On peut en effet supposer qu'il s'agit d'individus qui acceptent plus que la moyenne de donner des informations, la preuve étant qu'ils se sont inscrits pour répondre à des enquêtes en ligne. Concernant le biais du répondant, il est difficile de l'évaluer *a priori*, même si l'on peut penser qu'il sera moins important que dans la première expérience.

Dans le troisième cas, l'échantillon sera sélectionné aléatoirement au sein de la base de données d'une entreprise partenaire. Les participants recevront la demande de participation par email, sans avoir pour autant conscience de participer à une expérience. L'email envoyé, qui propose de participer à une opération particulière correspond en effet en tout point au type de courrier qu'ils reçoivent régulièrement de la part de ce fournisseur. Cette solution, tout en introduisant un biais de non réponse, fait donc entièrement disparaître le biais du répondant. De plus, l'auto-sélection s'opère sur l'attractivité de la proposition et non sur les problèmes de Respect de Vie Privée, ce qui réduit donc son impact dans le cadre de notre sujet d'étude. En outre, l'échantillon sélectionné se veut représentatif de la population de base (la base de données de l'entreprise). Enfin, sans être pour autant représentatif de la population des internautes français, il y a cependant des chances pour qu'il s'en rapproche, au moins sur certains critères sociodémographiques ou expérientiels (expérience d'Internet).

[136] Cette pratique n'est pas éthiquement incorrecte puisque les panelistes ont accepté au préalable de participer à des enquêtes.

Le tableau ci-dessous reprend les avantages et inconvénients liés à chaque type d'échantillon.

Tableau 8.3 Sélection des échantillons à interroger et types de biais à maîtriser

Expériences	Biais de non réponse	Biais de couverture	Biais du répondant
EXPERIENCE 1 - échantillon d'étudiants - questionnaire papier-crayon	Faible voire nul (population quasi « captive »)	Elevé (échantillon de convenance dont les caractéristiques peuvent être éloignées de la population de base)	Elevé (notamment à cause du type de population interrogée)
EXPERIENCE 2 - échantillon de panelistes (panel d'internautes) - questionnaire en ligne	Limité mais non nul (le refus de répondre n'est cependant pas lié au sujet de l'étude)	Faible (le panel est censé être représentatif de la population des internautes français)	Elevé (existe d'autant plus que les répondants ont choisi de participer à des études)
EXPERIENCE 3 - échantillon de clients d'un site Internet partenaire - les clients participent à une vraie opération organisée par le site	Elevé (certains des clients interrogés refuseront de participer à l'opération proposée et seront donc considérés comme des non-répondants)	Moyen (l'échantillon sera représentatif de la base de données de l'entreprise, elle-même représentative des internautes français sur certains critères)	Faible voire nul (les répondants n'ont pas conscience de participer à une expérimentation)

Comme l'indique ce tableau et conformément à l'objectif que nous nous étions fixés, le choix des échantillons, pour chaque phase expérimentale, permet donc de limiter, tour à tour, les principaux biais identifiés précédemment. Soulignons cependant que les éléments testés dans chaque expérience étant différents (cf. figure 8.1) et les biais contournés étant à chaque fois limités (par exemple, seul le biais de non-réponse est réellement réduit dans l'expérience 1), nous ne pourrons pas comparer entièrement les résultats entre eux ni les généraliser (sans prendre de précautions particulières) à l'ensemble de la population de base.

Nous abordons, à présent, le design et les caractéristiques du plan expérimental. Nous présentons, tout d'abord, le modèle que nous serons amenés à tester dans le cadre de ce travail doctoral, modèle qui constitue une « application » du cadre conceptuel général présenté au chapitre précédent (3.1). Nous évoquons ensuite le plan expérimental et surtout les facteurs qui seront manipulés dans les trois expériences que nous seront amenés à réaliser (3.2).

3 Présentation du modèle à tester et du plan expérimental

3.1 Le modèle à tester

Nous indiquons ci-après la démarche ayant conduit au modèle à tester, exposé à la figure 8.2. Comme nous l'avons déjà souligné, l'impact de 4 facteurs situationnels particuliers (présentés à la section 1) sera finalement testé dans le cadre de ce travail doctoral. Comme leur effet sur l'évaluation de la sollicitation correspond à la proposition P1 exposée au chapitre précédent, ils seront identifiés respectivement par les numéros 1.1 (Mentions de RVP), 1.2 (Familiarité), 1.3 (Quantité) et, enfin, 1.4 (Contexte).

De plus, nous rappelons que trois expériences seront menées successivement et que nous manipulerons conjointement deux ou trois des quatre facteurs choisis, ceci afin d'obtenir un design factoriel « raisonnable ». Ces trois expériences seront identifiées par des couleurs différentes au sein du modèle (rouge pour l'expérience 1, bleu pour la 2 et vert pour la 3). Comme nous l'avons déjà précisé, dans la première expérience, nous nous focaliserons sur la partie gauche du modèle (jusqu'à l'attitude face à la communication des données), tandis que dans la seconde, nous irons jusqu'à mesurer les intentions comportementales. Enfin, dans la troisième et dernière expérience, nous irons des intentions comportementales jusqu'à la réponse effective, en cas de sollicitation. Les variables étudiées respectivement dans l'une ou l'autre de ces expériences sont donc identifiées par la couleur correspondante. Ainsi, par exemple, les variables de confidentialité, sensibilité et pertinence perçues ne seront mesurées que dans l'expérience 1, tandis que la réponse à la sollicitation ne sera évaluée que dans la troisième. De même, le facteur « Mentions de RVP » (1.1) ne sera manipulé que dans la première expérimentation, tandis que les trois autres facteurs figureront au moins dans deux des trois expériences[137]. Les variables qui seront mesurées dans plusieurs expériences figurent en noir dans ce modèle. Nous y faisons également apparaître l'ensemble des variables dites externes qui seront contrôlées dans le cadre de ce plan expérimental[138] et qui seront exposées dans la section suivante. Enfin, nous rappelons que l'effet direct des intentions comportementales sur la réponse effective ne sera pas testé dans le cadre de ce travail, pour les raisons indiquées au chapitre précédent (cf. chapitre 7).

Le modèle à tester dans le cadre de cette thèse se présente donc de la manière suivante :

[137] Seul le facteur « Quantité » sera manipulé dans les trois expériences, conjointement à d'autres facteurs

[138] Certaines d'entre elles ne seront pas contrôlées à chaque fois, pour des raisons qui seront expliquées dans les chapitres correspondants (cf. chapitres 10, 11 et 12)

Figure 8.2 Modèle général à tester : Attitude et comportement du consommateur face à une sollicitation de ses données personnelles

Variables de contrôle

- Niveau d'implication vis-à-vis de la catégorie de produits (IMPL)
- Variables sociodémographiques (âge, sexe, CSP, niveau d'éducation)
- Expérience (EXP) et utilisation (UTIL) d'Internet
- Comportement habituel de réponse (COMP_HAB)
- Qualité perçue de la relation (QUALITE)

Etudié uniquement dans l'EXPERIENCE 1
Etudié uniquement dans l'EXPERIENCE 2
Etudié uniquement dans l'EXPERIENCE 3
Etudié dans plusieurs EXPERIENCES
Non testé dans la thèse

Puisqu'il s'agit d'un design expérimental en 3 phases, nous développerons ensuite 3 « sous-modèles » à tester (cf. figures 10.3, 11.2 et 12.3 présentés dans les chapitres correspondants) avec des hypothèses spécifiques à chacun d'eux. Ces hypothèses seront la traduction des propositions de recherche (notamment P1, P3i et P6) développées dans le chapitre précédent.
Nous présentons maintenant le plan expérimental, à travers la manipulation des facteurs dans chacune de ces expériences. Notre objectif n'est pas ici d'indiquer de quelle manière ils seront manipulés[139], mais plutôt de voir quels facteurs seront manipulés conjointement.

3.2 Le plan expérimental

Notre modèle à tester vise à mesurer l'effet de 4 facteurs principaux (Mentions de RVP sur le site et/ou la page du formulaire à remplir, Familiarité avec l'entreprise, Quantité de données demandées et Contexte de sollicitation) sur le processus de réponse du consommateur face à une sollicitation de ses données personnelles.
Nous évoquons maintenant les facteurs qui seront manipulés conjointement dans chacune de ces expériences et les raisons de ces choix. Nous précisons tout d'abord que l'objectif majeur du plan factoriel était de tester un maximum de relations de cause à effet, afin d'accroître la validité des résultats. Nous souhaitions aussi tester le plus grand nombre d'effets d'interaction possibles entre les facteurs, tout en nous pliant aux règles de décision décrites ci-dessous :
- Règle n°1 : conserver un design « raisonnable » en termes de facteurs manipulés conjointement et de nombre d'individus à interroger.
La majorité des chercheurs spécialistes de la logique expérimentale[140] plaident en effet pour la simplicité du design. Il serait ainsi préférable de réaliser plusieurs expériences en 2 x 2 plutôt qu'une seule « grande » expérience en 2 x 2 x 2 x 2. La principale raison à cela est qu'il est difficile d'interpréter les résultats d'une analyse de variance à plus de trois facteurs, notamment en ce qui concerne les effets d'interaction.
Nos expériences se limiteront donc à un maximum de trois facteurs manipulés conjointement, si possible à deux niveaux chacun et, au plus, à trois niveaux (cas du facteur « Contexte » qui comprendra 3 niveaux dans le cadre de l'expérience 2).
- Règle n°2 : respecter le principe d'orthogonalité des facteurs manipulés.
Il s'agit ici de s'assurer que les facteurs manipulés conjointement ne sont pas susceptibles de s'influencer mutuellement. Si cette règle se vérifie *a posteriori*, le chercheur doit cependant y réfléchir au préalable, afin d'être certain qu'il ne risque pas d'aboutir à une telle situation.

[139] Ceci sera expliqué dans les chapitres correspondant respectivement aux trois expériences (i.e. 10, 11 et 12)
[140] Citons notamment Calder Phillips, et Tybout (1982) et Perdue et Summers (1986)

Après avoir précisé nos objectifs et nos règles de décision, nous évoquons la démarche suivie, en commençant par détailler nos réflexions sur chaque facteur pris séparément et en poursuivant par le choix des facteurs finalement manipulés dans chacune des expériences.

Comme nous l'avons déjà indiqué, l'impact du facteur « Mentions de RVP » (qui correspond à la mise en œuvre de la Politique de RVP de l'entreprise) a déjà été étudié à maintes reprises dans la littérature, même s'il s'agissait d'un contexte différent (anglo-saxon vs. français ici). Nous avons donc décidé de ne manipuler ce facteur que dans une des 3 expériences, afin de nous concentrer davantage sur les autres facteurs, beaucoup moins étudiés.
Parmi les trois autres facteurs, celui correspondant à la « Quantité » nous semblait le plus « prometteur ». Nos entretiens qualitatifs indiquaient en effet l'existence probable d'un effet de saturation chez les répondants, ceux-ci acceptant de fournir quelques informations jugées nécessaires mais se montrant plus agacés s'il fallait dévoiler un grand nombre de données. De plus, ce facteur, très majoritairement étudié dans le contexte des enquêtes (sans que son effet soit d'ailleurs clairement démontré) ne l'était quasiment pas dans le cas d'une sollicitation de données personnelles. Nous avons donc choisi de le manipuler dans chacune des expériences. Enfin, les deux autres facteurs seront manipulés dans deux expérimentations sur les trois.

A priori, il ne paraît pas exister de problème d'orthogonalité entre les facteurs manipulés. Un doute est cependant apparu concernant l'interaction entre le facteur « Mentions de RVP » et le facteur « contexte ». Il nous semblait en effet plausible que les mentions relatives à la protection des données et figurant sur la page du formulaire à remplir soient adaptées au contexte de sollicitation. Elles seraient ainsi différentes selon qu'il s'agit d'un formulaire pour participer à un jeu-concours ou d'une inscription à une newsletter. Dans ces deux cas en effet, les informations à fournir aux consommateurs peuvent être différentes car l'objectif sous-jacent l'est aussi. Nous avons donc décidé de ne pas étudier ces deux facteurs conjointement.

Dans l'expérience 1, nous nous intéressons à la partie gauche du modèle à tester. De plus, nous ne mesurons que trois des quatre variables perceptuelles identifiées (confidentialité, sensibilité et pertinence) et évaluons leur impact sur l'attitude face à la communication des données, sans nous intéresser, pour l'instant, ni aux intentions comportementales (étudiées dans l'expérience 2) ni à la réponse (expérience 3). Les trois premiers facteurs (Mentions de RVP, Familiarité, Quantité) y seront donc manipulés conjointement, à deux niveaux chacun, soit un design factoriel inter-sujet en 2 x 2 x 2.

Dans l'expérience 2, nous nous focalisons sur le centre du modèle (de l'évaluation aux intentions comportementales, en passant par l'attitude). De plus, seule la valeur perçue (gain net à répondre), correspondant à l'évaluation de la sollicitation, sera concrètement mesurée[141]. Il nous semblait alors intéressant de manipuler les facteurs « Quantité » et « Contexte », qui nous paraissaient avoir un effet possible sur cette variable, l'effet de la « Familiarité » étant, selon nous, plus accessoire. Puisque nous ne manipulions que deux facteurs, nous avons choisi d'étudier trois contextes différents, soit un design inter-sujet en 2 x 3.

Enfin, dans l'expérience 3, nous nous intéressons surtout à la réponse du consommateur et à son impact sur l'intention de poursuivre la relation. Tous les facteurs nous semblaient alors déterminants, sachant que les « Mentions de RVP » étaient manipulées dans l'expérience 1 et que leur effet est déjà largement étudié dans la littérature. Les trois autres facteurs ont donc été manipulés conjointement, dans le cadre d'un design inter-sujet en 2 x 2 x 2[142].

Nous indiquons, dans la figure suivante, l'enchaînement des manipulations de facteurs :

Figure 8.3 Facteurs manipulés dans chacune des expériences

Le fait de manipuler ces différents facteurs conjointement nous permet, en outre, d'étudier tous les effets d'interaction possibles, hormis celui concernant les « mentions de RVP » x le « Contexte », dont nous avons justifié l'absence précédemment. Le tableau suivant présente l'ensemble des effets testés (directs ou d'interaction) dans le cadre de ce plan expérimental.

[141] Ce choix sera justifié dans les chapitres 10 et 11, correspondant respectivement aux expériences 1 et 2.
[142] En réalité, un quatrième facteur (i.e. « l'ordre des questions ») sera manipulé. Il n'est pas présenté ici dans la mesure où son effet ne sera pas analysé. Nous reviendrons sur les raisons de ce choix dans le chapitre 12.

Tableau 8.4 Ensemble des effets testés dans le cadre du plan expérimental

Effets testés		EXP 1	EXP 2	EXP 3	Nombre d'effets testés
Directs	Mentions de RVP	X			1
	Familiarité	X		X	2
	Quantité	X	X	X	3
	Contexte		X	X	2
D'interaction	RVP x Familiarité	X			1
	RVP x Quantité	X			1
	RVP x Contexte				0
	Familiarité x Quantité	X		X	2
	Familiarité x Contexte			X	1
	Quantité x Contexte		X	X	2
Total par expérience		6	3	6	15

Comme le montre ce tableau, certains effets directs sont ainsi testés deux ou trois fois, même s'il ne s'agit pas toujours des mêmes variables dépendantes. De plus, certains effets d'interaction sont également testés deux fois, ce qui permettra, par la suite, de confronter les résultats et d'assurer, le cas échéant, une meilleure validité interne.

Au final, le plan expérimental auquel nous avons abouti est résumé au tableau 8.5, page suivante. Celui-ci présente les facteurs expérimentaux manipulés dans chaque expérience et donne quelques indications supplémentaires, notamment concernant le type de design et d'échantillon utilisés dans chaque cas. Nous indiquons aussi les facteurs que nous « contrôlerons » dans chacune des phases ainsi que les objectifs de chacune d'elles.

Dans l'ensemble, chaque expérience visera donc à mettre l'individu dans une situation de sollicitation de ses données personnelles, qui soit la plus proche possible de la réalité. Après avoir placé le répondant dans un contexte spécifique, on lui présentera un formulaire, identique à celui qu'il aurait eu à remplir s'il s'était trouvé en ligne dans cette situation. Il faudra donc élaborer le scénario, le tester puis mettre en œuvre le plan expérimental.
La principale difficulté vient de la nécessité de bâtir un scénario crédible dans lequel les participants seront plongés, sans pour autant leur révéler ni le sujet de l'étude, ni les facteurs manipulés, ni les hypothèses sous-jacentes. Plusieurs pré-tests seront donc nécessaires pour parvenir à un scénario réaliste, tout en manipulant correctement les facteurs.

Tableau 8.5 Plan expérimental

Expérimentations Facteurs (niveaux)	EXP 1	EXP 2	EXP 3
MENTIONS DE RVP (Limitées / Nombreuses)	X (2)	- (limitées)	- (nombreuses)
FAMILIARITE (Faible / Elevée)	X (2)	- (élevée)	X (2)
QUANTITE DP (Faible / Elevée)	X (2)	X (2)	X (2)
CONTEXTE (Jeu-Concours / Newsletter / Espace client) (EXP 2) (JC / Enquête) (EXP 3)	- (jeu-concours)	X (3)	X (2)
Mise en situation	- sondage papier /crayon (copie écran d'un formulaire en ligne) - secteur d'activité : télécoms	- sondage en ligne (copie écran d'un formulaire en ligne) - secteur d'activité : télécoms	- situation réelle : invitation à répondre à un questionnaire par email - secteur d'activité : liste de cadeaux
Type d'expérimentation (nombre de facteurs)	Inter-sujets 2 x 2 x 2	Inter-sujets 3 x 2	Inter-sujets 2 x 2 x 2
Echantillon (taille prévue)	Etudiants 8 x 30 = 240	Panélistes 6 x 40 = 240	Clients « réels » Sollicités : 240 000
Facteurs « contrôlés »	- Contexte : JC	- Mentions RVP : limitées - Familiarité : élevée	- Mentions RVP : limitées
Objectifs	- mesurer les effets de ces facteurs sur trois variables de perception	- introduire un maximum de variables de contrôle - mesurer les intentions comportementales	- étudier le lien direct entre les facteurs et les variables dépendantes - accroître la validité des résultats (situation réelle)
Remarques	- peu de variables de contrôle (public aux caractéristiques homogènes)	- une seule variable perceptuelle mesurée (la valeur perçue)	- pas de mesure des variables médiatrices (perceptions et attitude) - échantillon important mais spécifique à une entreprise

Après avoir présenté les facteurs manipulés dans chacune des expériences, nous abordons à présent la notion de contrôle, si fondamentale dans le cadre d'une démarche expérimentale. Nous verrons, tout d'abord pourquoi il est nécessaire de contrôler (4.1) puis quelles sont les variables qu'il faut contrôler (4.2), avant d'évoquer les moyens de les contrôler (4.3).

4 Le contrôle des variables externes

4.1 L'importance de contrôler

Au cours d'une expérimentation, les variables non étudiées, que ce soit délibérément ou par inadvertance, ont un impact sur la généralisation des résultats (Cook et Campbell 1979). Le chercheur doit donc discerner celles qui pourraient exercer une influence sur la relation causale et dont la présence pourrait conduire à une autre interprétation des effets observés.
Ces variables externes, dites aussi variables parasites, doivent être contrôlées c'est-à-dire, soit mesurées puis comprises dans l'analyse statistique, soit maintenues constantes ou homogènes de façon à annuler leur influence sur la variable dépendante (pendant l'expérimentation ou post hoc). La relation de cause à effet est moins perturbée dans le cas d'un maintien constant (Lynch 1982), mais d'autres méthodes peuvent aussi être utilisées, en combinaison avec la première ou si celle-ci n'est pas envisageable. Il est évident qu'en recherche comportementale il existe un grand nombre de variables exogènes et qu'il est difficile (voire impossible) de toutes les prendre en compte (Calder, Phillips et Tybout 1981 et 1982). Cependant, l'identification préalable des variables externes est importante et doit faire l'objet d'une attention particulière de la part du chercheur. Il doit hiérarchiser l'importance de ces dernières pour essayer d'éliminer au mieux leur influence (Aronson et al. 1998). Cronbach (1975) conseille ainsi de mesurer le plus de variables parasites possibles pour ne pas se retrouver démuni au moment de l'analyse, ce qu'il appelle la prise de mesure « hors hypothèses ».

4.2 Les variables à contrôler

En nous aidant de la revue de littérature et de nos études préalables, nous avons pu identifier cinq catégories de variables devant être contrôlées dans le cadre de ce plan expérimental, dans la mesure où elles sont susceptibles d'interférer sur les résultats[143]. Il s'agit du niveau d'implication et des variables individuelles d'ordre sociodémographique et expérientiel d'une part, mais aussi du comportement habituel de réponse et de la qualité perçue de la relation, d'autre part.
Nous justifions l'intérêt de contrôler chacune de ces variables dans les paragraphes suivants.

[143] Pour rappel, nous avons préféré considérer ces variables comme des variables de contrôle, plutôt que comme d'éventuelles variables modératrices, pour les raisons déjà évoquées au chapitre 7 (paragraphe 3).

4.2.1 Le niveau d'implication

Le niveau d'implication d'un individu vis-à-vis d'un objet est lié au degré avec lequel il perçoit l'objet comme lié à lui ou comme un moyen d'atteindre des buts personnels. L'acceptation la plus courante du concept d'implication a trait à l'implication durable d'un consommateur vis-à-vis d'un objet (ou d'une classe de produits/services ou activités)[144].

De nombreux auteurs soulignent l'existence d'une relation forte entre l'implication et le décodage de l'information, démarche forcément importante dans le cadre d'une sollicitation de données personnelles. Ainsi, quand l'implication s'accroît, les individus auraient une plus forte motivation à comprendre les informations présentées et donneraient un sens plus élaboré à celles-ci. Au contraire, quand ce niveau d'implication est faible, les individus se fieraient davantage à des éléments périphériques pour prendre leur décision (Van Kenhove et al. 2002). Les recherches menées dans le cadre de la réponse à des enquêtes confirment l'impact majeur du niveau d'implication. Parmi les facteurs affectant le taux de réponse à des questionnaires administrés par voie postale, la sensibilité du sujet de l'enquête pour la personne sollicitée (que l'on peut considérer comme un proxi du niveau d'implication) s'avère un des (sinon le) critères les plus importants (Kanuk et Berenson 1975, Harvey 1987). Ce constat est confirmé par Van Kenhove et al. (2002) qui montrent que le taux de réponse est plus élevé pour les sujets impliquants que pour les sujets peu impliquants.

L'ensemble de ces raisons justifie pleinement le contrôle du niveau d'implication dans le cadre de notre plan expérimental. En réalité, cette variable ne sera contrôlée que dans les deux premières expériences, le contexte particulier de la dernière justifiant ce choix[145].

4.2.2 Les variables sociodémographiques

Les recherches passées en comportement du consommateur suggèrent que les différences individuelles sont captées par les variables attitudinales et comportementales (Ajzen et Fishbein 1980). Plusieurs auteurs suggèrent toutefois de contrôler ces variables, notamment celles qui concernent le profil sociodémographique du répondant, au moins par précaution.
Nous avons donc choisi de contrôler quatre variables sociodémographiques majeures, à savoir l'âge, le sexe, la catégorie socioprofessionnelle (CSP) et le niveau d'études.

[144] Nous reviendrons plus en détail sur la définition de ce concept quand nous présenterons l'instrument utilisé pour le mesurer (cf. chapitre 9).
[145] Nous reviendrons sur les justifications de ce choix dans le chapitre correspondant (i.e. chapitre 12)

L'ensemble de ces variables sera contrôlé dans les trois expériences, hormis certaines d'entre elles (âge, CSP et niveau d'études) dans le cadre de l'expérience 1, pour des raisons que nous préciserons dans le chapitre correspondant (i.e. chapitre 10).

Les variables liées à l'expertise sur Internet constituent un autre type de variables individuelles dont l'effet doit être contrôlé, étant donné leur incidence possible, compte tenu de notre terrain d'étude (application aux formulaires en ligne).

4.2.3 Les variables expérientielles

Même si tous les individus que nous serons amenés à interroger auront déjà utilisé Internet, ils ne l'auront pas expérimenté de la même manière, étant donné les différences de compétence dans ce domaine. L'expertise d'Internet est fréquemment analysée à travers 2 critères, que nous reprenons ici à savoir : l'ancienneté (en mois et/ou en années) et la fréquence d'utilisation (nombre de fois où l'individu accède au web et/ou à ses emails). Dès lors, même si certains auteurs estiment que l'utilisation de l'email n'est pas toujours un indicateur fiable de l'expérience d'Internet (Miyazaki et Fernandez 2000), nous contrôlerons quand même son impact. A l'inverse, pour des raisons de parcimonie (et de possible multicolinéarité entre les variables), nous ne contrôlerons pas l'effet de l'expérience d'achat en ligne, bien que son incidence sur le processus de réponse face à une sollicitation de données personnelles soit possible. En attestent à la fois la littérature sur le sujet (cf. chapitre 4) et les résultats issus de notre étude exploratoire descriptive (cf. chapitre 6). Phelps, D'Souza et Novak (2001) montrent ainsi que les gens ayant une expérience d'achat à distance (a fortiori sur Internet) ont été sensibilisés au fait que les entreprises collectent et utilisent leurs données personnelles et qu'ils sont par conséquent moins demandeurs de mesures de contrôle.

Plusieurs recherches justifient l'influence possible de l'expérience et de l'utilisation d'Internet dans le cadre de la réponse à une sollicitation de données personnelles. Nous n'indiquerons ici que les résultats les plus probants. Dans le domaine de la participation à une enquête, il a ainsi été montré que les personnes qui perçoivent la technologie Internet comme facile à utiliser sont davantage enclins à répondre (Parker 1992). On sait aussi que les individus experts se sentent moins vulnérables et donc davantage capables de faire face aux risques dans ce domaine. Ils seraient alors plus prêts que d'autres à échanger leurs données contre une réduction de prix (Cheskin Research 2000). De même, une personne experte aura de plus grandes connaissances (et compétences) en matière de sécurité et pourra donc savoir

rapidement si un site est sécurisé (Walczuch et al. 2001). Lynch et Beck (2001) segmentent, pour leur part, les internautes en 2 catégories : « ceux qui ont du pouvoir » (qui utilisent Internet chaque jour) et « ceux qui n'en n'ont pas » (qui l'utilisent moins d'une fois par jour). Ils constatent alors que ces derniers ont une attitude moins favorable vis-à-vis d'internet et sont donc plus préoccupés par les questions de RVP que les premiers. Enfin, Miyazaki et Fernandez (2001) notent que, plus la fréquence d'utilisation du Net augmente, plus les risques perçus en ligne diminuent. A l'opposé, Hoffman et al. (1999) montrent que les préoccupations pour le RVP s'accroissent avec le niveau d'activité en ligne.

L'ensemble de ces résultats confirme la nécessité de contrôler les variables d'expérience.

4.2.4 Le comportement passé

Pour Ehrenberg (1974), le comportement passé serait un plus grand prédicateur du comportement futur que ne le sont les attitudes. Cet effet a été prouvé en matière de réponse à un sondage. Ainsi, plus l'individu a fait l'objet de nombreuses requêtes par le passé et plus ses expériences dans ce domaine ont été difficiles, plus la probabilité que son attitude vis-à-vis des enquêtes soit défavorable est élevée (Goydner 1986). Singer, Groves et Corning (1999) montrent aussi que les croyances des individus à propos des sociétés de sondages sont liées à leurs expériences dans ce domaine. D'autres auteurs (dont Groves, Cialdini et Couper 1992) trouvent toutefois que ce type de facteur ne serait pas directement lié à la participation mais tendrait à produire une prédisposition psychologique qui affecterait la décision.
Ce type d'influence se manifesterait aussi en matière de sollicitation de données personnelles. Pour Bates (1964), les expériences personnelles guideraient le comportement dans toutes les actions qui sont associées au respect de la vie privée. Le rôle des expériences passées a déjà été analysé dans le contexte offline avec des résultats mitigés. Ainsi, pour Culnan (1993), l'expérience passée en matière d'invasion de sa vie privée ne serait pas clairement associée aux attitudes face à l'utilisation secondaire des données mais pourrait affecter la préoccupation pour le respect de la vie privée.

Dans la mesure où nous évaluons la préoccupation pour le respect de la vie privée dans le cadre de notre modèle, nous choisissons de ne pas contrôler l'expérience en matière d'invasion de la vie privée, afin de ne pas risquer de corrélations entre ces deux variables. Nous choisissons toutefois de contrôler la manière dont le consommateur a l'habitude de se comporter face aux demandes de données personnelles des entreprises, car on a vu, dans

l'étude exploratoire descriptive, qu'elle aurait un effet sur la réponse à une sollicitation particulière (cf. chapitre 6). Ce comportement habituel de réponse (COMP_HAB) pourrait donc intervenir sur la partie droite de notre modèle, au niveau des intentions comportementales de réponse (f et g) ou de la réponse effective (i). Il ne sera donc contrôlé que dans les deux dernières expériences puisque la première ne mesure que l'attitude.

4.2.5 La qualité perçue de la relation

La présence de cette variable de contrôle au sein du modèle à tester est principalement liée au fait qu'elle est susceptible d'influencer l'intention du consommateur de poursuivre la relation avec l'entreprise sollicitant les données, en atteste plusieurs recherches menées dans le domaine du marketing relationnel. Il a ainsi été montré que, en se fondant sur leur expérience avec l'entreprise, les consommateurs cherchent à savoir si les informations qu'ils lui ont données par le passé ont été utilisées de manière conforme à leurs attentes. Si c'est le cas, ils continuent la relation alors que, dans le cas contraire, ils peuvent au mieux se désengager, au pire engager un bouche à oreille négatif (Morgan et Hunt 1994). La qualité perçue de la relation pourrait donc avoir une incidence sur l'intention de poursuivre la relation avec la firme ayant sollicité les données, à travers la confiance que celle-ci engendre, amenant alors l'individu à établir des hypothèses plus ou moins favorables.
Cela justifie en tout cas son contrôle dans le cadre de nos expériences. En réalité, cette intention de poursuivre la relation n'est évaluée que dans les deux dernières expériences. Toutefois, cette variable ne sera contrôlée que dans l'expérience 3, puisque dans l'expérience 2, elle n'est pas susceptible d'influencer l'effet des facteurs manipulés.

Le tableau 8.6 indique les variables qui seront contrôlées dans chacune des expériences.

Tableau 8.6 Variables contrôlées dans chacune des expériences

Variables de contrôle	EXP 1	EXP 2	EXP 3
Implication	X	X	
Sociodémographiques	X (*)	X	X
Expérience et utilisation d'Internet	X	X	X
Comportement habituel de réponse		X	X
Qualité perçue de la relation			X

(*) Seul le sexe sera contrôlé, pour des raisons d'homogénéité de l'échantillon sur les autres variables

Après avoir indiqué les variables à contrôler, voyons de quelle manière ce contrôle s'opère.

4.3 Les modes de contrôle

A l'instar d'Herbert (2005 b), nous distinguons deux modes de contrôle principaux[146], présentés ci-dessous. Le contrôle de la situation expérimentale n'est pas évoqué ici car il s'agit d'une démarche générale visant à soigner la procédure, de façon à éliminer un maximum de source externe de variance pour mieux isoler l'effet des facteurs manipulés.

1. *Le contrôle par affectation aléatoire*

Il s'agit du contrôle des variables indépendantes par affectation aléatoire des unités expérimentales aux différents traitements. Même si cela n'est pas toujours possible, l'affectation aléatoire des unités expérimentales est une des règles d'or de l'expérimentation car elle garantit un meilleur contrôle des effets indésirables. Cependant, la randomisation parfaite, en suivant une table des nombres au hasard, n'est pas toujours possible, surtout dans les recherches de terrain en sciences sociales. Le chercheur peut cependant s'arranger pour « choisir » les unités qui reçoivent le traitement (Herbert 2005 b). Par conséquent, même dans des conditions difficiles, il est toujours possible de s'acquitter d'un certain degré de contrôle qui améliorera la qualité des inférences (Cook et Campbell, 1979).

2. *Le contrôle par « prise de conscience »*

Herbert (2005 b) désigne ainsi le mode de contrôle qui consiste à repérer et à prendre en compte (de façon active ou passive) les sources d'interférences pouvant troubler la relation de cause à effet. Cette prise en compte se traduit, soit par le maintien constant de la variable externe de façon à neutraliser les effets de sa variation, soit par sa simple mesure et son incorporation dans le traitement et l'analyse des données (analyse de covariance).

Outre ces variables individuelles qui seront contrôlées par ces deux modes, nous devrons aussi contrôler l'effet des facteurs non manipulés dans chacune des expérimentations. De façon classique, nous les fixerons à un niveau défini au départ et les maintiendrons constants à travers les conditions expérimentales, tout au long de l'expérience, de manière à annuler tout effet possible sur les variables médiatrices et/ou dépendantes mesurées.

[146] Nous ne tenons pas compte du contrôle par groupe de contrôle puisque cette démarche ne sera pas utilisée ici

Nous abordons, pour terminer ce chapitre sur la méthodologie de la recherche, la question de la validité du plan expérimental.

5 Les questions de validité

Le concept de validité introduit par Campbell et Stanley (1963) puis développé par Cook et Campbell (1979) est la source la plus intarissable de débats méthodologiques entre chercheurs en sciences sociales. Malgré les polémiques sur le sujet, la plupart des auteurs distinguent quatre types de validité, définis à l'origine par Cook et Campbell (1979).

Les deux premiers types portent sur l'ensemble des conclusions concernant la relation causale entre les variables manipulées et mesurées, il s'agit de :

- la **validité statistique** qui juge de la pertinence des inférences statistiques à partir des effets observés.
- la **validité interne** qui s'inquiète de savoir si les effets observés peuvent bien être attribués à une relation causale précise.

Les deux autres types de validité ne se préoccupent plus des relations entre les variables opérationnalisées mais plutôt de ce qu'elles représentent théoriquement. Nous y trouvons :

- la **validité de construit** qui concerne la généralisation possible des résultats empiriques à des construits et des relations théoriques d'un degré d'abstraction supérieur.
- la **validité externe** qui concerne la « généralisabilité », c'est-à-dire la possibilité de transposer les résultats à des populations, des situations et des moments différents.

Il est évident qu'aucun type de validité ne peut être négligé. Pris indépendamment l'un de l'autre, chacun, s'il ne peut être garanti ou conforté, posera des problèmes au chercheur, à des degrés divers et à des moments différents dans son processus de recherche. De plus, les validités fonctionnent généralement comme des vases communicants : le renforcement d'un type de validité entraîne l'affaiblissement d'un autre. Par exemple, utiliser un échantillon homogène peut consolider la validité interne mais amoindrir la validité externe.

Dans ces conditions, il est illusoire de s'attendre à ce qu'une recherche soit irréprochable à tous points de vue. Par conséquent, il revient au chercheur d'arbitrer et de gérer ses priorités. Cook et Campbell (1979) parlent de « *trade-off* » et hiérarchisent l'importance des validités en fonction des objectifs et du type de recherche. On peut toutefois noter que la validité interne revient toujours en première position dans la hiérarchie, quels que soient les enjeux de la recherche. De plus, elle rejoint la validité externe sur une préoccupation particulière : le

contrôle des variables contextuelles. En des termes simples, la validité interne permet en effet de s'assurer que l'on a bien mesuré la relation causale que l'on voulait mesurer et elle seule. Quand à la validité externe, elle est forte si les résultats sont robustes, c'est-à-dire réplicables, ce qui n'est possible que si les effets observés ne sont pas liés à une situation contextuelle (c'est-à-dire à un dispositif expérimental particulier, à un moment dans le temps ou à un échantillon de sujets spécifique).

Pour chaque type de validité, Cook et Campbell (1979) ont établi une liste des « menaces » qui peuvent leur porter préjudice. Nous revenons sur celles relatives à la validité interne qui sont les plus importantes à assurer dans le cadre de notre étude (5.1). Nous abordons ensuite la question de la validité externe (5.2).

5.1 La validité interne de la recherche

Il existe plusieurs menaces à la validité interne de la recherche que nous aborderons tour à tour en indiquant, notamment, les mesures que nous avons prises pour les circonscrire.
Nous résumons l'essentiel de cette démarche dans le tableau 8.7.

L'effet d'histoire : Ce sont toutes les modifications dans l'environnement du participant qui ont lieu en même temps que l'expérimentation. Elles ne sont pas forcément liées au phénomène étudié mais peuvent influencer son attitude, son comportement et donc ses réponses. Par définition, ces événements sont incontrôlables par le chercheur. Il doit donc, soit les isoler, soit les prendre en compte dans l'analyse de ses résultats. Dans le cadre de notre recherche, le participant pourrait être influencé, soit par une « affaire » d'invasion de la vie privée relatée par les médias[147], soit par une modification des dispositifs législatifs dans ce domaine. Toutefois, à notre connaissance, aucun scandale de grande ampleur n'a été relayé par la presse française pendant la période durant laquelle nos trois expérimentations ont été menées (de septembre 2005 à mars 2006). De plus, aucun changement législatif n'est intervenu à cette époque, la nouvelle loi sur l'Economie Numérique (LCEN) ayant été ratifiée mi 2004. Cette période était donc relativement neutre pour les participants, tout effet d'histoire étant dès lors circonscrit, d'autant que la durée de chaque expérimentation était relativement restreinte.

[147] Plusieurs « affaires » ont ainsi été relatées par les médias dans les années qui ont précédé notre étude, principalement aux Etats-Unis. Nous pouvons citer, entre autres, le cas du site Amazon qui a été soupçonné de pratiquer des prix différents d'après les informations détenues sur leurs clients.

L'effet de maturation : Il s'agit des modifications que le temps peut entraîner sur le sujet expérimental, qui sont indépendantes de la relation causale mais peuvent l'influencer. Le sujet peut avoir faim, froid, son humeur peut varier, il peut éprouver un sentiment de lassitude. Pour se prémunir de ces éléments (notamment de la lassitude), une de nos préoccupations majeures a été de réduire au maximum le temps de passation du questionnaire (d'où le choix d'échelles avec un nombre limité d'items[148]) et d'augmenter sa lisibilité. Seule l'expérience 3 s'écarte un peu de ce principe, notamment dans le cas du questionnaire long (qui comprenait 42 questions)[149]. Nous précisons toutefois qu'il s'agissait là d'une opération réelle et que les individus pouvaient, à tout moment, abandonner. Dans le cadre de notre dispositif expérimental, l'effet de maturation devait donc être réduit ou, au moins, limité.

L'effet de mortalité expérimentale : Ce biais est présent lorsque des sujets arrêtent de répondre au cours d'une expérience ou refusent de participer à nouveau lors d'un deuxième questionnement. Ils sabordent alors la distribution aléatoire ou modifient la composition des groupes, biaisant ainsi le principe de comparaison. Dans le cas de nos deux premières expériences, le questionnement n'ayant lieu qu'en une seule fois, cet effet est sérieusement limité. De plus, les sujets qui stopperont le questionnaire avant la fin seront supprimés de l'échantillon. Nous prévoirons dès lors un surplus de participants dans chaque groupe expérimental de façon à en conserver un nombre raisonnable en cas d'abandon. L'expérience 3 concerne, quant à elle, un dispositif en deux phases (l'expérimentation puis un post-test). Il est probable que certains individus ayant participé à la première phase refuseront de prendre part à la seconde. Des précautions particulières seront alors prises, notamment en ne retenant que les participants aux deux phases, malgré les biais que cela engendre. Nous reviendrons sur ce sujet dans le chapitre correspondant (i.e. chapitre 12).

L'effet d'instrumentation : il est lié à la manière dont les instruments de mesure sont utilisés. Un effet artificiel peut ainsi être obtenu si l'intervalle de réponses change entre le début et la fin du questionnaire ou si l'instrument est modifié entre le pré-test et le post-test. L'effet d'instrumentation peut aussi apparaître si le questionnaire n'est pas administré de la même manière ou dans les mêmes conditions à chaque fois. Cet effet est donc maîtrisable à condition de ne changer ni d'instrument, ni de mode de réponse, ni de mode d'administration au cours de l'expérimentation. Sa menace est donc écartée dans le cadre de notre dispositif

[148] Nous reviendrons sur ce sujet dans le chapitre suivant (9) qui présente les différents instruments de mesure.
[149] Les modalités précises de ce facteur seront présentées en détail dans le chapitre correspondant (chapitre 12)

expérimental puisque les collectes ont, à chaque fois, été réalisées sur une période de temps limitée (moins de 10 jours dans les expériences 1 et 2, moins de trois semaines dans la troisième), avec un même outil de collecte et dans les mêmes conditions d'administration.

L'effet d'examen ou de familiarité au test : Si un même phénomène est mesuré plusieurs fois auprès d'un même sujet, sa première réponse influencera les suivantes. Dans ce cas, il est possible d'enregistrer une variation de réponse qui n'est pas due à la manipulation du facteur mais au fait que le sujet réponde en s'appuyant sur le souvenir de ses réponses passées. Cet effet est éliminé ici puisque chaque répondant n'est confronté qu'à un seul type de traitement (designs inter-sujets), et ce, dans nos trois expériences.

L'effet de sélection : il existe lorsque les groupes expérimentaux qui sont comparés sont trop différents. Dans ce cas, c'est la différence entre les groupes qui peut être à l'origine d'une variation et pas la relation causale étudiée. Il y a aussi effet de sélection lorsque le chercheur prélève certains sujets pour les éliminer de l'analyse. La seule façon de se prémunir radicalement de cet effet est l'affectation aléatoire des sujets aux conditions expérimentales. Dans nos trois expériences, nous assignerons ainsi les individus aux différents traitements, de façon aléatoire. La répartition des sujets sera ensuite testée *a posteriori* (à l'aide d'un test de chi-deux), de façon à juger de l'homogénéité des groupes.

L'effet de contamination : Si les traitements impliquent l'utilisation d'informations spécifiques qui diffèrent d'un groupe à l'autre (comme des scénarii par exemple) et que les groupes échangent ces informations entre eux, la mesure de la relation causale peut être affectée et invalidée. Pour éviter cet inconvénient, les groupes peuvent être confinés en laboratoire. Sur le terrain, le chercheur peut aussi isoler ces groupes (par éloignement géographique ou recueils simultanés d'information). Cette menace ne pose aucun problème dans le cas des expériences 2 et 3 puisque les individus ne se connaissent pas et ne sont pas en contact. Dans le cas de l'expérience 1, les individus étant étudiants dans la même faculté, nous diversifierons les filières de recrutement de manière à éviter tout effet de contamination.

Le tableau suivant résume ces menaces et les mesures prises dans chaque cas pour y pallier.

Tableau 8.7 Les menaces à la validité interne et les moyens de s'en prémunir

Menaces à la validité interne	Définitions	Expériences pour lesquelles la menace est réelle			Moyens utilisés pour s'en prémunir
		EXP 1	EXP 2	EXP 3	
Effet d'histoire	Evénements autres que l'expérimentation survenant au cours de la recherche (Jones 1999)				- recherche d'événements éventuels susceptibles d'intervenir - période d'interrogation courte
Effet de maturation	Changements pouvant survenir chez les personnes questionnées au cours de la recherche (Jones 1999)			X	- réduction du temps de passation du questionnaire
Effet de mortalité expérimentale	Modification de la répartition des groupes suite à l'abandon ou au refus de participer à une deuxième étude			X	- questionnement en une seule fois - surplus de sujets dans chaque groupe - ne retenir que les participants aux deux phases (cas de l'expérience 3)
Effet d'instrumentation	Modification des instruments de mesure au cours du temps (Jones 1999)				- période d'interrogation courte - un seul outil de collecte - conditions d'administration identiques
Effet d'examen	Probabilité de changement due à la répétition d'un test ou d'une mesure				- un seul type de traitement par sujet (design inter-sujet)
Effet de sélection	Possibilité de différences entre les groupes expérimentaux				- affectation aléatoire des sujets aux différents traitements - homogénéité de la répartition testée *a posteriori* par un test de chi-deux
Effet de contamination	Possibilité que les sujets de l'expérience apprennent par d'autres l'objet de l'étude	X			- isolement des sujets (ou des groupes) - diversification des filières de recrutement (cas de l'expérience 1)

5.2 La validité externe de la recherche

La méthode expérimentale est souvent accusée d'être limitée en termes de validité externe. Les résultats qu'elle permet d'obtenir doivent en effet être analysés avec précaution car leur généralisation est souvent réduite. Evaluer la validité externe d'une recherche est complexe. Il ne faut pas seulement estimer le degré de réalisme de l'expérience ou la représentativité de la population étudiée, mais aussi considérer la robustesse des relations causales de la recherche (*i.e.* leur degré d'indépendance aux variables contextuelles) (McGrath et Brinberg 1983).

Pour répondre aux conditions de validité externe, Perrien, Chéron et Zins (1983) stipulent que le chercheur doit, au préalable, répondre positivement aux deux questions suivantes.

1. *Les caractéristiques des variables indépendantes sont-elles conformes à la réalité ?*

L'étude de l'impact de l'exposition à une sollicitation de données personnelles sur le processus de réponse rend nécessaire la manipulation de différents facteurs. Dans le cadre de notre dispositif expérimental, 4 d'entre eux ont été sélectionnés pour être manipulés alternativement au cours de trois expériences successives. Nous avons été particulièrement attentifs, lors de la mise en œuvre de ce dispositif, à faire en sorte que les facteurs manipulés soient le plus conforme possible à la réalité[150]. Nous avons également choisi des terrains d'étude (ou des secteurs d'activités) qui soient suffisamment impliquants pour les répondants. Enfin, dans la troisième et dernière expérience, nous avons mis en place un partenariat avec une entreprise réelle afin d'interroger de vrais clients, dans un contexte réel.

2. *Les unités tests choisies sont-elles représentatives de la population ?*

Nous avons déjà développé ce propos dans le paragraphe concernant le choix des échantillons interrogés (2.3). Nous ne nous étendrons donc pas sur la question. Pour rappel, la population interrogée dans la première expérience est constituée d'étudiants, ce qui limite la généralisation des résultats à ce type de population. Dans l'expérience 2, nous nous rapprochons du degré de représentativité en interrogeant un échantillon de panélistes. Enfin, dans l'expérience 3, le choix de clients réels pallie les limites de généralisation qui peuvent se produire lorsque des étudiants constituent les unités expérimentales.

[150] Nous revenons sur les précautions prises dans ce cadre dans les chapitres suivants (i.e. 10, 11 et 12)

Conclusion du chapitre 8

Le choix d'un design expérimental pour tester notre modèle se justifie à la fois par le type de relations à mesurer (de cause à effet) et par la nécessité d'être dans un environnement contrôlé, de par la diversité des variables susceptibles d'intervenir. Quatre facteurs principaux seront ainsi manipulés : les mentions de respect de la vie privée (RVP), la familiarité avec l'entreprise, la quantité de données demandées et le contexte de sollicitation.

Le plan expérimental choisi, comprenant trois expériences successives, nous permet de limiter tour à tour les principaux types de biais recensés (biais de non réponse, de couverture et du répondant). Nous choisissons, en outre, de diversifier le type d'échantillon interrogé (étudiants, panélistes et clients réels) afin de contrôler successivement chacun de ces biais.

Le modèle finalement testé, dans le cadre de ce travail doctoral, est présenté à la figure 8.2. Il sera décomposé ultérieurement en trois sous-modèles, correspondant à chacune des expériences menées (cf. chapitres 10, 11 et 12). Le plan expérimental mis en œuvre répond, quant à lui, à deux règles de base : conserver un design « raisonnable » et éviter tout risque d'orthogonalité entre les facteurs manipulés.

Le contrôle des variables externes est une opération nécessaire dans toute démarche expérimentale. Elle consiste à identifier l'ensemble des variables susceptibles d'interférer sur les résultats et à en contrôler l'effet, soit en les maintenant constantes, soit en les mesurant et en les incluant dans l'analyse statistique. L'affectation aléatoire des individus interrogés aux différents traitements s'avère, en outre, le meilleur moyen de contrôler les effets indésirables.

Plusieurs menaces peuvent réduire la validité interne ou externe d'une expérimentation. Ainsi, l'effet de maturation et de mortalité expérimentale est surtout préoccupant dans le cadre de la troisième expérience, tandis que l'expérience 1 fait courir le risque d'un effet de contamination. Après avoir identifié et défini ces menaces, nous avons indiqué les moyens mis en œuvre pour les limiter, dans chaque expérience où elles pouvaient intervenir.

Ayant présenté la méthodologie de la recherche, nous passons au chapitre suivant qui concerne le choix et le développement des instruments de mesure correspondant à l'ensemble des variables identifiées dans le modèle à tester.

Partie II – EXPLORER ET BATIR

Sollicitation de données personnelles
et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

CHAPITRE 9
CHOIX ET DÉVELOPPEMENT
DES INSTRUMENTS DE MESURE

« In using, evaluating or developing multi-item scale measures,
a number of guidelines and procedures are recommanded
to help ensure that the measure is as psychometrically sound as possible ».

Bearden, Netemeyer and Mobley (1993)

INTRODUCTION GÉNÉRALE

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

CONCLUSION GÉNÉRALE

Introduction du chapitre 9

Le modèle à tester et le plan expérimental visant à vérifier les propositions et hypothèses de recherche ayant été exposés dans les deux chapitres précédents, l'objectif de ce chapitre est de présenter l'opérationnalisation des principales variables de la recherche et de vérifier la qualité psychométrique des instruments de mesure utilisés.

Nous exposerons, dans une première section, la démarche de développement et de validation d'une échelle, en évoquant notamment les principaux critères de sélection d'un instrument de mesure, tels que définis par la littérature et appliqués au contexte de cette recherche (1.1). Sera ensuite abordé le processus que nous comptons suivre pour développer les instruments de mesure nécessaires à la vérification de nos propositions et hypothèses de travail (1.2). Nous terminerons cette section par une classification de l'ensemble des construits de la recherche, aboutissant au choix du type de méthode à privilégier (1.3).

La section suivante présente le résultat du développement de l'ensemble des instruments de mesure nécessaires à la mise en œuvre de l'expérience 1. Nous commencerons par aborder le processus de sélection des items et les résultats de l'analyse factorielle exploratoire, pour chacun des construits (2.1). Nous présentons, dans un second temps, les résultats de la réplication de l'analyse factorielle puis ceux qui concernent l'analyse confirmatoire (2.2).

La section 3 suit le même schéma que la section 2 mais concerne l'ensemble des instruments de mesure nécessaires à la mise en œuvre de l'expérience 2 et n'ayant pas été développés pour les besoins de l'expérience 1.

1 La démarche de développement et de validation des échelles de mesure

Nous verrons, dans un premier temps, sur quels critères se prend la décision de sélectionner (ou de créer) tel ou tel instrument de mesure (1.1), avant d'exposer le processus de développement et de validation d'une échelle de mesure (1.2). Nous terminerons cette section en présentant une typologie des instruments utilisés dans le cadre de cette recherche (1.3).

1.1 Les critères de sélection d'un instrument de mesure

Le choix des instruments de mesure doit répondre aux exigences de la recherche en termes méthodologiques (ou psychométriques) et opérationnels (Vernette 1991). Nous détaillons tour à tour ces deux points.

1.1.1 La qualité méthodologique des instruments de mesure

La qualité méthodologique (ou psychométrique) des instruments de mesure constitue le premier critère de choix. Si possible, les instruments de mesure doivent être élaborés à partir des recherches existantes. Ils peuvent alors, ou être repris tels quels, ou être adaptés au contexte étudié. Lorsqu'aucun instrument correspondant au construit à mesurer n'est disponible au sein de la littérature, soit parce que le construit est entièrement nouveau, soit parce que les instruments trouvés ne sont pas appropriés ou ont été bâtis dans un contexte si différent qu'ils ne sont pas généralisables à la situation étudiée, il convient d'élaborer un nouvel instrument. Dans notre cas, la majorité des construits étudiés sont soit nouveaux (cas de la sensibilité perçue par exemple[151]) soit suffisamment différents de la manière dont ils sont considérés habituellement (cas de la pertinence perçue par exemple[152]) pour justifier le recours à une procédure ad hoc. Nous nous sommes toutefois largement inspirés, à la fois de la littérature existante et de l'étude qualitative, pour générer les items correspondants à ces construits. Enfin, pour assurer une validité de contenu aux échelles nouvellement créées, nous avons soumis les items au jugement des experts et des consommateurs. Elles ont ensuite suivi une démarche de développement et de validation précise, qui sera présentée au point 1.2.

[151] Dans les recherches que nous avons étudiées, ce concept est toujours considéré comme un facteur à manipuler, sous la forme du type de données demandées. A notre connaissance, aucun chercheur n'a jusqu'ici évalué la sensibilité perçue d'une donnée ou d'un ensemble de données, par le biais d'une échelle de mesure.

[152] La pertinence est un concept déjà utilisé dans d'autres champs de recherche mais jamais mesuré en ce qui concerne la sollicitation de données personnelles.

1.1.2 La qualité opérationnelle des instruments de mesure

Par qualité opérationnelle, nous entendons la capacité de l'instrument à être compris et à générer une réponse qui reflète l'opinion réelle du répondant, tout en minimisant les capacités cognitives à mobiliser. Dans cette recherche, la qualité opérationnelle des instruments doit faire l'objet d'une attention particulière dans la mesure où la collecte de données suit une logique expérimentale, par le biais de questionnaires auto-administrés. Deux éléments ont conduit au choix des instruments de mesure : le principe de parcimonie d'une part et les caractéristiques de l'échelle de mesure d'autre part. Ils sont tour à tour présentés.

Le principe de parcimonie

Le principe de parcimonie est une caractéristique que les sciences chérissent depuis le 14ème siècle. Les chercheurs ont ainsi souvent indiqué que l'objectif de la science n'était pas seulement d'expliquer, de prédire et de comprendre le monde dans lequel nous vivons, mais surtout de le faire de la manière la plus efficiente possible. Lambert et Brittan (1970) discutent des raisons pour lesquelles le principe de parcimonie (ou de simplicité) est si important en science : « l'hypothèse la plus simple est souvent la plus élégante, la plus facile à comprendre, à se rappeler, à communiquer et à travailler avec ».

De nombreux chercheurs insistent sur les critères de non-redondance et de parcimonie lorsqu'il s'agit de choisir ou de créer un instrument de mesure. Cronin et Taylor (1992) expliquent ainsi qu'il n'est pas nécessaire d'utiliser l'échelle de SERVQUAL en 44 items quand un seul item suffit à mesurer la satisfaction du consommateur. Les instruments de mesure doivent donc être simples afin de permettre une administration rapide, qui minimise la fatigue et/ou la mortalité expérimentale (Vernette 1991).
Cette règle de parcimonie nous conduit alors à privilégier des mesures avec un faible nombre d'items (pas plus de 5 par construit). Ce choix se justifie d'autant plus que le nombre de construits à mesurer dans chaque expérience est élevé (jusqu'à six construits nouveaux pour l'expérience 1) et que nous souhaitons limiter autant que possible la durée de passation du questionnaire. De plus, la majorité des construits étudiés sont relativement « étroits » (au sens de Nunnally 1978), c'est-à-dire qu'ils sont simples et peuvent donc être facilement

appréhendés à l'aide d'un nombre limité d'items[153]. De plus, hormis la valeur perçue (par essence multidimensionnelle), tous les concepts mesurés dans la thèse sont unidimensionnels, ce qui réduit le nombre d'items nécessaires à la couverture du domaine du construit.

Les caractéristiques de l'échelle de mesure

Il s'agit ici de définir le type d'échelle utilisée et le nombre d'échelons à retenir.
Les échelles multi-items utilisées dans cette recherche sont souvent de type Likert et exceptionnellement de type sémantique différentiel, notamment pour l'échelle de valeur perçue. Comportant pour la plupart d'entre elles cinq échelons lors de la phase exploratoire, elles ont finalement été « transformées » en sept échelons, de manière à accroître la variance et la précision des réponses (Rieunier 2000) et à assurer la solidité des échelles (Darpy 1999). Les échelles d'intentions comportementales (intention de répondre, de mentir et de poursuivre la relation) utilisées pour l'expérience 2 ont cependant gardé cinq échelons, pour se conformer au format habituel de ce type d'échelle (de « oui certainement » à « non certainement pas »). Même si cela réduit l'ambition d'homogénéité au sein du questionnaire, ce choix est essentiellement dicté par un souci de cohérence avec la littérature. De plus, dans la mesure où l'ensemble des intentions comportementales ont été mesurées sur une même page du questionnaire, nous ne pensons pas que cela ait pu gêner en quoi que ce soit les répondants.

Le recours au point neutre (choix d'un nombre d'échelons impair) se justifie par l'idée que les répondants peuvent ne pas avoir d'opinion tranchée ou avoir une opinion « moyenne » sur certains des items proposés. Bien que ce choix ne les encourage pas à se prononcer pour ou contre l'argument avancé, il offre l'avantage de conduire à une image plus représentative de la réalité de leurs perceptions.

Voyons, à présent, le processus de développement et de validation d'un instrument de mesure.

1.2 Développement et validation d'un instrument de mesure

Nous présentons ci-après les étapes suivies pour développer les instruments de mesure créés spécifiquement pour les besoins de la recherche.

[153] Ce principe ne nous semble toutefois pas devoir conduire à une vision mono item de la mesure qui mènerait à une ultra-simplification, fort dommageable, selon nous, à la validité des instruments et donc de la recherche.

La construction de ces échelles s'est faite en suivant l'essentiel de la démarche préconisée par Churchill (1979), enrichie des travaux de Rossiter (2002)[154]. Nous distinguons ainsi 3 phases importantes, comprenant chacune plusieurs étapes. Celles-ci sont présentées dans le tableau ci-dessous et seront brièvement exposées dans les paragraphes suivants.

Tableau 9.1 Etapes de développement et de validation d'un instrument de mesure

Phase	Démarche de Churchill (1979)	Techniques utilisées
1	Spécifier le domaine du construit	Définition du concept
	Générer un échantillon d'items	- Rédaction d'un ensemble d'items - Pré-test qualitatif et quantitatif auprès de consommateurs (une dizaine maximum) - Soumission à des experts
2	Collecter des données	Collecte 1 (pré-test)
	Purifier les mesures	- Analyse factorielle exploratoire (AFE) - Statistiques descriptives et calcul de l'alpha de Cronbach
3	Collecter des données	Collecte 2 (données récoltées au cours de l'expérimentation correspondante)
	Estimer la fiabilité	- Analyse confirmatoire (AFC) - Calcul de l'alpha de Cronbach et du rhô de Jöreskog
	Estimer la validité	- Estimation de la validité convergente - Estimation de la validité discriminante - Estimation de la validité prédictive

1.2.1 Spécification du domaine du construit et génération des items

Première étape dans le développement d'une échelle de mesure, la spécification du domaine du construit consiste à présenter une définition claire du concept et à en délimiter les contours. Nous nous aiderons pour cela, à la fois de la littérature existante, et des résultats de l'étude qualitative exploratoire. Cette définition est importante, d'une part pour ne pas risquer

[154] Nous nous démarquerons en effet légèrement de cette démarche dans la mesure où, devant développer un grand nombre d'instruments de mesure simultanément, ceux-ci correspondants en outre à des concepts relativement étroits, nous faisons clairement le choix d'une mesure simple de ces construits. Nous limiterons donc, dès le départ, le nombre d'items générés et conservés, malgré les limites auxquelles cette décision conduit et l'apparente contradiction avec la démarche préconisée par Churchill qui veut que la purification de la mesure soit de nature « statistique ». Le fait que, malgré cela, nous cherchions à établir et à vérifier de manière objective les qualités psychométriques des instruments développés vient, selon nous, réduire le « biais » engendré par ce choix et conduit à ne pas remettre en cause les échelles créées.

d'inclure des items ne correspondant pas au concept et d'autre part pour aider les experts à se faire une opinion sur les items proposés. Une fois le construit défini, il s'agit alors de générer un maximum d'items en s'aidant pour cela de la littérature existante, des verbatims recueillis lors des entretiens semi-directifs et de l'intuition. Précisons que, pour un même concept, nous nous sommes efforcés de développer des items suffisamment différents entre eux pour éviter les risques de redondance, cause de surestimation de l'alpha. L'ensemble de ces items est alors soumis à un échantillon de consommateurs de manière qualitative (recueil de leurs ressentis en face à face) puis quantitative (pré-test). Un premier ensemble d'items est alors soit supprimé soit reformulé en tenant compte des remarques indiquées par les répondants.
Ce nouvel échantillon d'items est alors soumis à l'évaluation d'experts[155] qui sont, ou des spécialistes du domaine de recherche, ou experts en matière de développement d'échelle. Aidés en cela par la définition du construit issue de l'étape précédente, nous leur demandons d'évaluer chaque item proposé selon une classification en 3 points :
- TR (Très Représentatif) : l'item représente bien le construit qu'il est censé mesurer ; il peut donc être conservé tel quel ou avec des modifications mineures
- MR (Moyennement représentatif) : l'item ne représente pas complètement le construit qu'il est censé mesurer ; il faut envisager des modifications substantielles voire le supprimer
- PR (Peu ou Pas Représentatif) : l'item ne mesure pas le construit ; il faut le supprimer
Nous demandions aussi aux experts, en plus de cette évaluation, de faire des commentaires pour chaque item (notamment pour ceux jugés moyennement représentatifs) et de proposer d'éventuelles améliorations.

Pour information, l'ensemble des instruments développés pour les besoins de la première expérimentation (confidentialité, sensibilité et pertinence perçue, attitude face à la communication des données) totalisait 36 items. Suite au pré-test auprès des consommateurs, 26 ont été conservés et proposés pour évaluation aux experts et l'objectif fixé était de retenir les items remportant le plus d'adhésion[156]. Nous avons alors conservé les items ayant été jugés par tous les experts comme très représentatifs (TR) ou n'ayant pas plus de 2 « moyennement représentatif » (MR). Tout item ayant au moins un jugement « peu représentatif » (PR) était immédiatement supprimé. Au final, 16 items ont été conservés, soit une moyenne de 4 items par concept, ces items faisant alors l'objet d'une analyse factorielle exploratoire dont nous exposons brièvement les principes dans le paragraphe suivant.

[155] 7 experts ont été consultés pour les construits de l'expérience 1 ; 5 d'entre eux ont accepté de répondre.
[156] Conformément au principe de parcimonie, nous avions fixé un maximum de 5 items à retenir par concept, pour ne pas alourdir l'instrument de collecte et donc risquer une forte mortalité expérimentale. Nous avons déjà expliqué à plusieurs reprises l'origine de ce choix, notamment à la page précédente (en note de bas de page).

1.2.2 Analyse factorielle exploratoire et purification des mesures

La démarche suivie pour tester la validité et la fiabilité des instruments de mesure lors de l'analyse factorielle exploratoire (AFE) est classique. Nous ne nous étendrons donc pas dessus, renvoyant les lecteurs intéressés aux principaux ouvrages dans ce domaine. Nous indiquons, dans le tableau suivant (tableau 9.2) les principales étapes suivies ainsi que les techniques et/ou critères utilisés pour prendre les décisions relatives à cette démarche.

Tableau 9.2 Etapes suivies lors de l'analyse factorielle exploratoire (AFE)

N°	Etapes de l'analyse factorielle exploratoire	Techniques et critères utilisés
1	Vérifier l'adéquation des données pour l'analyse factorielle (AF)	- Test de Bartlett ($p < 0,05$) - Test de KMO (valeur > 0,5)
2	Choisir le type d'analyse factorielle à mener	- Analyse factorielle sans rotation si le construit est unidimensionnel - Avec rotation s'il est multidimensionnel (orthogonale ou oblique)
3	Analyser la qualité de représentation des variables (« communalité »)	- Pourcentage de variance expliquée par chaque variable > 0,5 (en dessous de ce seuil[157], la variable est exclue et une nouvelle AF est menée)
4	Analyser la variance expliquée par les facteurs	- Facteur retenu si valeur propre de l'axe > 1 (règle de Kaiser) - Facteurs retenus si variance cumulée > 60% (Malhotra 1993)
5	Analyser les corrélations entre les variables et les axes	- Corrélation > 0,5 - Différence de contribution entre l'axe principal et un autre axe > 0,3
6	Estimer la fiabilité des items	- Alpha de Cronbach > 0,7 (Fornell et Larker 1981, Nunnally 1978)

Concernant l'estimation de l'alpha, au cours d'une méta-analyse menée sur ce coefficient, Peterson (1995) suggère d'évaluer sa qualité selon le type de recherche effectuée, exploratoire ou confirmatoire (cf. tableau 9.3 ci-dessous). Là encore, nous préconisons une relative souplesse dans l'application de ces règles. Ainsi, un item détériorant (légèrement) l'alpha pourrait être retenu s'il contribue (fortement) à la validité de contenu du construit.

[157] Nous nous réservons cependant le droit de conserver un item ayant une qualité de représentation très légèrement en dessous du seuil communément admis de 0,5 (jusqu'à 0,47), à condition que cela se justifie théoriquement et/ou empiriquement.

Tableau 9.3 Interprétation du Coefficient Alpha de Cronbach selon Peterson (1995)

Coefficients	Tests réalisés	Valeurs obtenues	Interprétation
Alpha de Cronbach global	Test de cohérence interne du construit	0,6 < alpha < 0,8 alpha > 0,8 alpha > 0,9	Cas d'une recherche exploratoire Cas d'une recherche confirmatoire Risque de redondance entre les items
Alpha de Cronbach par item	Test de cohérence interne du construit si un item est enlevé	Si alpha sans item > alpha global	L'item détériore l'alpha global => l'item doit être exclu et une nouvelle analyse factorielle doit être réalisée

1.2.3 Analyse confirmatoire et propriétés psychométriques

L'analyse confirmatoire (AFC) a pour but de confirmer la structure factorielle sous-jacente aux données. Elle fournit également des indices d'ajustement et des valeurs d'indicateurs qui permettent d'apprécier la validité et la fiabilité de l'échelle construite. Elle doit s'effectuer sur une seconde collecte de données et est précédée d'une réplication de l'analyse factorielle exploratoire, afin de s'assurer que la structure factorielle du concept étudié est bien celle découverte précédemment. Au besoin, il peut être nécessaire de procéder à une nouvelle épuration des items, sur la base de leur « communalité » ou de leur saturation.

Dans notre cas, cette analyse confirmatoire sera réalisée à l'aide du logiciel AMOS 5.0. Les étapes à suivre pour mener à bien une telle analyse sont présentées dans le tableau suivant :

Tableau 9.4 Etapes de réalisation d'une analyse confirmatoire (AFC)

N°	Etapes de l'analyse confirmatoire	Démarche
1	Vérification préalable à l'analyse des données	- Elimination des valeurs manquantes - Vérification de la taille de l'échantillon - Vérification de la normalité des données
2	Spécification du modèle	- Utilisation du modèle conceptuel à tester - Utilisation des résultats issus de la réplication de l'analyse factorielle exploratoire
3	Identification des contraintes	- Choix de la méthode d'estimation - Définition des paramètres à calculer - Fixation des contraintes à certains paramètres
4	Test du modèle	- Examen des indices d'ajustement
5	Vérification des propriétés psychométriques de l'échelle	- Estimation de la fiabilité - Estimation de la validité de trait - Estimation de la validité prédictive

Nous revenons brièvement ci-dessous sur les spécificités de chacune de ces étapes.

➢ ***Etapes 1 et 2 : Vérifications préalables et spécification du modèle***

L'étape 1 consiste à vérifier que les données peuvent faire l'objet d'une analyse factorielle. Dans notre cas, les observations ayant trop de données manquantes ont été supprimées de l'analyse tandis que les valeurs manquantes « ponctuelles » ont été remplacées par la moyenne des observations pour le traitement correspondant[158]. De plus, l'échantillon est supérieur à 200 individus (environ 240 pour les expériences 1 et 2), ce qui constitue une taille suffisante pour mener à bien ce type d'analyse. Enfin, nous vérifions systématiquement la normalité des données par l'examen des indices d'asymétrie et d'aplatissement.

La deuxième étape a consisté, pour sa part, à spécifier le modèle à tester sous AMOS et à sélectionner les items retenus à l'issue de la réplication de l'analyse factorielle exploratoire.

➢ ***Etape 3 : Identification des contraintes***

L'étape 3 concerne principalement la fixation des contraintes pour que le modèle puisse « tourner » (le nombre de degrés de libertés doit ainsi être supérieur ou égal à 1). Dès lors, toute échelle disposant de 3 items (ou d'un nombre inférieur) ne peut pas bénéficier d'une telle analyse[159], sauf à pouvoir la valider conjointement avec une autre échelle du modèle. Nous ne pourrons donc mener une telle analyse confirmatoire que sur les échelles disposant d'un minimum de 4 items et sur les échelles de 3 items pouvant être analysées conjointement (à condition que le nombre de degrés de liberté soit supérieur ou égal à 1). Nous analyserons ainsi conjointement les trois échelles de perception (confidentialité, sensibilité et pertinence) d'une part et les échelles d'attitude et de préoccupation pour le RVP d'autre part, dans le cadre de l'expérience 1. Pour l'expérience 2, nous analyserons conjointement les échelles d'attitude et d'intention de poursuivre la relation.

Cette troisième étape est aussi l'occasion de choisir la méthode d'estimation des paramètres. Nous retiendrons la méthode du maximum de vraisemblance (M*aximum Likelihood* ou ML) qui est proposée par défaut par les logiciels. Celle-ci repose sur l'hypothèse de multi-normalité des variables, vérifiée par l'examen des indices d'asymétrie et d'aplatissement et par le calcul du coefficient de Mardia (coefficient de concentration multivarié). D'autres méthodes d'estimation peuvent s'affranchir de cette hypothèse contraignante, mais nécessitent alors un échantillon important (West, Finch et Curran 1995). Notre taille d'échantillon

[158] Plutôt que de calculer une moyenne globale, nous calculons une moyenne pour tous les individus affectés au traitement expérimental correspondant.

[159] Le nombre de degrés de liberté est alors égal à zéro et le calcul des indices d'ajustement ne peut pas s'effectuer.

correspond davantage à une condition de bonne application du maximum de vraisemblance. De plus, il s'avère que cette méthode d'estimation fournit de bons résultats, même quand l'hypothèse de multi-normalité est violée (Olsson et al. 2000). Ainsi, pour des items mesurés sur une échelle de Likert disposant d'au moins 4 échelons[160] et des coefficients d'asymétrie et d'aplatissement dans les limites acceptables, « il est tout à fait justifié de recourir à une estimation des paramètres par le maximum de vraisemblance » (Evrard Pras et Roux 2003).

➢ ***Etape 4 : Test du modèle***

L'étape 4 consiste à vérifier la qualité du modèle par le biais d'indices d'ajustement qui permettent d'évaluer dans quelle mesure le modèle théorique reproduit correctement les données collectées. Il existe trois sortes d'indices qui correspondent à des perspectives différentes de l'évaluation : les indices absolus, incrémentaux et de parcimonie.

Les indices disponibles étant nombreux, nous retiendrons ceux utilisés le plus souvent par les autres chercheurs. Notre choix s'est finalement porté sur les indices suivants : GFI ; AGFI ; RMSEA ; RMR (indices absolus) ; CFI (indice incrémental) et χ^2 / ddl (indice de parcimonie). Nous présentons ci-dessous les caractéristiques de ces indices et les valeurs critiques retenues.

Tableau 9.5 Les indices d'ajustement retenus : caractéristiques et valeurs critiques

Indices	Explications	Seuil critique
GFI (Goodness of Fit)	Mesure la part relative des variances et covariances présentes dans l'échantillon et prédites par la matrice théorique	> 0,9 (Sharma 1996)
AGFI (Adjusted Goodness of Fit)	Indice ajusté du GFI : ajuste cette prédiction selon les degrés de liberté	> 0,8 (Sharma 1996)
RMSEA (Root Mean Squared Error of Approximation)	Estime la manière dont le modèle se rapproche de la matrice de covariance de la population	< 0,08 (Sharma 1996)
RMR ou SRMR (Root Mean squared Residual)	Donne une estimation sur les résidus, c'est-à-dire sur les écarts entre les données et les résultats calculés	Proche de 0 (Jöreskog et Sobom 1984)
CFI (Comparative Fit Index)	Indicateur permettant de comparer un modèle exploratoire à un modèle nul en mesurant la diminution relative du manque d'ajustement	> 0,9 (Bentler, 1990)
χ^2 / ddl (chi-deux / degrés de liberté)	Indice permettant de mesurer le degré de parcimonie absolue d'un modèle mais aussi de distinguer, parmi plusieurs modèles alternatifs, lequel est le plus parcimonieux (Jöreskog, 1969)	Le plus bas possible

[160] Dans notre recherche, nous avons recours à des échelles à 5 ou 7 échelons.

➢ *Etape 5 : Vérification des qualités psychométriques des échelles*

L'étape 5 consiste à vérifier la qualité psychométrique des instruments de mesure.
Elle concerne principalement 4 éléments :

- l'estimation de la fiabilité
- l'estimation de la validité convergente (VC)
- l'estimation de la validité discriminante (VD)
- l'estimation de la validité prédictive (VP)

Nous pourrions aussi ajouter la validité de contenu qui s'assure de la correspondance entre les items et le construit qu'ils sont censés mesurer. Faute de critères statistiques, ce type de validité s'apprécie qualitativement. Dans notre cas, le fait que les items aient été soumis à un panel d'experts pour juger de leur représentativité doit renforcer l'adéquation des mesures aux construits qu'ils sont censés représenter. Une telle démarche est donc de nature à améliorer la validité de contenu des échelles développées dans le cadre de cette recherche.
Nous évoquerons successivement, dans les points suivants, les caractéristiques de chacun des autres éléments de qualité et les méthodes utilisées pour la démontrer.

La fiabilité de l'échelle de mesure

La fiabilité reflète dans quelle mesure les items composant l'échelle sont pertinents pour décrire le construit sous-jacent. Dans la phase exploratoire, nous mesurons cette fiabilité par l'Alpha de Cronbach. Cet indicateur présente toutefois un certain nombre de lacunes :
- il augmente avec le nombre d'indicateurs, ce qui incite à l'ajout d'indicateurs redondants
- il suppose que les indicateurs soient parfaitement corrélés avec le construit (Bollen 1989)

Par conséquent, l'alpha ne serait pas toujours un indicateur « fiable » de la fiabilité de l'échelle. Alternative intéressante, le Rhô de Jöreskog (1971) est de plus en plus utilisé en recherche. Il est en effet moins sensible au nombre d'indicateurs et est plus adapté aux méthodes d'équations structurelles car les termes d'erreur sont intégrés de manière explicite (Didellon et Valette-Florence 1996, Roussel et al. 2002). Dans le cadre de l'analyse confirmatoire, la fiabilité de l'échelle est donc généralement mesurée par le Rhô de Jöreskog, qu'on peut ensuite comparer à la valeur de l'alpha déduite de l'analyse exploratoire.

Notons qu'une forte cohérence interne entre les items montre que ces derniers mesurent la « même chose » et co-varient ensemble mais n'indique pas si ces items sont de « bonnes » mesures de la variable latente. Une échelle fiable ne serait donc pas nécessairement valide. La

question de la validité est traitée dans les paragraphes qui suivent. Deux types de validité sont examinés : la validité de trait (convergente et discriminante) et la validité prédictive.

La validité de trait (convergente et discriminante)

La validité de trait indique que l'opérationnalisation d'un concept mesure bien ce qu'elle est censée mesurer. Elle se décompose en validité convergente et validité discriminante (Zaltman Pinson et Angelmar 1973). La première correspond à la capacité d'une mesure à fournir des résultats proches de ceux d'autres mesures du même trait. La seconde est la capacité de la mesure à fournir des résultats différents de mesures d'autres traits.

Pour démontrer la validité convergente, deux méthodes peuvent être utilisées :

- le test-t associé à chacune des contributions factorielles : une valeur du t supérieure à 1,96 indique une bonne validité convergente
- la variance moyenne extraite (ρVC) : théoriquement, le pourcentage de la variance moyenne des items doit être supérieur ou égal à 0,5. Dans ce cas, en effet, les items sont mieux représentés par le construit que par les termes d'erreur.

Nous utiliserons surtout cette seconde approche, préconisée par Fornell et Larker (1981).

Pour tester la validité discriminante, plusieurs méthodes sont disponibles. Nous retiendrons, en priorité, l'approche préconisée par Fornell et Larker (1981), qui consiste à vérifier que la variance partagée entre les construits est inférieure à la variance moyenne partagée entre un construit et sa mesure. Il existe une seconde approche, recommandée par Roussel et al. (2002), qui consiste à comparer un modèle où les corrélations sont libres à un modèle où les corrélations entre les variables latentes sont fixées à 1. Nous y aurons recours en cas de doute.

La validité prédictive de l'échelle de mesure

Enfin, la validité prédictive concerne l'aptitude d'un instrument de mesure à prédire la valeur d'un autre concept. Il faut donc disposer d'un concept pouvant faire office de conséquence pour pouvoir vérifier la validité prédictive d'une mesure. On peut aussi l'estimer par rapport à une mesure directe et globale du concept sous-jacent, comme l'a fait Fleck-Dousteyssier (2006) dans sa thèse. Dans tous les cas, il s'agit de démontrer l'existence d'un lien (fort) entre le concept dont on a créé la mesure et le concept conséquence.

Pour rappel, dans le cadre de l'analyse confirmatoire, la méthode d'estimation des paramètres que nous avons retenue (*Maximum Likelihood*) requiert des distributions multi-normales, ce qui sera vérifié à travers le coefficient de Mardia (fourni par le logiciel AMOS 5.0). Or, si ce coefficient dépasse un certain seuil[161], l'estimation peut être sérieusement biaisée : cela a tendance à accroître la valeur du chi-deux et à décroître les valeurs des écarts-types. Il n'est donc pas certain que les paramètres estimés par cette méthode soient fiables. Un moyen de « contourner » ce problème est alors de recourir à un *bootstrap*[162] qui consiste à estimer à nouveau les paramètres, non pas sur la base d'une hypothèse de multi-normalité mais sur un ré-échantillonnage empirique fondé sur le remplacement de données. Cette méthode offre alors l'avantage d'obtenir des estimations des paramètres non (ou moins) biaisées.
Même en l'absence de problème de multi-normalité, recourir à un *bootstrap* peut être utile pour obtenir une estimation des erreurs standards des paramètres du modèle, en plus des estimations « classiques » fournies par la méthode du Maximum de Vraisemblance.

Pour toutes ces raisons, lors des analyses confirmatoires qui seront menées, nous réaliserons une première estimation des paramètres (et de l'ajustement du modèle) par la méthode du maximum de vraisemblance, avec examen du coefficient de multi-normalité. Nous procéderons ensuite à une seconde analyse, avec un *bootstrap* sur 2000[163] échantillons. Ceci nous fournira alors des informations qui viseront à évaluer la précision des estimations établies par la méthode du maximum de vraisemblance, notamment en cas de déviance par rapport à la multi-normalité. A noter que, dans les deux cas, les indices d'ajustement du modèle et les estimations des paramètres non standardisées sont identiques. Nous obtenons par contre de nouvelles estimations des écarts-types (colonne S.E.) qu'il est possible de comparer avec celles obtenues par le maximum de vraisemblance. Les résultats doivent alors être proches, le cas contraire signifiant un problème de normalité qu'il conviendra de corriger.

Pour ne pas alourdir les résultats qui sont déjà très riches, nous nous contenterons toutefois, dans les pages qui suivent, d'indiquer les principaux résultats obtenus et de signaler, le cas échéant, les problèmes identifiés ainsi que la manière dont nous les avons contournés.

---

[161] Il existe en réalité deux seuils, le premier traduisant une anormalité sans conséquence sur les estimations (anormalité modérée), le second correspondant à un véritable problème d'anormalité. Tous les auteurs ne sont pas d'accord sur ces « seuils », certains évoquant respectivement les valeurs de + ou – 2 puis + ou – 10, tandis que d'autres sont à + ou – 3 et + ou – 16.
[162] Nous préférons conserver le terme anglo-saxon, faute de terme français équivalent et suffisamment parlant.
[163] Nous choisissons ce chiffre de 2000, bien qu'il ait été démontré que l'accroissement de la fiabilité des résultats n'était que marginal au-delà de quelques centaines.

1.3 Typologie des construits de la recherche et type d'analyse conduite

L'objectif de ce paragraphe est de présenter l'ensemble des construits de la recherche nécessitant le développement d'une échelle de mesure et d'indiquer le type de démarche qui sera conduite pour chacun d'eux, eu égard aux propriétés psychométriques à vérifier.

L'examen du cadre conceptuel laisse apparaître 13 concepts à mesurer : les quatre variables de perception (confidentialité, sensibilité, pertinence, valeur), l'attitude face à la communication des données, la préoccupation pour le RVP, l'intention de répondre et de mentir, l'intention de poursuivre la relation, la réponse ainsi que le niveau d'implication, la qualité perçue de la relation et le comportement habituel de réponse[164] (variables de contrôle). Nous mettons tout de suite de côté la réponse (variable binaire, introduite dans l'expérience 3) qui sera mesurée par l'action réalisée par les participants (remplissage ou non du formulaire). Nous évacuons aussi le comportement habituel de réponse (introduit dans les expériences 2 et 3), mesuré par plusieurs items distincts, et qui sera présenté dans les chapitres correspondants.

L'ensemble des construits restants (11) peuvent être regroupés selon plusieurs critères :
- le fait d'être uni ou multidimensionnel (nombre de dimensions)
- le fait d'être mono ou multi-items (nombre d'items)
- leur statut dans le cadre de la recherche (l'expérience au cours de laquelle ils sont étudiés)
- le fait qu'il y ait reprise / adaptation d'une échelle existante ou développement ad hoc

Le tableau suivant indique l'appartenance des construits aux différents segments identifiés :

Tableau 9.6 Identification des concepts de la recherche selon plusieurs critères

Construits	Nombre de dimensions	Nombre d'items	Statut (N° Exp)	Reprise (R), Adaptation (A) ou construction (C)
Confidentialité	1	Plusieurs (Pl)	1	C
Sensibilité	1	Pl	1	C
Pertinence	1	Pl	1	C
Valeur	> 1	Pl	2	C
Attitude	1	Pl	1 et 2	C

[164] Nous n'évoquons pas ici les variables sociodémographiques et celles concernant l'expérience d'Internet qui sont mesurées classiquement sans que le développement d'une mesure ad hoc soit nécessaire.

Préoccupation	1 (2)[165]	Pl	1, 2 et 3	A
Int. de répondre	1	1	2	C
Int. de mentir	1	1	2	C
Int. de poursuivre la relation	1	Pl	2 et 3	A
Implication	1	Pl	1 et 2	R/A
Qualité	1	Pl	3	C

Les concepts mesurés par un seul item (intention de répondre et de mentir) ne sont pas concernés par la réalisation d'une analyse factorielle exploratoire puis confirmatoire. Les instruments utilisés seront donc présentés dans les chapitres correspondant aux résultats de l'expérience pour laquelle ils sont mesurés. De plus, seuls les instruments de mesure multi-items utilisés dans les expériences 1 et 2 seront présentés dans ce chapitre. Dans l'expérience 3 en effet, seules trois échelles de mesure multi-items seront utilisées : la préoccupation pour le respect de la vie privée (déjà développée pour les expériences 1 et 2), l'intention de poursuivre la relation (déjà été développée pour l'expérience 2) et la qualité perçue de la relation (variable de contrôle introduite uniquement dans l'expérience 3). Nous évoquerons donc la qualité psychométrique de cette dernière dans le chapitre présentant les résultats de cette expérience (cf. chapitre 12).

La démarche suivie pour valider les échelles de mesure adaptées et/ou développées pour les besoins de ce travail doctoral s'est faite en 2 temps. Nous nous sommes d'abord intéressés aux concepts utilisés pour l'expérience 1 puis à ceux nécessaires à l'expérience 2 (et n'ayant pas déjà été utilisés dans l'expérience 1). Nous avons à chaque fois procédé à deux collectes de données séparées, la première faisant l'objet d'une analyse exploratoire, la seconde de l'analyse confirmatoire. De plus, nous avons cherché à vérifier au maximum la validité de chaque instrument[166], tout en étant parfois limités par le nombre d'items dévolus à chacun ou par l'absence de construit à introduire conjointement (notamment pour l'analyse prédictive).
Le tableau ci-dessous reprend l'ensemble des construits pour lesquels un instrument de mesure a été adapté ou créé ainsi que les différentes analyses auxquelles nous avons pu procéder pour chacun d'eux.

[165] Ce construit, dont l'instrument de mesure est adapté d'une échelle existante à deux dimensions, n'aboutira finalement qu'à une dimension. Il sera présenté en détail dans la section suivante.
[166] La fiabilité des instruments a été, pour sa part, systématiquement vérifiée.

Tableau 9.7 Démarche adoptée pour chacun des construits de la recherche

<table>
<tr><th>Construits</th><th>N° Exp.</th><th>Collecte 1 (AFE)</th><th>Collecte 2 (AFC)</th><th>VC</th><th>VD</th><th>VP[167]</th></tr>
<tr><td colspan="7">Etape 1 : construits de l’expérience 1</td></tr>
<tr><td>Confidentialité</td><td>1</td><td rowspan="5">Pré-test 1
N = 308</td><td rowspan="5">Oui
N = 232</td><td>Oui</td><td>Oui</td><td>Oui</td></tr>
<tr><td>Sensibilité</td><td>1</td><td>Oui</td><td>Oui</td><td>Oui</td></tr>
<tr><td>Pertinence</td><td>1</td><td>Oui</td><td>Oui</td><td>Oui</td></tr>
<tr><td>Attitude</td><td>1 et 2</td><td>Oui</td><td>Oui</td><td>Oui</td></tr>
<tr><td>Préoccupation</td><td>1, 2 et 3</td><td>Oui</td><td>Oui</td><td>Oui</td></tr>
<tr><td>Implication</td><td>1 et 2</td><td>Pré-test 2
N = 90</td><td>Non</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="7">Etape 2 : construits de l’expérience 2 (non utilisés dans l’expérience 1)</td></tr>
<tr><td>Valeur</td><td>2</td><td rowspan="2">Pré-test 3
N = 93</td><td rowspan="2">Oui
N = 252</td><td>Oui</td><td>Oui</td><td>Oui</td></tr>
<tr><td>Int. de poursuivre la relation</td><td>2 et 3</td><td>Oui</td><td>Oui</td><td>Non</td></tr>
</table>

Comme nous le constatons à travers ce tableau, nous ne mènerons pas d’analyse confirmatoire pour le concept d’implication, parce qu’il est mesuré par 3 items et que la validité de l’échelle (reprise de la littérature) a déjà été démontrée. Nous ne vérifierons pas non plus la validité prédictive de l’intention de poursuivre la relation, par manque de concept « conséquence ».

Dans les paragraphes suivants, nous présentons les résultats du développement et/ou de l’adaptation des échelles de mesure nécessaires aux expériences 1 (section 2) et 2 (section 3).

2 Le développement des instruments de mesure nécessaires à l’expérience 1

Nous débutons la présentation par la spécification du domaine du construit et par les résultats de l’analyse exploratoire (2.1), pour chacun des concepts utilisés dans l’expérience 1. Nous abordons successivement les trois concepts de perception (confidentialité, sensibilité et pertinence), l’attitude face à la communication des données, la préoccupation pour le respect de la vie privée et enfin l’implication envers la catégorie de produit (variable de contrôle). Nous poursuivons par l’analyse confirmatoire (précédée d’une réplication de l’analyse exploratoire) suivie d’un examen de la fiabilité et de la validité des instruments obtenus (2.2).

[167] Abréviations utilisées respectivement pour validité convergente (VC), discriminante (VD) et prédictive (VP)

2.1 Spécification du domaine du construit et analyse exploratoire

2.1.1 Les échelles de confidentialité, sensibilité et pertinence

La confidentialité perçue des données sollicitées

Plusieurs auteurs insistent sur le fait qu'il ne faut pas confondre vie privée et confidentialité. Ainsi, si la vie privée correspond au « droit de l'individu de décider si et de quelle manière il est susceptible de dévoiler aux autres ses pensées, opinions, sentiments et faits personnels » (Mayer 2002), la confidentialité fait plutôt référence aux restrictions imposées quant à la manière dont les informations dévoilées pourront être transmises par la suite. Plus précisément, il y a confidentialité si toute transmission de données visant à identifier l'individu ou susceptible de lui causer du tort est interdite et/ou soumise à autorisation préalable (Mayer 2002). Par conséquent, la confidentialité perçue peut être définie comme le fait, pour un individu, d'estimer que les données dévoilées ne tomberont pas entre les mains de tiers mal intentionnés et/ou qu'elles ne seront utilisées qu'après accord de l'intéressé.

Dans la mesure où nous n'avons trouvé, au sein de la littérature, aucune échelle satisfaisante capable de mesurer ce construit, nous avons décidé de développer nous-mêmes les items, en nous inspirant des verbatims recueillis au cours de l'étude qualitative préalable. Un item intéressant issu d'une échelle développée par Miyazaki et Krishnamurthy (2002) pour mesurer la « favorabilité perçue des pratiques des entreprises en matière de traitement de données » est cependant repris tel quel (après traduction et rétro-traduction). Après évaluation par les experts, les 4 items retenus pour former l'échelle de confidentialité (format Likert en 7 points) sont présentés dans le tableau suivant.

Tableau 9.8 Echelle de mesure initiale de la confidentialité perçue (CONFID)

Items	Développement
CONFID 1 : Je crains que cette entreprise n'utilise mes données sans mon accord préalable (item inversé)	Créé
CONFID 2 : Je pense que l'entreprise respectera la confidentialité de mes données personnelles	Créé
CONFID 3 : Je crois que l'entreprise considère la confidentialité comme quelque chose d'important	Adapté de la littérature
CONFID 4 : J'ai peur que l'entreprise communique ces données à des partenaires ou à des tiers	Créé

Avant de mener l'analyse factorielle, nous procédons à une analyse descriptive univariée sur ces 4 items. Les résultats montrent que l'item CONFID 4 souffre d'un manque de variance, qui nous conduit à le supprimer. L'analyse factorielle exploratoire menée sur les 3 items restants donne les résultats suivants (KMO = 0,668 ; sig. du test de Bartlett = 0,000) :

Tableau 9.9 Résultat de l'analyse factorielle exploratoire sur l'échelle de confidentialité

Items	Qualité de la représentation	Saturation (sur l'axe 1)
CONFID 2 : Je pense que l'entreprise respectera la confidentialité de mes données personnelles	0,717	0,847
CONFID 3 : Je crois que l'entreprise considère la confidentialité comme quelque chose d'important	0,668	0,817
CONFID 1 : Je crains que cette entreprise n'utilise mes données sans mon accord préalable (inversé)	0,579	0,761
% de variance expliquée	65,49%	
Alpha de Cronbach	0,74	

La qualité de représentation des items, tout comme leur saturation, sont satisfaisantes (> 0,5). L'alpha est également convenable (0,74), surtout pour une échelle en construction.

La sensibilité perçue des données sollicitées

Dans le langage courant, on parle de point sensible pour désigner un endroit « que le moindre contact rend douloureux ou fait souffrir ». On dit aussi que quelque chose est sensible quand il requiert « une attention et des précautions particulières, à cause des réactions possibles » (Dictionnaire Le Petit Robert). Dans la littérature sur le respect de la vie privée, le concept de sensibilité est fréquemment employé, sans qu'il n'existe pour autant une définition précise de ce terme. Weible (1993) définit la sensibilité d'une donnée comme « le niveau de préoccupation qu'un individu peut ressentir pour un type de donnée dans une situation particulière ». Gandy (1993) indique que certaines personnes considèrent comme sensible toute donnée qui, si elle était effacée ou partagée, pourrait leur causer du tort. Toutefois, il explique aussi qu'il est souvent difficile de différencier ce qui est blessant de ce qui est simplement ennuyeux. L'idée de définir une donnée sensible en référence aux conséquences qu'elle est susceptible de causer semble donc délicate à mettre en œuvre. Finalement, si, comme le souligne Hill (1995), des tentatives ont essayé de différencier l'information sensible de celle qui ne l'est pas, un consensus n'a toujours pas été trouvé. En réalité, la majorité des auteurs évoquent la sensibilité d'une donnée en référence à la catégorie à laquelle elle

appartient (i.e. sociodémographique, financier, …). Ils considèrent ainsi qu'une donnée financière est forcément plus sensible qu'une donnée sociodémographique. Cette idée se retrouve d'ailleurs dans la définition « juridique » du terme, la loi dite informatique et libertés qualifiant de sensibles (et interdisant alors toute collecte ou traitement à leur égard) « les données à caractère racial ou ethnique, les opinions politiques, philosophiques ou religieuses, etc …) ». En nous basant sur les entretiens semi-directifs menés auprès de consommateurs (internautes) français, il apparaît toutefois que l'idée de segmenter les données de cette façon serait finalement peu opportune : la sensibilité serait en effet à la fois individuelle (propre à chaque personne) et situationnelle (susceptible d'évoluer selon le contexte). De plus, les personnes interrogées semblent faire davantage référence au fait de « pouvoir garder de telles données pour soi », idée qui ressort de la majorité des entretiens. En référence à cette dernière, le concept de sensibilité perçue des données sollicitées peut ainsi se définir comme le fait, pour un individu, de considérer les données demandées comme des données intimes, qu'il ne souhaite révéler qu'à un nombre limité d'interlocuteurs. La sensibilité se différencie donc bien de la pertinence (concept défini dans le paragraphe suivant), dans la mesure où un individu peut considérer une donnée comme intime mais trouver normal qu'on la lui demande à certaines occasions (par exemple : le numéro de carte de crédit lors d'un achat en ligne ou la taille et le poids sur un site spécialisé dans les régimes).

Nous n'avons pas trouvé, au sein de la littérature, d'échelle susceptible de mesurer efficacement ce concept. La plupart des auteurs l'évaluent en effet en référence à la fréquence avec laquelle l'individu accepte de fournir les données à différents interlocuteurs, que ce soit dans le domaine interpersonnel (parents, familles, amis ou étrangers) ou dans un cadre plus commercial (entreprises). Nous avons donc développé notre propre échelle, en nous basant sur les verbatims issus de l'étude qualitative.

Après évaluation par les experts, les 4 items retenus pour former l'échelle de sensibilité (format Likert en 7 points) sont présentés dans le tableau suivant.

Tableau 9.10 Echelle de mesure initiale de la sensibilité perçue (SENS)

Items	Développement
SENS 1 : Globalement, les données demandées sont trop personnelles	Créé
SENS 2 : Je considère que les informations qui me sont demandées sont privées	Créé
SENS 3 : certaines de ces informations, je les réserve à mes proches	Créé
SENS 4 : l'entreprise n'a pas à connaître certaines de ces informations	Créé

Avant de mener l'analyse factorielle, nous procédons à une analyse descriptive univariée sur ces 4 items. Les résultats montrent que l'item SENS 4 n'a pas une distribution normale (coefficient d'asymétrie = - 1,326), ce qui nous conduit à le supprimer. L'analyse factorielle exploratoire, menée sur les 3 items restants, donne les résultats suivants (KMO = 0,651 ; sig. du test de Bartlett = 0,000) :

Tableau 9.11 Résultat de l'analyse factorielle exploratoire sur l'échelle de sensibilité

Items	Qualité de la représentation	Saturation (sur l'axe 1)
SENS 1 : Globalement, les données demandées sont trop personnelles	0,809	0,903
SENS 2 : Je considère que les informations qui me sont demandées sont privées	0,805	0,900
SENS 3 : certaines de ces informations, je les réserve à mes proches	0,555	0,745
% de variance expliquée	72,65%	
Alpha de Cronbach	0,80	

La qualité de représentation des items, tout comme leur saturation, sont satisfaisantes (> 0,5). L'alpha est également satisfaisant (0,80), surtout pour une échelle en construction.

La pertinence perçue des données sollicitées

Le concept de pertinence a été utilisé dans d'autres domaines de recherche (le parrainage ou l'extension de marque par exemple) pour désigner l'idée de cohérence ou de « *fit* » entre deux notions ou entités. Aaker et Keller (1990) parlent ainsi de pertinence entre une marque mère et un nouveau produit tandis que McDonald (1991) évoque la pertinence entre le parrain et l'entité parrainée. Couramment, on emploie ce terme pour désigner « ce qui est exactement adapté à l'objet dont il s'agit » ou encore « ce qui dénote du bon sens, de la justesse d'esprit ». En matière de sollicitation de données personnelles, si l'idée de pertinence des données sollicitées a déjà été évoquée par certains chercheurs, le concept n'a jamais été réellement défini, ni même mesuré. D'après ce qui ressort des analyses de notre étude qualitative préalable, il apparaît que les consommateurs sont particulièrement soucieux de la légitimité de l'entreprise à demander chacune des informations se trouvant sur le formulaire à compléter. Ainsi, alors qu'ils trouvent normal qu'on demande leur nom et leur adresse, la plupart sont plus sceptiques quand il s'agit de recueillir le numéro de téléphone ou la date de naissance. Toute information demandée incite donc le consommateur à se poser la question de la finalité de sa présence sur le formulaire et à être particulièrement attentif à tout ce qui serait inutile.

Le concept de pertinence perçue des données sollicitées peut ainsi se définir comme le fait, pour un individu, d'estimer que (toutes) les données demandées sont nécessaires à la réalisation de l'échange.

Comme pour les concepts précédents, nous n'avons trouvé, dans la littérature, aucune échelle visant à mesurer le concept de pertinence perçue des données sollicitées. Nous avons donc développé nos propres items, sur la base des verbatims recueillis lors de l'analyse qualitative préalable. Comme dans le cas de la confidentialité perçue, un item intéressant issu d'une échelle développée par Fusilier et Hoyer (1980) pour mesurer la perception d'invasion de la vie privée est repris et adapté au contexte présent. Après évaluation par les experts, les 4 items retenus pour former l'échelle de pertinence (format Likert en 7 points) sont présentés dans le tableau suivant.

Tableau 9.12 Echelle de mesure initiale de la pertinence perçue (PERT)

Items	Développement
PERT 1 : Fournir ces données à cette entreprise me semble justifié	Créé
PERT 2 : Il me semble normal que l'entreprise me demande toutes ces informations	Créé
PERT 3 : La collecte de ces données ne me semble pas nécessaire (item inversé)	Adapté de la littérature
PERT 4 : Il n'est pas nécessaire que l'entreprise connaisse toutes ces informations	Créé

Avant de mener l'analyse factorielle, nous procédons à une analyse descriptive univariée sur ces 4 items. Les résultats montrent que l'item PERT 4 souffre d'un manque de variance. Introduit au sein d'une première analyse factorielle, il est aussi le moins bien représenté des 4 items, ce qui nous conduit à le supprimer. Une nouvelle analyse factorielle menée sur les 3 items restants donne les résultats suivants (KMO = 0,772 ; sig. du test de Bartlett = 0,000) :

Tableau 9.13 Résultat de l'analyse factorielle exploratoire sur l'échelle de pertinence

Items	Qualité de la représentation	Saturation (sur l'axe 1)
PERT 1 : Fournir ces données à cette entreprise me semble justifié	0,726	0,852
PERT 3 : La collecte de ces données ne me semble pas nécessaire (item inversé)	0,629	0,793
PERT 2 : Il me semble normal que l'entreprise me demande toutes ces informations	0,762	0,873
% de variance expliquée	70,53%	
Alpha de Cronbach	0,79	

La qualité de représentation des items tout comme leur saturation est satisfaisante (> à 0,5). L'alpha est également satisfaisant (0,79), surtout pour une échelle en construction.

2.1.2 L'échelle d'attitude face à la communication des données

L'attitude est une variable complexe puisqu'elle est considérée comme multidimensionnelle par certains chercheurs et unidimensionnelle par d'autres. Il n'y a donc pas d'unanimité quant à la dimensionnalité de ce concept et finalement deux approches coexistent (Bagozzi 1989, Darpy et Volle 2003). La première considère que l'attitude comporte trois composantes : cognitive (les croyances), affective (un jugement affectif global) et conative (prédispositions comportementales). La seconde considère l'attitude comme un concept unidimensionnel et se focalise sur la composante affective, les composantes cognitives et conatives étant respectivement considérées comme des antécédents et des conséquences (Lutz 1991). Dans la mesure où notre modèle conceptuel est processuel, l'attitude jouant le rôle de médiateur entre une évaluation d'une part et des intentions comportementales d'autre part, nous nous situons clairement dans cette seconde approche et considérons alors l'attitude comme un construit unidimensionnel, essentiellement basé sur l'affect.

Fishbein et Ajzen (1975) soulignent que l'attitude est la réaction ou l'évaluation de l'individu à l'égard d'un objet sur un continuum favorable/défavorable ou aime/déteste. Pour d'autres théoriciens, l'attitude représente une anticipation ou une prédisposition à ajuster son comportement à une situation spécifique ou bien encore une réponse conditionnée à un stimulus (LaPiere 1934). Dans le domaine de la sollicitation de données personnelles par le biais d'un formulaire à remplir, l'attitude de l'individu face à la communication de ses données peut donc se définir comme le degré d'agrément relatif à la divulgation de ces informations.

Il n'existait, au sein de la littérature, aucune échelle mesurant l'attitude face à la communication des données, la majorité des recherches évaluant directement le comportement. Il existe par contre des mesures de l'attitude dans de nombreux autres domaines. En s'inspirant de ces dernières et des verbatims issus de notre étude qualitative, nous avons ainsi créé plusieurs items soumis à l'évaluation des experts. Les 4 items retenus pour former l'échelle d'attitude face à la communication des données (format Likert en 7 points) sont présentés dans le tableau suivant.

Tableau 9.14 Echelle de mesure initiale de l'attitude (ATT)

Items	Développement
ATT 1 : Cela ne me pose aucun problème de donner ces informations	Créé
ATT 2 : Je me sens mal à l'aise à l'idée de remplir ce formulaire (inversé)	Créé
ATT 3 : Je suis favorable à l'idée de répondre à ces questions	Créé
ATT 4 : Devoir fournir ces données ne me plaît pas particulièrement (inversé)	Créé

Avant de mener l'analyse factorielle, nous procédons à une analyse descriptive univariée sur ces 4 items. Les résultats montrent que l'item ATT 4 souffre d'un manque de variance (coefficient d'asymétrie = 0,936). Introduit au sein d'une première analyse factorielle, il est aussi le moins bien représenté des 4 items, ce qui nous conduit à le supprimer. Une nouvelle analyse factorielle menée sur les 3 items restants donne les résultats suivants (KMO = 0, 723 ; sig. du test de Bartlett = 0,000) :

Tableau 9.15 Résultat de l'analyse factorielle exploratoire sur l'échelle d'attitude

Items	Qualité de la représentation	Saturation (sur l'axe 1)
ATT 1 : Cela ne me pose aucun problème de donner ces informations	0,751	0,867
ATT 2 : Je me sens mal à l'aise à l'idée de remplir ce formulaire (inversé)	0,751	0,866
ATT 3 : Je suis favorable à l'idée de répondre à ces questions	0,809	0,899
% de variance expliquée	77,02%	
Alpha de Cronbach	0,85	

La qualité de représentation des items, tout comme leur saturation, sont satisfaisantes (> 0,5). L'alpha est très satisfaisant (0,85), sans qu'il n'y ait pour autant redondance entre les items.

2.1.3 L'échelle de préoccupation pour le respect de la vie privée

Dans un précédent chapitre, nous avons déjà défini la préoccupation pour le respect de la vie privée (on devrait ajouter dans son aspect informationnel) comme le niveau de préoccupation qu'un individu pouvait ressentir face aux menaces qui pèsent sur lui quand les entreprises sollicitent et/ou utilisent des données personnelles le concernant. S'agissant d'un concept central au sein de la littérature, plusieurs échelles visant à le mesurer ont déjà été développées. Si certains chercheurs (notamment Smith et al. 1996 et Stewart et Segars 2002) considèrent ce

construit comme multidimensionnel, en référence à l'ensemble des pratiques susceptibles de constituer une menace (i.e. collecte, utilisation secondaire, accès aux données, …) d'autres auteurs préfèrent mesurer une « préoccupation générale », ce qui a l'immense avantage de limiter le nombre d'items dévolus à l'échelle (pour exemple 15 items sont nécessaires pour l'échelle de Smith et al. 1996). Nous nous situons clairement dans cette seconde perspective, notre objectif étant de simplifier au maximum les mesures de manière à ne pas saturer le répondant et à éviter ainsi tout biais dans les réponses et tout risque de mortalité expérimentale. Plusieurs échelles correspondant à cet objectif ont été trouvées dans la littérature. Trois d'entre elles ont été retenues, pour leur correspondance avec la définition que nous avons donnée de ce concept (présentée ci-dessus), pour leurs qualités psychométriques et enfin pour la faiblesse du nombre d'items (4 à 6 maximum). L'échelle en 5 items proposée par Culnan (1993) a cependant rapidement été écartée de par la longueur des items auxquels on aboutissait, après traduction de l'anglais vers le français. Pour exemple, un des items était formulé de la manière suivante : « les entreprises ne devraient pas utiliser les données personnelles des consommateurs dans un autre but que ce pour quoi elles ont été collectées ». De plus, la formulation indirecte des items ne nous semblait pas de nature à différencier réellement le niveau de préoccupation des individus. L'échelle en 4 items de Farag et Krishnan (2003) a également été écartée par crainte d'un manque de variance dans les réponses et d'un certain effet de halo. Les items étaient en effet formulés de la manière suivante : « il est important que je sois informé du type d'informations … / que l'entreprise m'informe de la durée … / que je sache pourquoi l'entreprise veut collecter … / que je sache si l'entreprise est susceptible d'utiliser les données … ». La troisième et dernière échelle proposée par Dinev et Hart (2003) a été retenue (en réalité adaptée) car elle a l'avantage de correspondre parfaitement à la définition proposée. Elle présente cependant deux inconvénients (auxquels nous nous proposons de remédier) : d'une part, comme l'échelle précédente, la formulation des items nous fait craindre un effet de halo ; d'autre part, elle est bidimensionnelle (deux dimensions à 3 items chacune). La première dimension concerne l'aspect « disponibilité des informations » (*PCI : Privacy Concern for Information availability*). Elle est formulée de la manière suivante : « je suis préoccupé(e) par le fait qu'une personne puisse trouver des informations concernant … mes date et lieu de naissance et le nom de mes parents (item 1) ; mon adresse et mon numéro de téléphone personnel et professionnel (item 2), l'adresse et le numéro de téléphone de mes résidences actuelles et précédentes (item 3) ». Cet aspect nous semble davantage lié à la sensibilité des données qu'à la préoccupation pour le respect de la vie privée. De plus, nous estimons qu'il existe une certaine redondance entre les items, surtout entre les deux derniers. Pour « gommer » cet

effet, nous décidons donc de regrouper l'ensemble de ces 3 items en un seul, formulé de la manière suivante : « je suis inquiet à l'idée que des entreprises puissent détenir des données que je considère comme privées ». Nous nous inspirons également des 3 items issus de la seconde dimension - nommée « utilisation abusive » - (*PCA : Privacy Concern for information Abuse*) pour formuler les 3 items suivants. Nous nous permettons cependant de modifier certains d'entre eux pour être plus facilement accessibles à un public de « non-initiés ». Nous modifions également le début des items et tournons l'un d'eux en sens inverse (item inversé) pour réduire tout risque d'effet de halo. Enfin, nous supprimons l'aspect « internet » car nous souhaitons mesurer une préoccupation « générale », sans indication du type de média concerné. Cette « nouvelle » échelle en 4 items est ensuite soumise à évaluation auprès des experts[168] puis pré-testée auprès de consommateurs pour s'assurer de la compréhension des items. Plusieurs améliorations ont dû être apportées avant d'aboutir à l'échelle finale. L'échelle d'origine et l'échelle retenue après transformation et « pré-validation » sont présentées dans le tableau suivant :

Tableau 9.16 Echelle de mesure de la préoccupation pour le respect de la vie privée (PREOC)

Echelle initiale en 6 items (Dinev et Hart, 2003)	Echelle après transformation (4 items)
I am concerned that a person can find the following information about : … my date and place of birth and the name of my parents (PCI 1) … addresses and telephones of my home/workplace (PCI 2) … address and telephones of my current and previous residence (PCI 3)	PREOC 1 : Je suis inquiet(e) à l'idée que des entreprises puissent détenir des données que je considère comme privées PREOC 2 : L'idée que mes données personnelles puissent être utilisées sans ma permission préalable ne m'inquiète pas particulièrement (inversé)
PCA 1 : I am concerned that the information I submit (on the internet) could be misused PCA 2 : I am concerned about submitting information on the Internet because it could be used in a way I did not foresee PCA 3 : I am concerned about submitting information (on the Internet) because of what others might do with it	PREOC 3 : Je suis préoccupé(e) par le fait que les informations que je donne aux entreprises puissent être utilisées à mon insu PREOC 4 : Cela m'ennuie de devoir divulguer des données personnelles aux entreprises

Avant de mener l'analyse factorielle, nous procédons à une analyse descriptive univariée sur ces 4 items. Les résultats montrent que l'item PREOC 4 souffre d'un manque de variance. Introduit au sein d'une première analyse factorielle, il est aussi le moins bien représenté des 4

[168] Tous les items seront jugés pertinents par les experts mais nécessitant certaines modifications

items, ce qui nous conduit à le supprimer. Une nouvelle analyse factorielle menée sur les 3 items restants donne les résultats suivants (KMO = 0, 729 ; sig. du test de Bartlett = 0,000) :

Tableau 9.17 Résultat de l'analyse factorielle exploratoire sur l'échelle de préoccupation

Items	Qualité de la représentation	Saturation (sur l'axe 1)
PREOC 1 : Je suis inquiet(e) à l'idée que des entreprises puissent détenir des données que je considère comme privées	0,781	0,884
PREOC 2 : L'idée que mes données personnelles puissent être utilisées sans ma permission préalable ne m'inquiète pas particulièrement (inversé)	0,742	0,862
PREOC 3 : Je suis préoccupé(e) par le fait que les informations que je donne aux entreprises puissent être utilisées à mon insu	0,792	0,890
% de variance expliquée	77,20%	
Alpha de Cronbach	0,85	

La qualité de représentation des items tout comme leur saturation est satisfaisante (> à 0,5). L'alpha est très satisfaisant (0,85), sans qu'il y ait pour autant redondance entre les items.

2.1.4 L'échelle d'implication envers la catégorie de produits

Valette-Florence (1989, p 58) rapporte une définition de l'implication proposée par plusieurs chercheurs : « l'implication est un état non observable de motivation, d'excitation ou d'intérêt [..] créé par des variables externes (la situation, le produit ..) et internes (le moi, les valeurs). Elle entraîne [..] certaines formes de recherche de produit, de traitement de l'information et de prise de décision ». L'implication d'un individu dans une catégorie de produits a aussi été définie par Strazzieri (1994) comme « une disposition durable d'un individu à allouer de l'énergie à tout ce qui se rapporte à l'objet, due au potentiel de stimulation de cet objet, associé à sa pertinence personnelle pour l'individu »[169].

Les mesures d'implication les plus couramment utilisées en France sont l'échelle des profils d'implication (Laurent et Kapferer 1986), le *Personal Involvement Inventory* (Zaichowsky 1985) et l'échelle de Pertinence-Intérêt-Attirance (Strazzieri 1994). Nous utiliserons cette dernière car il s'agit d'une mesure unidimensionnelle qui bénéficie à la fois d'une très bonne cohérence interne et d'une forte valeur prédictive (Christau et Strazzieri 1996, Rieunier 2000).

[169] Traduit de l'anglais par Cristau et Strazzieri (1996)

Par ailleurs, elle possède un nombre d'items beaucoup moins important que les deux autres (6 contre respectivement 16 et 20). L'échelle originale de Pertinence-Intérêt-Attirance (PIA) comporte deux énoncés relatifs à chaque dimension (cf. tableau 9.18). Afin d'alléger au maximum le questionnaire, nous avons cependant choisi de recourir à la procédure suivie par Volle (1996) et Rieunier (2000) qui consiste à ne conserver qu'un item par dimension. Rieunier (2000) rapporte aussi les coefficients d'association observés par plusieurs chercheurs sur cette échelle. Les 3 items retenus sont alors ceux qui montrent les plus forts coefficients d'association. L'échelle d'implication (IMPL) finalement utilisée est présentée dans le tableau 9.18. A noter que nous utilisons le terme de « sujet » au lieu des termes classiques d'« activité » ou de « produit » car nous cherchons à mesurer l'implication du répondant à l'égard de la catégorie « téléphonie mobile », secteur d'activité choisi comme application dans le cadre des expériences 1 et 2[170].

Tableau 9.18 L'échelle d'implication (PIA) originale (Strazzieri 1994) et l'échelle simplifiée

Echelle PIA originale (6 items)		Echelle PIA simplifiée (3 items)	
Dimensions	Items originaux	Items conservés	Dimensions
Pertinence 1 Pertinence 2	C'est un(e) .. qui compte beaucoup pour moi C'est un(e) .. auquel j'accorde une importance particulière	C'est [un sujet] auquel j'accorde une importance particulière	Pertinence 2 (IMPL 1)
Intérêt 1 Intérêt 2	J'aime particulièrement parler de .. On peut dire que c'est un(e) .. qui m'intéresse	On peut dire que c'est [un sujet] qui m'intéresse	Intérêt 2 (IMPL 2)
Attirance 1 Attirance 2	Je me sens particulièrement attiré par .. Le seul fait de me renseigner sur .. est un plaisir	Je me sens particulièrement attiré(e) par [tout ce qui a trait à la téléphonie mobile]	Attirance 1 (IMPL 3)

Cette échelle écourtée est pré-testée auprès de 90 étudiants de Master à l'Université de Paris Dauphine. L'analyse factorielle de l'échelle d'implication (PIA, que nous appelons pour notre part IMPL) donne les résultats suivants (KMO = 0, 682 ; sig. du test de Bartlett = 0,000) :

Tableau 9.19 Résultat du pré-test de l'échelle PIA (IMPL) sur la « téléphonie mobile »

Items	Qualité de la représentation	Saturation (sur l'axe 1)
IMPL 1 : C'est [un sujet] auquel j'accorde une importance particulière	0,614	0,784
IMPL 2 : On peut dire que c'est [un sujet] qui m'intéresse	0,671	0,819

[170] La justification de ce choix sera abordée dans les chapitres correspondants (chapitres 10 et 11).

IMPL 3 : Je me sens particulièrement attiré(e) par [tout ce qui a trait à la téléphonie mobile]	0,700	0,837
% de variance expliquée	66,18%	
Alpha de Cronbach	0,74	

La validité interne et la fiabilité de cette échelle sont donc globalement satisfaisantes, même si l'alpha est assez bas, compte tenu du fait qu'il s'agit d'une échelle ayant démontré de bonnes qualités psychométriques dans d'autres recherches. Le fait qu'elle ait été administrée en fin de questionnaire pourrait expliquer ce résultat, qui reste cependant acceptable.
Cette échelle étant limitée à 3 items, elle ne fera pas l'objet d'une analyse confirmatoire.

Après avoir mené cette analyse exploratoire sur l'ensemble des échelles correspondant aux concepts utilisés pour l'expérience 1, nous donnons maintenant les résultats de l'analyse confirmatoire réalisée sur un nouvel échantillon de 232 individus.

2.2 L'analyse confirmatoire

La présentation des résultats de l'analyse confirmatoire se déroule en deux temps. Nous commençons par les résultats issus de la réplication de l'analyse exploratoire sur le nouvel échantillon. Nous indiquons ensuite les résultats des modèles de mesure estimés par le logiciel AMOS 5.0, avant d'aborder la question de la validité et de la fiabilité. Nous rappelons que cette analyse confirmatoire concerne l'ensemble des concepts utilisés pour l'expérience 1 (confidentialité, sensibilité, pertinence, attitude et préoccupation), hormis l'implication pour laquelle nous nous contentons de réaliser une réplication de l'analyse factorielle exploratoire.

2.2.1 La réplication de l'analyse exploratoire

Les résultats de la réplication de l'analyse exploratoire sont présentés brièvement ci-après et résumés dans le tableau ci-dessous (cf. tableau 9.20). L'objectif est simplement de vérifier la structure factorielle des échelles, avant de les introduire sous AMOS.
Globalement les résultats de cette nouvelle analyse factorielle menée sur les 6 concepts concernés (confidentialité, sensibilité, pertinence, attitude, préoccupation et implication) confirment l'unidimensionnalité de chaque concept, la bonne qualité de représentation des items (hormis l'item SENS 3 qui est alors supprimé) et conduit à accepter provisoirement la fiabilité de chaque échelle (alphas de Cronbach compris entre 0,70 et 0,87).

Tableau 9.20 Résultat de la réplication de l'analyse factorielle exploratoire

Concept	Items (après expl.)	Items (avant AMOS)	KMO / Test de Bartlett	% de variance	Qualité de la représentation Extraction min - max	α de Cronbach (avant AMOS)
Confidentialité (CONFID)	3	3	0,673 / 0,000	69,9%	0,60 – 0,78	0,77
Sensibilité (SENS)	3	2	0,609 / 0,000	63,8%	0,42[171] – 0,75	0,70 / 0,78[172]
Pertinence (PERT)	3	3	0,722 / 0,000	74,8%	0,73 – 0,77	0,83
Attitude face à la communication (ATT)	3	3	0,653 / 0,000	73,0%	0,59 - 0,83	0,81
Préoccupation RVP (PREOC)	3	3	0,736 / 0,000	79,4%	0,77 – 0,81	0,87
Implication (IMPL)	3	3	0,704 / 0,000	73,5%	0,67 – 0,77	0,82

Une analyse factorielle multivariée (KMO = 0, 849 ; sig. du test de Bartlett = 0,000) menée sur les trois variables de perception introduites simultanément (soit 8 items au total) laisse apparaître quelques problèmes possibles (cf. tableau 9.21 ci-dessous). En effet, malgré une qualité de représentation et une saturation systématiquement supérieure à 0,5 pour l'ensemble des items ; on constate que les items CONFID 1 (inversé) et PERT 2, tout en étant corrélés avec le facteur correspondant, saturent aussi sur d'autres axes, la différence de contribution entre l'axe principal et l'autre axe étant tout juste égale à 0,3 pour PERT 2 (0,717 contre 0,410) et légèrement inférieure à 0,3 pour CONFID 1 (0,616 contre 0,340).

Tableau 9.21 Résultat de l'analyse factorielle multivariée sur les trois variables de perception

Items	Facteur 1	Facteur 2	Facteur 3	Qualité de représentation
PERT1	**0,873**			0,815
PERT3_i	**0,740**		0,371 (*)	0,749
PERT2	**0,717**		-0,410	0,727
CONFID2		**0,890**		0,826
CONFID3		**0,847**		0,778
CONFID1_i	0,313	**0,616**	-0,340	0,593
SENS1			**0,851**	0,822
SENS2			**0,844**	0,801

(*) Coefficients inférieurs à 0,3 omis

Nous conservons néanmoins, dans un premier temps, l'ensemble des 8 items correspondant aux trois variables de perception, nous réservant le droit de les éliminer du modèle de mesure défini sous AMOS, en cas de mauvaise estimation.

[171] Cette qualité de représentation correspond à l'item SENS 3 qui sera supprimé.
[172] Le premier alpha est calculé sur les 3 items, le second sur les 2 items restants. Nous donnons ce résultat pour information, dans la mesure où seuls 2 items composent l'échelle.

2.2.2 L'analyse confirmatoire et la qualité des échelles

L'analyse confirmatoire proprement dite est réalisée grâce à l'estimation de différents modèles de mesure sous AMOS 5.0. Comme toutes nos échelles sont en 3 items (voire en 2 items pour la sensibilité), nous ne pouvons estimer la qualité de chacune d'elle séparément. Nous décidons donc de procéder en 2 temps. Nous estimons un premier modèle composé des trois variables de perception, puis, dans un second modèle, nous introduisons conjointement l'attitude et la préoccupation pour le RVP. Nous précisons que les vérifications préalables à l'estimation des modèles, la spécification des modèles et l'identification des contraintes ont été faites en suivant le descriptif présenté au point 1.2.3 de ce chapitre. Nous résumons brièvement ces opérations dans le tableau 9.22 ci-dessous.

Tableau 9.22 Etapes préalables au test du modèle de mesure sous AMOS

	Modèle 1	Modèle 2
Vérifications préalables	- élimination ou remplacement des valeurs manquantes - vérification de la taille de l'échantillon : 232 individus - vérification de la normalité des données (cf. tableau 9.23)	
Spécification du modèle	3 variables de perception (confidentialité, sensibilité et pertinence, soit 8 items)	attitude et préoccupation (soit 6 items)
Identification des contraintes	- choix de la méthode d'estimation : maximum de vraisemblance - aucune contrainte particulière (hormis les contraintes « classiques »)	

Le tableau 9.23 présente les résultats visant à vérifier la normalité des données. Il montre qu'on peut considérer les données comme normales, les coefficients d'asymétrie et d'aplatissement étant respectivement inférieurs aux seuils de 1 et 1,5. Nous rappelons que nous aurons recours au maximum de vraisemblance pour estimer le modèle[173].

Tableau 9.23 Vérification de la normalité des données

Concepts et items		Coefficient d'asymétrie	Coefficient d'aplatissement
CONFID	CONFID 1	- 0,495	- 0,908
	CONFID 2	- 0,076	- 0,850
	CONFID 3	0,098	- 0,362

[173] Nous renvoyons les lecteurs au paragraphe 1.2.3 pour la justification du choix de cette méthode.

SENS	SENS 1 SENS 2	- 0,081 - 0,517	- 0,996 - 0,781
PERT	PERT 1 PERT 2 PERT 3	0,145 0,510 - 0,236	- 0,757 - 0,625 - 0,926
ATT	ATT 1 ATT 2 ATT 3	0,346 0,013 0,240	- 1,081 - 1,157 - 0,661
PREOC	PREOC 1 PREOC 2 PREOC 3	- 0,560 0,695 - 0,484	- 0,470 - 0,613 - 0,628

Modèle 1 : les trois variables de perception (confidentialité, sensibilité et pertinence).

Un premier modèle comprenant les 3 variables de perception (soit les 8 items conservés) est testé sous AMOS. Ce modèle ne présente pas des indices d'ajustement très convaincants (χ^2 / dl = 2,69 ; GFI = 0,950 ; AGFI = 0,895 ; RMSEA = 0,086, RMR = 0,241 ; CFI = 0,828). En particulier, le RMR et le CFI sont respectivement au dessus et en-dessous des limites acceptables. L'AGFI et le RMSEA sont, quant à eux, tout juste aux niveaux acceptables.

Un deuxième modèle est alors testé en supprimant les 2 items « problématiques » (CONFID 1 et PERT 2), identifiés lors de la réplication de l'analyse factorielle exploratoire[174]. Ce nouveau modèle (cf. figure 9.1) présente des indices d'ajustement tout à fait satisfaisants (χ^2 / dl = 1,23 ; GFI = 0,989 ; AGFI = 0,962 ; RMSEA = 0,032 ; RMR = 0,062 ; CFI = 0,991). De plus, le calcul du coefficient de Mardia (3,964) montre une absence de problème important en termes de multi-normalité. Ceci permet d'assurer une certaine « fiabilité » au niveau de l'estimation des paramètres, qui est confirmée par les résultats du *bootstrap* (cf. Annexe 9.1).
L'observation des contributions factorielles montre que la saturation des items sur leurs facteurs respectifs est globalement bonne puisqu'elle est systématiquement supérieure à 0,5.
Ce modèle « final », retenu pour son caractère parcimonieux, présente donc les 3 variables de perception (confidentialité, sensibilité et pertinence), mesurée chacune par deux items (CONFID2 et CONFID3, SENS1 et SENS2, PERT1 et PERT3).

[174] Une étape intermédiaire visant à supprimer un des 2 items est testée sans succès : la qualité du modèle s'améliore de manière substantielle sans parvenir toutefois à un niveau pleinement acceptable.

Figure 9.1 Modèle de mesure des trois variables de perception et indices d'ajustement

χ^2 / dl	GFI	AGFI	RMSEA	RMR	CFI
1,23	0,989	0,962	0,032	0,062	0,991

Afin de vérifier les propriétés psychométriques de ces instruments de mesure et notamment d'établir leur validité prédictive, l'impact des 3 variables de perception (confidentialité, sensibilité et pertinence) sur l'attitude face à la communication des données (variable conséquence dans le modèle à tester) est vérifié, au sein d'un nouveau modèle (cf. figure 9.2).

Ce modèle (cf. Figure 9.2), qui comprend 9 items (6 pour les variables de perception et 3 pour l'attitude), présente des indices d'ajustement satisfaisants (χ^2 / dl = 1,44 ; GFI = 0,971 ; AGFI = 0,937 ; RMSEA = 0,044, RMR = 0,118, CFI = 0,955). Le calcul du coefficient de multi-normalité (11,519) dénonce une anormalité modérée à importante (selon le seuil retenu) et justifie le recours à une analyse par *bootstrap*. Les résultats montrent une absence de violation importante, les écarts-types estimés par le *bootstrap* étant, dans l'ensemble, proches de ceux initialement calculés par le Maximum de Vraisemblance (cf. Annexe 9.2).

Nous procédons alors à l'examen de la fiabilité et de la validité des échelles sur les estimations issues de ce modèle.

Figure 9.2 Impact des trois variables de perception sur l'attitude et indices d'ajustement

χ^2 / dl	GFI	AGFI	RMSEA	RMR	CFI
1,44	0,971	0,937	0,044	0,118	0,955

Nous examinons tour à tour la fiabilité, la validité de trait (convergente et discriminante) puis la validité prédictive, en suivant les principes présentés au point 1.2.3 de ce chapitre.

La fiabilité de l'ensemble de ces échelles s'avère satisfaisante avec des coefficients de cohérence interne (ρ) respectifs de : CONFID = 0,795 ; SENS = 0,796 ; PERT = 0,773 ; ATT = 0,836. Concernant la validité convergente, les coefficients (ρVC) sont acceptables pour l'ensemble des construits : CONFID = 0,660 ; SENS = 0,661 ; PERT = 0,630 ; ATT = 0,634. Concernant la validité discriminante, l'examen des coefficients du modèle de mesure indique que le coefficient de validité convergente (ρVC) est systématiquement supérieur au carré du coefficient de corrélation entre les construits pris deux à deux (r^2_{ij}) (cf. tableau 9.24). La validité discriminante est donc clairement établie entre les concepts qui permettent de qualifier l'évaluation de la sollicitation (confidentialité, sensibilité et pertinence) d'une part et entre ces critères et l'attitude face à la communication des données d'autre part.

Tableau 9.24 Qualités psychométriques des échelles de perceptions et d'attitude

Variable latente	fiabilité		Validité convergente	Validité discriminante (r^2_{ij})		
	Alpha de Cronbach	Rhô de Jöreskog	ρVC	Sensibilité	Pertinence	Attitude
Confidentialité	0,77	0,795	0,660	0,194	0,270	0,058
Sensibilité	0,78	0,796	0,661		0,578	0,036
Pertinence	0,83	0,773	0,630			0,152
attitude	0,81	0,836	0,634			

Ce modèle conduit aussi à montrer la validité prédictive des échelles correspondant aux variables de perception (confidentialité, sensibilité et pertinence), dans la mesure où elles ont chacune un impact significatif sur l'attitude (respectivement 0,24 ; 0,39 et 0,20 ; $p \leq 0,05$).

Cette première analyse conduit donc à accepter l'unidimensionnalité, la validité et la fiabilité des échelles de confidentialité, de sensibilité et de pertinence. Les concepts correspondant à l'évaluation de la sollicitation sont ainsi mesurés en 2 items chacun.

Modèle 2 : Attitude face à la communication des données et Préoccupation pour le RVP

Concernant les concepts utilisés dans l'expérience 1, un dernier modèle comprenant l'attitude face à la communication des données et la préoccupation pour le RVP, soit 6 items (3 chacun), est également testé. Il présente des indices d'ajustement tout à fait satisfaisants (χ^2 / dl = 0,57 ; GFI = 0,993 ; AGFI = 0,983; RMSEA = 0,000 ; RMR = 0,042). L'examen du coefficient de Mardia (9,810) montre une absence de violation forte de la normalité, qui sera confirmée par les résultats du *bootstrap*.

La fiabilité des échelles de mesure correspondant aux concepts d'attitude et de préoccupation est satisfaisante avec des coefficients de cohérence interne (ρ) respectifs de : ATT = 0,830 et PREOC = 0,876. Concernant la validité convergente, les coefficients (ρVC) sont acceptables pour ces deux construits : ATT = 0,627 et PREOC = 0,702. Concernant la validité discriminante, l'examen des coefficients du modèle de mesure indique que le coefficient de validité convergente (ρVC) est systématiquement supérieur au carré du coefficient de corrélation entre ces deux construits (r^2_{ij}) = 0,311. La validité discriminante est donc démontrée entre l'attitude face à la communication des données et la préoccupation pour le respect de la vie privée. Enfin, nous testons la validité prédictive de la préoccupation (cf. Figure 9.3) en examinant le lien causal entre la préoccupation et l'attitude, ce dernier

apparaissant bien comme un concept « conséquence » au sein du modèle à tester. L'impact significatif de la préoccupation sur l'attitude étant démontré (- 0,558 ; p< 0,01), la validité prédictive de l'échelle de préoccupation pour le respect de la vie privée (PREOC) est donc clairement établie[175].

Figure 9.3 Impact de la préoccupation pour le RVP sur l'attitude et indices d'ajustement

χ^2 / dl	GFI	AGFI	RMSEA	RMR
0,57	0,993	0,983	0,000	0,042

Cette première analyse confirmatoire conduit donc à accepter l'unidimensionnalité, la validité et la fiabilité des échelles développées dans le cadre de l'expérience 1. Les concepts correspondant à l'évaluation de la sollicitation (confidentialité, sensibilité et pertinence) sont ainsi mesurés en 2 items chacun, tandis que les concepts d'attitude face à la communication des données et de préoccupation pour le RVP sont mesurés en 3 items.

3 Le développement des instruments de mesure nécessaires à l'expérience 2

Nous débutons la présentation par la spécification du domaine du construit et par les résultats de l'analyse exploratoire (3.1), pour les concepts faisant l'objet d'un développement ou d'une adaptation d'échelle dans le cadre de l'expérience 2. Nous abordons successivement les instruments de mesure relatifs à la valeur perçue (VAL) et à l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données (IPR). Nous poursuivons par l'analyse confirmatoire (précédée d'une réplication de l'analyse factorielle exploratoire) suivie d'un examen de la fiabilité et de la validité des instruments obtenus (3.2).

[175] Rappelons que nous testerons la validité prédictive de l'échelle d'attitude dans le cadre de l'expérience 2.

3.1 Spécification du domaine du construit et analyse exploratoire

3.1.1 L'échelle de valeur perçue

Dans le chapitre présentant notre cadre conceptuel, nous avons défini la valeur perçue comme une différence entre les bénéfices (ce qui est reçu) et les coûts (ce qui est donné) liés à la transaction (Zeithaml 1988, Bolton et Drew 1991). Plus précisément, dans le cadre qui nous intéresse ici, il s'agit d'une différence entre les bénéfices et les coûts liés à la fourniture d'informations personnelles à un marchand Internet. Si plusieurs échelles de valeur perçue existent au sein de la littérature, aucune d'entre elles n'était réellement adaptable au contexte étudié ici, en particulier à cause de la spécificité des bénéfices et des coûts engagés dans ce type de transaction. Nous avons donc décidé de développer une nouvelle échelle, propre à l'évaluation de la valeur perçue à fournir des données personnelles, en nous inspirant à la fois des verbatims récoltés lors de l'étude qualitative mais aussi de l'ensemble des résultats et échelles relevés dans la littérature.

Dans un premier temps, nous avons pensé nous inspirer des échelles d'équité trouvées dans la littérature, puisque ce concept repose, lui aussi, sur une comparaison coûts/bénéfices. Nous avons alors élaboré une première liste d'items (9 au total) soumis à l'évaluation de 5 experts. Ces items sont axés sur la comparaison des coûts et bénéfices entre le consommateur et l'entreprise. Deux de ces items sont ainsi formulés de la façon suivante : « dans cette situation, l'échange avec l'entreprise n'est pas équilibré » ; « je pense que l'entreprise gagne plus que moi à collecter ces données ». 4 items provenaient de l'échelle développée par Oliver et Swan (1989), un était repris de l'échelle de Ployart et Ryan (1997), les 4 autres étant développés par nos soins. Sur ces 9 items, 5 d'entre eux ont été considérés par les experts comme très représentatifs du construit sous-jacent (la valeur perçue). Ces 5 items ont alors été pré-testés auprès de consommateurs, de manière qualitative tout d'abord puis quantitativement. Les résultats de ce pré-test (analyse factorielle exploratoire) ont été très décevants : un item apparaît mal représenté (communalité très inférieure à 0,5) et trois autres items ont une distribution très anormale (coefficients d'asymétrie et d'aplatissement respectivement supérieurs à 1 et 1,5). Parmi les explications possibles, nous pensons que le consommateur perçoit toujours un déséquilibre quand il s'agit de comparer ses propres contributions et bénéfices à ceux de l'entreprise, ce qui se reflète dans le manque de variance obtenu sur les 3 items. Nous décidons donc d'abandonner cette piste et d'élaborer une échelle de valeur perçue fondée sur une comparaison coûts/bénéfices en « interne » c'est-à-dire au niveau de l'individu, sans qu'il ne se compare pour autant à d'autres intervenants potentiels.

Une nouvelle échelle est alors élaborée. Pour ce faire, nous commençons par réfléchir aux différents types de coûts et bénéfices liés à la fourniture d'informations personnelles en général, et sur Internet en particulier. Nous nous aidons en cela à la fois de la littérature existante mais aussi des résultats issus de notre étude qualitative. Trois types de coûts et trois types de bénéfices sont ainsi identifiés (cf. Annexe 9.3).

Parmi les coûts, on distingue ainsi :

- les coûts monétaires : que l'on peut considérer comme quasi nuls sur Internet
- les coûts d'énergie et d'opportunité : essentiellement limités à la charge cognitive liée au remplissage du questionnaire
- les coûts psychologiques : essentiellement liés aux craintes relatives au respect de la vie privée et à la sécurité lors de la transaction (et donc très importants ici)

Pour les bénéfices, nous reprenons une classification couramment utilisée dans la littérature, notamment dans le cadre de l'adhésion à un programme relationnel (Mimouni 2002), contexte qui nous semble proche de celui étudié ici puisqu'il s'agit d'un cas possible de sollicitation de données personnelles. Nous distinguons alors :

- les bénéfices utilitaires : liés à une réduction des coûts de recherche et/ou de décision ou à un gain monétaire (par exemple, des bons d'achat ou de réduction)
- les bénéfices hédoniques : liés à une dimension plaisir, au sentiment de faire une bonne action ou à l'espérance d'un gain (les deux premiers étant cependant relativement limités)
- les bénéfices de reconnaissance : liés à un meilleur ciblage ultérieur des offres

Sur cette base, nous élaborons une nouvelle liste d'items (une quinzaine au total) censés représenter ces différents types de coûts et bénéfices. Nous nous attendons ainsi à former une échelle multidimensionnelle, chacune des dimensions pouvant correspondre aux catégories de coûts et bénéfices. Nous choisissons un format à supports sémantiques (en 7 points) plutôt que le format Likert parce que nous le trouvons particulièrement adapté au construit à mesurer (une différence coûts/bénéfices) et, qu'en plus, il présente une efficacité globale satisfaisante (Vernette 1991). Cette nouvelle liste est soumise à évaluation auprès de 3 experts, dans le but de la réduire aux items les plus représentatifs du construit à mesurer (la valeur perçue) et d'assurer ainsi une validité de contenu satisfaisante. Sur les 16 items soumis à évaluation (8 pour les coûts, autant pour les bénéfices), 12 d'entre eux sont ainsi retenus (7 pour les coûts, 5 pour les bénéfices), les uns directement, les autres après modifications préalables. Ces 12 items retenus pour former l'échelle de valeur perçue (format sémantique différentiel en 7 points) sont présentés au tableau 9.25, donné ci-dessous.

Tableau 9.25 Echelle de mesure initiale de la valeur perçue (VAL)

Items	Coûts	Bénéfices
Répondre aux questions de ce formulaire vous paraît :		
1 simple / compliqué (inversé)	X (énergie)	
2 risqué / pas risqué	X (psycho.)	
3 utile / inutile		X (util. et recon.)
4 facile / difficile (inversé)	X (énergie)	
5 ennuyeux / pas ennuyeux	X (énergie)	
6 apporter beaucoup / ne rien apporter		X (util. et recon.)
7 bénéfique / pas bénéfique		X (util. et recon.)
8 agréable / désagréable		X (hédonique)
9 une perte de temps / mériter d'y consacrer du temps	X (énergie)	
10 avoir des avantages / n'avoir aucun avantage		X (util. et recon.)
11 avoir des effets négatifs / avoir des effets positifs	X (psycho.)	
12 coûter beaucoup / ne rien coûter	X (mon. et psy.)	

Avant de mener l'analyse factorielle, nous procédons à une analyse descriptive univariée sur ces 12 items. Les résultats du pré-test mené auprès de 93 étudiants montrent que les items 1 (simple/compliqué) et 4 (utile/inutile) posent des problèmes de normalité qui nous conduisent à les supprimer (indices d'asymétrie et d'aplatissement respectivement de 1,674 et 1,756 pour l'item 1 et de 1,948 et 2,844 pour l'item 4).

Du fait de la définition même de ce concept (différence entre des bénéfices et des coûts qui sont, par essence, opposés) et de la construction de l'échelle sous-jacente (items correspondant à plusieurs sous-dimensions de coûts et de bénéfices), nous soupçonnons, d'une part, l'existence de plusieurs dimensions et, d'autre part, que celles-ci soient fortement corrélées entre elles. Nous décidons alors de mener une analyse factorielle avec rotation oblique (puisque nous souhaitons permettre aux facteurs d'être corrélés entre eux[176]) sur les 10 items restants. Il existe plusieurs types de rotations obliques possibles, parmi lesquelles nous choisissons Promax[177], qui offre l'avantage d'être à la fois rapide et conceptuellement simple. Cette méthode vise en effet à trouver une matrice qui ait une structure facile à interpréter.

Cette analyse donne les résultats suivants (KMO = 0,723 ; sig. du test de Bartlett = 0,000) :

[176] Les résultats issus de la matrice de corrélations entre les facteurs (corrélation proche de 0,4 pour deux d'entre eux) confirment la pertinence de ce choix, tout comme les conclusions issues de l'analyse confirmatoire, présentée plus loin dans ce chapitre.

[177] Par mesure de comparaison, nous menons aussi l'analyse avec rotation oblique de type « Oblimin ». Nous trouvons que les résultats pour ces deux types de méthodes ne sont pas fondamentalement différents entre eux.

Tableau 9.26 Résultat de l'analyse factorielle exploratoire sur l'échelle de valeur perçue

Items	Facteur 1	Facteur 2	Facteur 3	Qualité de représentation
VAL 7	**0,988**	(*)		0,864
VAL 6	**0,892**			0,779
VAL 3	**0,712**			0,535
VAL 10	**0,686**	0,399		0,481
VAL 5		**- 0,881**		0,716
VAL 8		**0,815**		0,658
VAL 9		**- 0,581**	0,298	0,519
VAL 2			**0,878**	0,744
VAL 11			**0,769**	0,643
VAL 12			**0,630**	0,487
% variance expliquée	35,25	16,27	11,79	

Méthode de rotation : Promax avec normalisation de Kaiser

(*) Coefficients inférieurs à 0,2 omis

La qualité de représentation des items est satisfaisante, hormis pour les items 10 (avoir des avantages/n'avoir aucun avantage) et 12 (coûter beaucoup/ne rien coûter) avec respectivement 0,481 et 0,487. Nous conservons néanmoins ces 2 items du fait de la proximité avec le seuil communément admis de 0,5 et pour ne pas compromettre (en tout cas immédiatement) la validité de contenu de l'échelle. La saturation des items est également satisfaisante (systématiquement supérieure à 0,5), hormis pour les items 10 (avantages) et 9 (perte de temps/mérite d'y consacrer du temps) qui saturent sur 2 axes et pour lesquels la différence de contribution entre l'axe principal et l'autre axe est légèrement inférieure à 0,3 (respectivement 0,287 et 0,283). L'analyse factorielle laisse ainsi apparaître 3 facteurs ayant une contribution respective de 35% ; 16% et 12%, soit un pourcentage global de variance expliquée de 63,3%. Ces trois facteurs semblent correspondre respectivement à une dimension « bénéfices » ou « intérêt » (avec les items 7, 6, 3 et 10), une dimension « coût d'énergie » (items 5, 8 et 9) et une dimension « coût monétaire et psychologique » (items 2, 11 et 12).

Ces résultats nous conduisent donc à conclure à l'existence de 3 dimensions visant à mesurer le concept de valeur perçue. L'ensemble des 10 items retenus feront ensuite l'objet d'une analyse confirmatoire afin de contrôler la présence de ces trois dimensions.

3.1.2 L'échelle d'intention de poursuivre la relation

Pour Zeithalm et al. (1996), l'engagement est un indicateur de la qualité perçue du service apporté par l'entreprise. Cette notion d'engagement, que nous avons largement utilisée pour définir le concept d'intention de poursuivre la relation dans le chapitre 7, correspond à tout type de comportement qui vise à rester dans l'échange (Halaby et Weakliem 1989). En référence au concept d'engagement, l'intention de poursuivre la relation peut alors se définir comme « la croyance selon laquelle la relation avec l'entreprise continuera à l'avenir » (Bennett et Gabriel 2001). Rappelons que nous étudions ici une relation « particulière » dans la mesure où elle concerne et le consommateur et une entité commerciale (l'entreprise représentée par le site qu'elle met à disposition des visiteurs et de ses clients). Selon Zeithaml et al. (1996), dans le cadre d'une cyber-relation, cette croyance selon laquelle la relation continuera se manifesterait alors par le fait de revenir sur le site, d'acheter sur le site, par des remarques positives sur le site et par le fait de recommander le site aux autres.

Il existe, au sein de la littérature, de multiples échelles visant à mesurer l'engagement et pouvant donc s'appliquer au concept d'intention de poursuivre la relation. Nous avons cherché dans les études existantes une échelle courte, simple et valide pouvant s'appliquer facilement au contexte étudié ici (la sollicitation de données personnelles et ses conséquences sur les relations futures avec l'entreprise sollicitante). Parmi l'ensemble des instruments de mesure trouvés, nous avons retenu l'échelle de Moore et Benbasat (1991), reprise et adaptée par Plouffe et al. (2001) puis par Wallet-Wodka (2003), visant à mesurer respectivement l'intention d'adoption d'une innovation, d'une innovation virtuelle (système d'information et carte de paiement électronique) et d'un intermédiaire d'achat virtuel (site d'agence de voyages). Bien que paraissant *a priori* éloignée du concept que nous cherchons à mesurer (intention de poursuivre la relation), la lecture des items confirme son adaptation au contexte étudié d'une part et au construit d'autre part. De plus, il s'agit d'une échelle courte (3 ou 4 items) ayant des qualités psychométriques satisfaisantes. Nous choisissons donc de reprendre ces items, tout en adaptant leur formulation au cas étudié ici. Ces items sont alors soumis à évaluation auprès d'experts et de consommateurs afin d'assurer une validité de contenu satisfaisante. Suite à cette évaluation, les 4 items proposés (après reformulation des items d'origine) sont conservés pour former l'échelle d'intention de poursuivre la relation (IPR), présentée au tableau 9.27 ci-dessous.

Tableau 9.27 Echelle de mesure de l'intention de poursuivre la relation (IPR)

Echelle de Plouffe et al. en 4 items (2001)	Echelle de Wallet-Wodka en 3 items (2003)	Echelle retenue en 4 items
Je serai intéressé(e) de continuer à utiliser ce système Je m'arrangerai pour détenir cette carte de façon permanente Je ne vois pas vraiment l'intérêt de continuer à utiliser cette carte de paiement (inversé) Je recommanderai à mes amis d'obtenir cette carte pour faire leurs achats	Je pense que je continuerai à utiliser ce site d'agence de voyages dans le futur Je m'arrangerai pour acheter des voyages sur ce site d'agence de voyages Je ne vois pas vraiment l'intérêt de revenir sur ce site d'agence de voyages dans le futur (inversé) (omis)	(IPR 1) Je pense que je continuerai à utiliser les services de [cet opérateur] dans le futur (IPR 2) Je pense que je resterai chez [cet opérateur] pendant longtemps[178] (IPR 3) Je ne vois pas l'intérêt de participer aux opérations proposées[179] par [cet opérateur] dans le futur (inversé)[180] (IPR 4) Je recommanderai à mes amis de se rendre sur ce site afin de bénéficier des mêmes offres que moi

Pour rappel, l'échelle d'intention de poursuivre la relation doit être utilisée dans le cadre des expériences 2 et 3. Malgré notre volonté initiale de conserver les 4 items ci-dessus dans les deux cas, plusieurs modifications ont cependant dû intervenir que nous présentons et justifions rapidement ci-dessous. Au final, quelques différences existent entre les échelles utilisées dans chacune des expériences. L'expérience 2 simule une sollicitation de données personnelles par un opérateur de téléphonie mobile (d'où le terme « cet opérateur »), tandis que dans l'expérience 3, il est fait référence à un programme de fidélité appelé Elisa (« cet opérateur » est alors remplacé par « le programme Elisa »). La seconde différence est liée au format de réponse retenu. Dans l'expérience 2, plusieurs intentions comportementales (autres que celles liées à la poursuite de la relation) sont mesurées, en utilisant un format en 5 échelons (de « oui certainement » à « non certainement pas »). Pour ne pas perturber les répondants, nous conservons ce format pour l'échelle d'intention de poursuivre la relation (alors que nous utiliserons un format Likert en 7 points dans l'expérience 3). Lors du pré-test des échelles utilisées pour l'expérience 2, l'évaluation de l'item IPR 3 (inversé) sur l'échelle allant de « oui certainement » à « non certainement pas » s'est avérée délicate, voire quasi impossible. Par conséquent, nous décidons de ne pas inverser cet item qui est alors formulé de la manière suivante « je participerai aux opérations proposées par cet opérateur dans le futur ». L'inversion est cependant conservée dans l'expérience 3.

[178] Suite aux recommandations des experts et à l'évaluation des consommateurs, l'item initialement formulé « je m'arrangerai pour acheter mon prochain forfait chez cet opérateur » a été reformulé ainsi.
[179] La formulation « participer aux opérations proposées par … » a été préférée à celle de « revenir sur le site de … » dans la mesure où pour un opérateur de téléphonie mobile, les fréquences de retour et de consultation du site sont très aléatoires et très espacées. Cette formulation était également plus adaptée pour l'expérience 3.
[180] Dans l'expérience 2, l'item IPR 3 ne sera pas inversé, pour les raisons exposées ci-après.

Cette échelle en 4 items est soumise à un pré-test mené auprès de 93 étudiants. Avant de réaliser l'analyse factorielle, nous procédons à une simple analyse descriptive univariée. Les résultats montrent que l'ensemble des items a une distribution normale. Les 4 items retenus sont soumis à une analyse factorielle (sans rotation puisque l'échelle est unidimensionnelle) qui donne les résultats suivants (KMO = 0,641 ; sig. du test de Bartlett = 0,000) :

Tableau 9.28 Résultat de l'AFE sur l'échelle d'intention de poursuivre la relation (IPR)

Items	Qualité de la représentation	Saturation (sur l'axe 1)
IPR 1 Je pense que je continuerai à utiliser les services de [cet opérateur] dans le futur	0,506	0,554
IPR 2 Je pense que je resterai chez [cet opérateur] pendant longtemps	0,803	0,873
IPR 3 Je participerai aux opérations proposées par [cet opérateur] dans le futur (non inversé)	0,816	0,889
IPR 4 Je recommanderai à mes amis de se rendre sur ce site afin de bénéficier des mêmes offres que moi	0,823	0,903
% de variance expliquée	73,71%	
Alpha de Cronbach	0,74	

La qualité de représentation des items tout comme leur saturation est satisfaisante, hormis pour l'item IPR 1 qui atteint tout juste le seuil acceptable de 0,5. L'alpha est satisfaisant (0,74), proche de celui obtenu par Wallet-Wodka (2003) sur l'échelle à 3 items (0,78).

Les résultats des analyses exploratoires menées sur les échelles de valeur perçue et d'intention de poursuivre la relation se révèlent donc satisfaisants et conduisent à accepter provisoirement la tridimensionnalité de l'échelle de valeur perçue et l'unidimensionnalité de celle d'intention de poursuivre la relation. Les analyses confirmatoires présentées ci-après visent à vérifier ces résultats et à s'assurer de la qualité psychométrique des instruments développés.

3.2 L'analyse confirmatoire

Comme précédemment, nous commençons par présenter les résultats issus de la réplication de l'analyse exploratoire sur un nouvel échantillon. Nous indiquons ensuite les résultats des modèles de mesure estimés par le logiciel AMOS 5.0, avant d'aborder la question de la validité et de la fiabilité. Nous rappelons que cette analyse confirmatoire concerne les deux construits ayant fait l'objet d'un développement ou d'une adaptation d'échelle dans le cadre de l'expérience 2 (la valeur perçue et l'intention de poursuivre la relation). Nous débutons la

présentation par la réplication de l'analyse factorielle exploratoire sur un nouvel échantillon de données (collecte 2) puis présentons le modèle de mesure et les propriétés psychométriques de ces deux échelles.

3.2.1 La réplication de l'analyse exploratoire

La réplication de l'analyse factorielle exploratoire s'est donc opérée sur un nouvel échantillon composé de 252 internautes panélistes (échantillon sur lequel a été menée l'expérience 2). Nous présentons successivement les résultats concernant la valeur perçue (VAL) et l'intention de poursuivre la relation (IPR).

La réplication de l'analyse factorielle sur l'échelle de valeur perçue

Rappelons que la première analyse factorielle, menée sur un échantillon de 93 étudiants, aboutissait à la formation de 3 facteurs nommés respectivement « bénéfices » ou « intérêt » (items 7, 6, 3 et 10), « coût d'énergie » (items 5, 8 et 9) et « coût monétaire et psychologique » (items 2, 11 et 12). L'ensemble de ces items est soumis à une nouvelle analyse factorielle avec rotation Promax (du fait de la corrélation entre les facteurs), précédée d'une analyse descriptive univariée, visant à vérifier la normalité des distributions pour chacune des variables concernées. Les résultats de cette analyse descriptive confirment l'absence de problèmes de normalité, les coefficients d'asymétrie et d'aplatissement étant inférieurs aux seuils respectifs de 1 et de 1,5 (cf. tableau 9.29).

Tableau 9.29 Coefficients d'asymétrie et d'aplatissement (*) pour les items de valeur perçue

Items	Coefficient d'asymétrie	Coefficient d'aplatissement
VAL 2 : risqué	0,013	0,883
VAL 3 : utile	0,064	0,925
VAL 5 : ennuyeux	0,043	1,064
VAL 6 : apporter beaucoup	0,148	0,752
VAL 7 : bénéfique	0,036	0,640
VAL 8 : agréable	0,003	0,460
VAL 9 : perte de temps	0,016	0,799
VAL 10 : avantages	0,381	0,669
VAL 11 : effets négatifs	0,047	0,325
VAL 12 : coûter beaucoup	0,504	0,539

* Résultats donnés en valeur absolue

L'analyse factorielle (avec rotation Promax) menée sur ces 10 items (KMO = 0,880 ; sig. du test de Bartlett =0,000) laisse apparaître 2 facteurs (correspondant respectivement aux notions d'intérêt et de coût déjà identifiées précédemment) bénéficiant d'une corrélation de -0,5. A noter que la séparation initiale en « coût d'énergie » et « coût monétaire et psychologique » n'apparaît plus ici, ceux-ci se regroupant pour ne former qu'une seule et même catégorie. Ces deux facteurs, nommés « intérêt » et coût », expliquent respectivement 49,5% et 13,9% de la variance soit un total de 63,4% de variance expliquée. A la lecture des résultats, on constate que l'item « agréable » (VAL 8) a une qualité de représentation (0,461) inférieure au seuil acceptable de 0,5 ; il sera donc supprimé de l'analyse. De plus, les items « perte de temps » (VAL 9), « effets négatifs » (VAL 11) et ennuyeux (VAL 5) saturent sur les deux facteurs, la différence de contribution étant très inférieure à 0,3 (respectivement 0,244 ; 0,209 et 0,200). Ces trois items sont donc également supprimés de l'analyse.

Tableau 9.30 Résultat de la réplication de l'analyse factorielle sur l'échelle de valeur perçue

Items	Facteur 1	Facteur 2	Qualité de représentation
VAL 7 : bénéfique	**0,884**	(*)	0,715
VAL 6 : apporter beaucoup	**0,883**		0,668
VAL 10 : avantages	**0,828**		0,651
VAL 3 : utile	**0,816**		0,673
VAL 8 : agréable	**0,640**		*0,461*
VAL 9 : perte de temps	***- 0,613***	*0,369*	0,613
VAL 12 : coûter beaucoup		**0,945**	0,719
VAL 2 : risqué		**0,820**	0,652
VAL 11 : effets négatifs	*- 0,399*	***0,608***	0,641
VAL 5 : ennuyeux	*- 0,314*	***0,514***	0,524
% Variance expliquée	49,49	13,88	

Méthode de rotation : Promax avec normalisation de Kaiser

(*) Coefficients inférieurs à 0,2 omis

Une nouvelle analyse factorielle (avec rotation Promax) est menée sur les 6 items restants (KMO = 0,788 ; sig. du test de Bartlett = 0,000). Cette fois-ci, la qualité de représentation est supérieure à 0,5 pour tous les items et chaque item sature bien sur un seul facteur (cf. tableau 9.31). On retrouve les deux facteurs précédents (intérêt et coût), représentant respectivement 52,4% et 20,9% de variance expliquée, soit un total de 73,3%. La dimension « intérêt » est mesurée par 4 items (« apporter beaucoup », « bénéfique », « avantages » et « utile ») tandis que la dimension « coût » n'en comprend que 2 (« coûte beaucoup » et « risqué »).

Nous vérifions la fiabilité des échelles ainsi constituées en calculant l'alpha de Cronbach. Nous obtenons 0,86 pour le facteur « intérêt »[181] et 0,70[182] pour le facteur « coût ».

Tableau 9.31 Résultat de la dernière analyse factorielle sur l'échelle de valeur perçue

Items	Facteur 1	Facteur 2	Qualité de représentation
VAL 7 : bénéfique	**0,881**	(*)	0,742
VAL 6 : apporter beaucoup	**0,863**		0,735
VAL 3 : utile	**0,814**		0,685
VAL 10 : avantages	**0,807**		0,684
VAL 12 : coûter beaucoup		**0,931**	0,814
VAL 2 : risqué		**0,810**	0,739
% Variance expliquée	52,41	20,91	
Alpha de Cronbach	0,86	0,70	

Méthode de rotation : Promax avec normalisation de Kaiser

(*) Coefficients inférieurs à 0,2 omis

La réplication de l'analyse factorielle sur l'échelle d'intention de poursuivre la relation

Après avoir vérifié la normalité des distributions correspondant aux 4 items de l'échelle d'intention de poursuivre la relation (coefficients d'asymétrie et d'aplatissement respectivement inférieurs à 1 et 1,5), nous menons une nouvelle analyse factorielle (KMO = 0,664 ; sig. du test de Bartlett = 0,000) qui donne les résultats suivants (cf. tableau 9.32) :

Tableau 9.32 Réplication de l'analyse factorielle sur l'intention de poursuivre la relation

Items	Qualité de la représentation	Saturation (sur l'axe 1)
IPR 1 Je pense que je continuerai à utiliser les services de [cet opérateur] dans le futur	0,501	0,708
IPR 2 Je pense que je resterai chez [cet opérateur] pendant longtemps	0,632	0,795
IPR 3 Je participerai aux opérations proposées par [cet opérateur] dans le futur (non inversé)	0,519	0,721
IPR 4 Je recommanderai à mes amis de se rendre sur ce site afin de bénéficier des mêmes offres que moi	0,610	0,781
% de variance expliquée	66,56%	
Alpha de Cronbach	0,73	

[181] La suppression d'un des items fait systématiquement baisser l'alpha, signe de l'apport significatif de chacun d'eux.

[182] L'alpha est donné à titre indicatif puisqu'il n'est calculé que sur 2 items.

La qualité de représentation des items est acceptable, bien que celle correspondant aux items IPR 1 et IPR 3 atteigne tout juste le seuil de 0,5. La saturation des items est également satisfaisante puisque systématiquement supérieure à 0,5. La fiabilité de l'échelle est convenable avec un alpha de 0,73.

3.2.2 L'analyse confirmatoire et la qualité des échelles

Après avoir répliqué l'analyse factorielle exploratoire afin de s'assurer de la structure des échelles, nous arrivons à présent à l'analyse confirmatoire qui procède en deux temps : l'estimation d'un modèle de mesure (par le biais d'indices d'ajustement) et la vérification des qualités psychométriques des instruments de mesure. Nous présentons ces résultats successivement pour les deux échelles concernées (valeur perçue et intention de poursuivre la relation). Nous ne revenons pas sur les étapes préalables à l'estimation du modèle, celles-ci étant identiques à celles réalisées pour les échelles relatives à l'expérience 1 et présentées à la section 2. Nous rappelons qu'il n'existe ici aucune valeur manquante, que la taille de l'échantillon est suffisante (N = 252), que la normalité des données a été vérifiée et que nous utilisons la méthode du maximum de vraisemblance.

Analyse confirmatoire de l'échelle de valeur perçue

Nous estimons un premier modèle comprenant les 6 items correspondant aux deux facteurs (« intérêt » et « coût ») relatifs à l'échelle de valeur perçue. Ce modèle (cf. Figure 9.4) présente des indices d'ajustement satisfaisants (χ^2 / dl = 0,84 ; GFI = 0,991 ; AGFI = 0,997 ; RMSEA = 0,000, RMR = 0,042).

L'analyse des contributions factorielles montre que les saturations des items sur leurs facteurs respectifs sont bonnes (supérieures à 0,63). De plus, la forte corrélation entre les deux dimensions (ou facteurs) (- 0,63), soupçonnée dès l'analyse exploratoire (d'où le choix d'une rotation oblique), est confirmée ici. En outre, ceci fait penser à l'existence d'un facteur de second ordre, qui correspondrait au construit de valeur perçue (définie comme la différence entre les coûts et bénéfices).

Figure 9.4 Modèle de mesure des deux dimensions de valeur perçue et indices d'ajustement

χ^2 / dl	GFI	AGFI	RMSEA	RMR
0,84	0,991	0,997	0,000	0,042

Nous testons alors un nouveau modèle qui intègre un facteur de second ordre, nommé « valeur » (cf. Figure 9.5). Celui-ci est mesuré par un seul item, intégré dans le questionnaire, correspondant à une « valeur globale » (VAL) et formulé de la façon suivante : « j'estime qu'il y a plus de risque à répondre que de bénéfices à en retirer ». Ce nouveau modèle présente des indices d'ajustements tout à fait satisfaisants (χ^2 / dl = 1,15 ; GFI = 0,985 ; AGFI = 0,965 ; RMSEA = 0,024 ; RMR = 0,084 ; CFI = 0,998).

L'examen du coefficient de multi-normalité (20,988) fait cependant penser à un problème de normalité excessive au sein du modèle et justifie le recours à un *bootstrap*, afin de nous assurer que l'estimation des paramètres n'est pas biaisée. L'estimation des écarts-types et leur comparaison avec ceux estimés par la méthode « classique » indiquent qu'il n'existe pas de biais important. Afin de conforter ce résultat, nous calculons un nouvel indice d'estimation de la validité du modèle, appelé « Bollen-Stine ». Les résultats donnés par AMOS indiquent que le modèle estimé par *bootstrap* est meilleur dans 575 échantillons (sur les 2000 testés), soit une probabilité de 575/2000 = 0,713, correspondant à l'estimation globale du modèle (*p-value*). Au seuil usuel de 5%, on peut donc conclure qu'il ne faut pas rejeter ce modèle. Cette valeur est d'ailleurs supérieure au seuil de significativité correspondant à l'estimation par la méthode du maximum de vraisemblance (chi-deux = 13,79 ; dl = 12 ; p= 0,314).

Ayant démontré la « fiabilité » de ce modèle, nous pouvons analyser plus en détail l'estimation des différents paramètres. La corrélation entre les items et la variable latente qu'ils sont censés mesurer est donnée par les coefficients de régression et par l'estimation des poids standardisés (ou corrélations multiples au carré[183]). Les résultats de l'estimation des paramètres (fournis en Annexe 9.4) montrent que les coefficients de régression et les SMC sont supérieurs à 0,5 hormis pour l'item « risqué » dont le SMC est proche du niveau acceptable (0,480). Ce modèle est donc satisfaisant au regard des indices d'ajustement et de l'estimation des paramètres. Nous pouvons alors nous baser dessus pour évaluer les qualités psychométriques de l'échelle de valeur perçue.

Figure 9.5 Modèle de mesure de l'échelle de valeur perçue et indices d'ajustement

χ^2 / dl	GFI	AGFI	RMSEA	RMR	CFI
1,15	0,985	0,965	0,024	0,084	0,998

La vérification des qualités psychométriques de l'échelle de valeur perçue concerne successivement la fiabilité, la validité de trait (convergente et discriminante) et la validité prédictive. La fiabilité de chacune des dimensions de l'échelle (intérêt et coût), vérifiée précédemment par l'alpha de Cronbach, est confirmée par les résultats du calcul du Rhô de

[183] En anglais SMC pour *Squared Multiple Correlations*

Jöreskog qui est respectivement de 0,86 et 0,76. La validité convergente, censée vérifier que la variance partagée entre un construit et sa mesure est supérieure à la variance due aux termes d'erreur est aussi vérifiée, la valeur du rhô de convergence étant supérieure à 0,5 pour chacune des dimensions (respectivement 0,614 et 0,619). La validité discriminante vérifie que la variance partagée entre les construits est inférieure à la variance moyenne partagée entre un construit et sa mesure, pour chacune des dimensions. Elle est également établie, le coefficient de validité convergente étant supérieur au carré du coefficient de corrélation entre les deux dimensions ($\rho VC > 0,397$). Le tableau 9.33 présente l'ensemble des résultats concernant la fiabilité et la validité de trait de l'échelle. Enfin, la validité prédictive est acquise par la validation d'un lien significatif entre chacune des dimensions et la mesure de valeur globale. En effet, les valeurs SMC indiquent que plus de 60% de la variance de la valeur globale est attribuée à ses deux dimensions « intérêt et « coût ».

Tableau 9.33 Qualités psychométriques des dimensions de l'échelle de valeur perçue

Variable latente	fiabilité		Validité convergente	Validité discriminante (r^2_{ij})	
	Alpha de Cronbach	Rhô de Jöreskog	ρVC	Intérêt	Coût
Intérêt	0,86	0,864	0,614		
Coût	0,70	0,759	0,619	0,397	

Les qualités psychométriques de l'échelle de valeur perçue mesurée en 6 items (4 pour la dimension « intérêt » et 2 pour la dimension « coût ») ayant été clairement établies, passons aux résultats concernant l'échelle d'intention de poursuivre la relation.

Analyse confirmatoire de l'échelle d'intention de poursuivre la relation

Nous estimons un premier modèle comprenant les 4 items relatifs à cette échelle d'intention de poursuivre la relation. Ce modèle présente des indices d'ajustement très médiocres (χ^2 / dl = 20,80 ; GFI = 0,921 ; AGFI = 0,607 ; RMSEA = 0,281, RMR =0,095 ; CFI = 0,840). L'analyse des contributions factorielles prouve que les saturations des items sur leurs facteurs respectifs sont bonnes (supérieures à 0,55). Concernant la variance des termes d'erreur, l'examen des covariances des résidus standardisés montre que l'item IPR 1 pose problème. Il faut donc le supprimer du modèle qui devient alors tout juste identifié (le nombre de degré de liberté est égal à zéro), ce qui empêche le calcul des indices d'ajustement.

Pour prouver la qualité de l'échelle d'intention de poursuivre la relation (désormais réduite aux 3 derniers items), nous devons donc estimer un nouveau modèle qui intègre l'intention de poursuivre la relation avec un autre construit. Nous choisissons l'échelle d'attitude, dont les qualités psychométriques ont déjà été établies pour l'expérience 1 (cf. paragraphe 2.2.2) et qui est à nouveau utilisée dans le cadre de l'expérience 2.

Ce nouveau modèle (cf. Figure 9.6), comprenant 6 items (3 pour l'intention de poursuivre la relation, autant pour l'attitude), présente cette fois-ci des indices d'ajustement convenables (χ^2 / dl = 2,65 ; GFI = 0,970 ; AGFI = 0,921 ; RMSEA = 0,081, RMR =0,091 ; CFI = 0,977).

Le calcul de coefficient de Mardia (5,243) permet de conclure en l'absence d'anormalité excessive susceptible de biaiser l'estimation des paramètres, ce qui est confirmé par le *bootstrap*. Sur la base de ce modèle (cf. Annexe 9.5), nous estimons alors conjointement la fiabilité et la validité de trait de ces deux instruments de mesure, même si ces qualités ont déjà été vérifiées précédemment pour l'attitude. Nous pouvons également estimer la validité prédictive de l'échelle d'attitude en remplaçant le lien de corrélation entre les deux variables latentes par un simple lien direct (allant de l'attitude à l'intention de poursuivre la relation[184]). La validité prédictive de l'échelle d'intention de poursuivre sa relation ne pourra quant à elle être vérifiée, faute de construit susceptible de faire office de « variable conséquence ».

Figure 9.6 Modèle de mesure de l'attitude et de l'intention de poursuivre la relation

χ^2 / dl	GFI	AGFI	RMSEA	RMR	CFI
2,65	0,970	0,921	0,081	0,091	0,977

[184] Même si ce lien n'est pas présent au sein de notre modèle à tester, nous souhaitions vérifier l'existence d'une telle relation. Les résultats issus de l'expérience 2 et présentés au chapitre 11 confirmeront ce résultat.

Le tableau 9.34 ci-dessous résume les qualités psychométriques des échelles d'attitude et d'intention de poursuivre la relation. Ainsi, la fiabilité de ces deux échelles (ATT et IPR), vérifiée précédemment par l'alpha de Cronbach, est confirmée par les résultats du calcul du Rhô de Jöreskog qui est respectivement de 0,82 et 0,71. La validité convergente est aussi vérifiée, la valeur du rhô de convergence étant supérieure à 0,5 pour les 2 construits (respectivement 0,622 et 0,503), même s'il atteint tout juste le seuil acceptable pour l'intention de poursuivre la relation. Concernant la validité discriminante, le rhô de convergence est supérieur au carré du coefficient de corrélation entre les deux construits, même si ce n'est que très légèrement (0,503 > 0,49). Cette différence étant très faible, nous décidons de recourir à la méthode préconisée par Roussel et al. (2002), afin d'établir avec plus de certitude la validité discriminante entre ces 2 construits. Nous comparons alors un modèle où les corrélations entre les deux variables latentes sont fixées à 1 (modèle contraint) à un modèle où les corrélations sont libres (modèle « libre »). L'examen des indices d'ajustement pour ces deux modèles révèle que le modèle libre est meilleur que le modèle contraint. Par ailleurs, la différence de chi-deux entre ces deux modèles est significative ($\chi^2_{\text{modèle contraint}}$ = 35,565 et $\chi^2_{\text{modèle libre}}$ = 21,165 ; χ^2_{diff} = 14,399 ; p = 0,000), ce qui permet de conclure qu'il y a bien validité discriminante entre ces deux construits.

Enfin, la validité prédictive de l'échelle d'attitude est acquise par la validation d'un lien significatif (0,70) entre cette dernière et l'intention de poursuivre la relation.

Tableau 9.34 Qualités psychométriques de l'échelle d'intention de poursuivre la relation

Variable latente	fiabilité		Validité convergente	Validité discriminante (r^2_{ij})	
	Alpha de Cronbach	Rhô de Jöreskog	ρVC	IPR	Attitude
IPR	0,70	0,715	0,503		
Attitude	0,81	0,825	0,622	0,490	

Les qualités psychométriques de l'échelle d'intention de poursuivre la relation composée de 3 items ont donc été établies, excepté ce qui concerne la validité prédictive.

Cette seconde analyse confirmatoire conduit donc à accepter la validité et la fiabilité des échelles développées dans le cadre de l'expérience 2. Le concept de valeur perçue est ainsi mesuré en 6 items correspondant à deux dimensions, l'une relative aux bénéfices et l'autre aux coûts, tandis que le concept d'intention de poursuivre la relation est mesuré en 3 items.

Conclusion du chapitre 9

Pour conclure ce chapitre, l'analyse de la littérature et les études menées auprès d'experts et de consommateurs ont permis de retenir des instruments de mesure simples et s'administrant rapidement pour éviter les biais de réponse et la mortalité expérimentale.

Le processus de création et/ou d'adaptation des instruments de mesure utilisés dans le cadre des expériences 1 et 2 a ensuite fait l'objet d'une attention particulière, afin de développer des échelles fiables et valides.

La réalisation d'une analyse exploratoire et confirmatoire sur la quasi-totalité des échelles développées a permis de conclure à la fiabilité et à la validité de l'ensemble des instruments de mesure créés.

En ce qui concerne les instruments de mesure développés pour les besoins de l'expérience 1, les trois variables de perception s'avèrent bien unidimensionnelles et sont mesurées par 2 items chacune, tandis que l'attitude face à la communication des données et la préoccupation pour le respect de la vie privée ont chacune 3 items.
Enfin, concernant les instruments de mesure utilisés dans l'expérience 2 et non développés pour les besoins de l'expérience 1, l'échelle de valeur perçue apparaît bidimensionnelle (la dimension « intérêt » comprend 4 items contre 2 seulement pour celle de « coût »), tandis que l'échelle d'intention de poursuivre la relation est mesurée en 3 items.

La validité et la fiabilité de l'ensemble des instruments de mesure ayant été démontrées, nous pouvons à présent évoquer la mise en œuvre et les résultats issus des 3 expériences menées successivement.

CONCLUSION DE LA PARTIE II

Cette seconde partie, qui visait essentiellement à bâtir le cadre conceptuel général de cette recherche a permis de répondre aux questions suivantes :

1. Comment un individu perçoit-il la sollicitation de ses données personnelles ?

L'étude qualitative, menée auprès de consommateurs français, a permis d'identifier quatre critères sur lesquels repose prioritairement l'évaluation d'une sollicitation : il s'agit de la confidentialité, de la sensibilité et de la pertinence perçues des données sollicitées, auxquelles s'ajoute la perception de la valeur (ou du gain net) à répondre.

2. Comment les consommateurs français gèrent-ils leurs données personnelles ?

L'étude descriptive a permis de segmenter la population des internautes français en quatre catégories qui se distinguent sur des éléments de nature sociodémographique, expérientielle et comportementale. Elle souligne aussi l'importance de deux facteurs qui semblent plus particulièrement influencer la décision de fournir les renseignements : la familiarité avec la firme sollicitant les données et le contexte dans lequel celles-ci sont demandées.

3. Peut-on modéliser le processus de décision du consommateur confronté à une sollicitation de ses données personnelles ?

Les résultats préalables ont permis de bâtir le cadre conceptuel général de cette recherche qui suit une logique essentiellement processuelle.

4. Quelle méthodologie utiliser pour tester les différentes hypothèses de recherche ?

Le chapitre 8 a été l'occasion de justifier le choix d'une méthodologie expérimentale pour tester nos hypothèses de recherche.

5. Peut-on s'appuyer sur des instruments de mesure fiables et valides ?

L'examen des qualités psychométriques des instruments de mesure mobilisés pour cette recherche conduit à conclure à l'absence de problème de validité ou de fiabilité.

PARTIE III

EXPLIQUER

Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne

Analyse empirique des résultats

INTRODUCTION DE LA PARTIE III

Nous venons d'exposer, dans la partie II, le cadre conceptuel auquel nous avons abouti ainsi que la méthodologie choisie pour le tester, tout en nous assurant de la qualité des instruments de mesure qu'elle mobilise.

Nous abordons, dans cette troisième et dernière partie (intitulée « EXLIQUER »), l'aspect empirique de notre travail.
En cela, la principale question (d'ordre empirique) à laquelle nous souhaitons apporter une réponse est la suivante :

Les données que nous serons amenés à collecter nous permettent-elles de valider nos hypothèses de recherche et ainsi le modèle conceptuel présenté ?

Pour répondre à cette question majeure, trois expériences successives seront mises en œuvre, correspondant aux trois chapitres de cette partie III. Chacune d'elles se focalise plus particulièrement sur une variable clé du processus de réponse.

Nous réaliserons ainsi une première expérience sur la mesure de l'attitude du consommateur face à la communication de ses données personnelles (chapitre 10). Elle sera menée auprès d'un échantillon d'étudiants, par le biais d'un questionnaire papier.

L'expérience 2 (chapitre 11) sera réalisée en ligne, auprès de panélistes. Elle visera, pour l'essentiel, à évaluer les intentions comportementales de réponse.

Enfin, l'expérience 3 (chapitre 12) a fait l'objet d'un partenariat avec un site réellement implanté sur le web. Cette collaboration fut l'occasion de nous intéresser à la partie comportementale du processus : la réponse à la sollicitation. Plus de 240 000 clients « réels » seront ainsi sollicités pour participer à cette expérience « grandeur nature ».

Partie III – EXPLIQUER

Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne

Analyse empirique des résultats

CHAPITRE 10

EXPLIQUER L'ATTITUDE DU CONSOMMATEUR FACE À LA COMMUNICATION DE SES DONNÉES PERSONNELLES (ÉTUDE 3)

« There is general agreement that attitude represents a summary evaluation of a psychological object captured in such attribute dimensions as good-bad, harmful-beneficial, pleasant-unpleasant and likable-dislikable ».

Ajzen (2001)

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

Introduction du chapitre 10

La présentation du modèle conceptuel général de la recherche (chapitre 7), suivie de l'explication du plan expérimental en 3 phases (au chapitre 8) visant à tester l'ensemble des propositions et hypothèses émises, nous amène à réaliser une première expérience, centrée sur la partie gauche du modèle, de l'exposition de l'individu à la sollicitation (facteurs situationnels), à l'attitude de ce dernier face à la communication de ses données.

Cette expérience sera menée auprès d'un échantillon d'étudiants soumis à une situation de sollicitation de données personnelles, par le biais d'un formulaire Internet à remplir. Plusieurs facteurs seront alors manipulés afin de tester leur impact respectif sur l'évaluation que l'individu porte à la sollicitation.

La définition des objectifs assignés à cette expérimentation (phase 1) nous a conduits à choisir les facteurs à manipuler (phase 2), qui ont ensuite fait l'objet d'une réflexion approfondie et de nombreux pré-tests avant d'aboutir à une opérationnalisation[185] qui soit satisfaisante (phase 3). Le contexte de notre recherche reposant sur une sollicitation de données personnelles sur Internet, nous avons ensuite créé des pages reprenant le formulaire à remplir (phase 4), tout en étant particulièrement attentifs au réalisme de ce dernier.
Par la suite, nous avons élaboré puis pré-testé l'instrument de collecte (phase 5) qui a été administré (en format papier-crayon) auprès de plus de 230 étudiants (phase 6).
Après avoir vérifié l'ensemble des conditions nécessaires au test des hypothèses, conditions qui concernent notamment, la validité du plan d'expérience, la vérification des conditions expérimentales et l'établissement de la qualité psychométrique des instruments de mesure utilisés (phase 7), nous avons pu procéder à l'analyse des données.
L'analyse des résultats constitue la phase ultime (phase 8) de cette procédure aux termes de laquelle nous avons été en mesure d'indiquer (et de commenter) les hypothèses validées.

L'ensemble de la procédure suivie pour mener à bien cette première expérience est résumé à travers la figure suivante (cf. figure 10.1) :

[185] Il est difficile, voire impossible, de trouver le terme d'opérationnalisation dans un dictionnaire français. Il peut néanmoins être considéré comme un néologisme plus à même d'exprimer le sens qu'on veut lui donner (i.e. manière d'opérationnaliser une variable) que des expressions, plus correctes, comme celle de « mise en œuvre ». Il est, de plus, couramment employé par les chercheurs en gestion, notamment dans le domaine expérimental. Nous emploierons donc, et ce, de manière quelque peu abusive, ce terme dans la suite de ce document.

Figure 10.1 Phases suivies pour mener l'expérimentation (*)

(*) Nous occultons volontairement la sélection, le développement et la validité des échelles de mesure, nécessaires au test des hypothèses et inclus dans l'instrument de collecte, cette phase ayant été largement abordée dans le chapitre précédent (cf. chapitre ç).

1 Objectifs et apports de l'expérience 1

Dans cette première expérience, nous nous focaliserons sur la partie gauche de notre modèle à tester (cf. figure 10.2 ci-dessous), à savoir l'impact de l'exposition à la sollicitation (facteurs situationnels) sur l'évaluation de celle-ci (par le biais de variables perceptuelles) et l'effet de cette dernière sur l'attitude du consommateur face à la communication de ses données.

Figure 10.2 Modèle simplifié testé dans l'expérience 1 [186]

(*) Variables qui ne seront pas étudiées dans le cadre de cette première expérience

Le choix des facteurs situationnels et des variables perceptuelles étudiées dans ce cadre répond à une volonté de combler des manques identifiés au sein de la littérature et de tester certains liens pressentis à travers les résultats des études exploratoires (cf. chapitres 5 et 6).

Ainsi, de leur étude empirique menée auprès de consommateurs chinois, Zhang, Wang et Shen (2001) tirent les conclusions suivantes, sans avoir pu, pour autant, les vérifier :
- si le fait d'être client de l'entreprise sollicitant les données n'est pas une motivation pour encourager un individu à répondre, le fait de ne pas l'être pourrait constituer un frein majeur ;
- il serait préférable de ne pas inclure trop de questions dans le formulaire (sauf si elles sont indispensables à la connaissance du client) car celui-ci risque de quitter le site sans répondre.
L'impact possible de ces 2 facteurs (i.e. familiarité avec l'entreprise sollicitant les données et quantité de données demandées) sur le processus de réponse a été « confirmé » par les résultats de l'étude exploratoire descriptive présentée au chapitre 6.

Pour Caudill et Murphy (2000), le respect de la vie privée (dans son aspect informationnel) dépend essentiellement de deux dimensions : 1) la quantité et la profondeur des informations

[186] Le modèle complet sera présenté dans la suite de ce chapitre (cf. figure 10.3, paragraphe 3.1).

collectées, 2) le degré de connaissance de l'individu quant à la collecte et à l'utilisation de telles informations, auquel de nombreux auteurs ajoutent la dimension « choix ». Ces deux dernières dimensions (connaissance et choix) représentent d'ailleurs deux des principes à respecter par les entreprises souhaitant mettre en place une politique de traitement des données éthique et/ou réglementaire[187]. Or, lors de la navigation sur Internet, l'ensemble de ces dimensions peuvent être mises en péril, ce qui ne serait pas sans incidence sur la manière dont l'individu perçoit la sollicitation de ses données.
L'objectif de cette expérience sera donc de tester les conclusions formulées par ces auteurs qui ne les ont, jusqu'à présent, jamais vérifiées empiriquement.

Nous nous intéresserons donc ici à l'impact de trois facteurs situationnels majeurs[188] :

1) les mentions de RVP figurant sur le formulaire
2) la familiarité du consommateur avec l'entreprise sollicitant les données
3) la quantité de données demandées

Plus précisément, nous analyserons l'effet de ces facteurs sur trois des perceptions identifiées dans le cadre conceptuel (la confidentialité, la sensibilité et la pertinence perçues des données sollicitées). Nous verrons aussi l'impact de ces perceptions et de la préoccupation pour le respect de la vie privée sur l'attitude de l'individu face à la communication de ses données.

2 Le design expérimental

Le design proposé dans le cadre de cette première expérience est un plan factoriel complet de type 2 x 2 x 2 (trois facteurs manipulés [Mentions RVP, Familiarité et Quantité], chacun à deux niveaux), en inter-sujet[189] (chaque répondant n'est confronté qu'à un seul des 8 traitements possibles). Puisque le quatrième facteur présenté dans le cadre conceptuel (le contexte de sollicitation) n'est pas manipulé ici, nous choisissons de le fixer à un niveau identique pour tous les répondants, sous la forme d'un jeu-concours. Ce contexte offre en effet une occasion idéale de mettre les individus dans une situation de sollicitation de données, situation qui s'avère à la fois courante et réaliste.

[187] Si l'adoption de tels principes dépend du bon vouloir des entreprises outre-Atlantique, en Europe et notamment en France, il s'agit de principes légaux et donc obligatoires. Seule leur mise en œuvre (la manière de les appliquer concrètement) est laissée à l'appréciation des entreprises.
[188] La manière dont ces facteurs seront opérationnalisés sera développée au point 2.1
[189] En anglais « *between subjects* »

Nous exposerons, dans un premier temps, la manière dont les facteurs choisis ont été manipulés (2.1) avant d'évoquer les variables qui seront contrôlées (2.2).

2.1 Les facteurs manipulés

Comme indiqué précédemment, trois facteurs seront manipulés dans le cadre de cette première expérience : les mentions de RVP, la familiarité avec l'entreprise sollicitant les données et la quantité de données demandées. Nous présentons les choix opérés concernant le type et le niveau des manipulations (2.1.1) avant d'aborder l'opérationnalisation (2.1.2).

2.1.1 Type de manipulation et niveau des facteurs manipulés

Par principe, une expérimentation fait l'objet d'un scénario visant à mettre le répondant dans une situation propice à le faire réagir. Dans le cas présent, il s'agit donc de soumettre l'individu à une sollicitation de données, par le biais d'un formulaire à remplir sur Internet.
Un premier choix s'offrait alors à nous. Nous pouvions en effet mettre le répondant sur Internet, lui faire parcourir un (ou plusieurs) site(s) et le diriger vers une page contenant un formulaire sollicitant des données personnelles. Cette solution conduisait elle-même à un nouveau choix, celui d'un (ou de) site(s) réel(s) ou fictif(s). La solution du site réel a été rapidement écartée du fait de la difficulté à maîtriser le processus de navigation et à contrôler certaines variables exogènes. Le choix du site fictif, bien que séduisant, a également été rejeté car nous nous intéressons uniquement à la réaction du consommateur une fois celui-ci arrivé sur la page du formulaire. Faire naviguer le répondant sur un site, même fictif, faisait donc courir le risque d'allonger le processus (et donc le temps global de l'expérimentation), sans pouvoir contrôler pour autant certaines variables exogènes (comme la vitesse de connexion), sauf à prendre des dispositions particulières (interroger tous les répondants dans une même salle par exemple). De plus, se posait aussi la question de la manière de « diriger » le répondant sur la page proposée (quelles instructions donner ? comment être sûr qu'il s'y rendrait ?). Bien que réalisable, cette stratégie du site fictif demandait donc un ensemble de précautions (notamment d'ordre technique) qui ne justifiait pas, selon nous, l'investissement nécessaire à sa mise en œuvre, tout en posant, en outre, de multiples problèmes de contrôle.

Nous nous sommes donc tournés vers une solution, à la fois simple et « économique » qui répondait, en plus, parfaitement à l'objectif que nous nous étions fixé et qui consistait à ne créer que la page contenant le formulaire à remplir[190]. Il s'agissait alors de bâtir un scénario crédible amenant l'individu à s'imaginer être sur Internet en train de naviguer et d'arriver sur une page comprenant un formulaire à compléter, page qui leur serait fournie sous la forme d'une copie-écran et qui sera désignée ci-après sous le nom de « page formulaire ».

Deux types de manipulations sont couramment utilisés par les chercheurs adoptant une logique expérimentale. Il s'agit, d'une part, de la manipulation par évènement (absence ou présence d'un évènement ou d'un « signe » particulier) et, d'autre part, de la manipulation par instruction (phrase du scénario visant à indiquer au répondant dans quel cas il se trouve).
Dans le cadre de cette expérience, nous aurons ainsi recours à ces deux types de manipulations. Plus spécifiquement, les mentions de RVP et la quantité de données demandées feront l'objet d'une manipulation par évènement : ces deux facteurs seront concrètement manipulés par les différences (visuelles) apportées sur la page du formulaire (copie-écran) qui sera présentée aux répondants. A l'opposé, la familiarité sera manipulée par le biais d'instructions données aux répondants concernant le type d'entreprise les sollicitant (concrètement le nom du site sur lequel ils sont censés se trouver) et le type de relation qu'ils entretiennent avec elle (l'ancienneté de leur relation). Dans les paragraphes suivants, nous détaillons plus précisément l'ensemble de ces manipulations pour chacun des facteurs.

Une condition nécessaire à une « bonne » expérimentation est l'obtention de variance, que l'on cherche alors à expliquer par les facteurs manipulés. Kerlinger (1986) recommande de faire en sorte que les traitements soient les plus différents possibles, puisque cela donne une meilleure chance d'avoir de la variance. Nous avons cherché à respecter ce principe, tout en restant le plus réaliste possible. Nous nous sommes ainsi largement inspirés des formulaires « réels » trouvés sur Internet pour élaborer notre page, de manière à plonger les répondants dans une situation qui leur paraisse la plus proche possible de la réalité. L'objectif était d'obtenir des réponses qui correspondent au maximum à ce qu'aurait été leur réaction s'ils s'étaient réellement trouvés sur Internet face à ce formulaire à remplir. Lors de l'élaboration et de la mise en œuvre du design expérimental, nous avons donc été particulièrement attentifs au réalisme de la situation et à l'absence (au moins à la réduction) de biais inhérents à la

[190] Au total, 8 pages différentes ont été créées correspondant aux 8 traitements expérimentaux. Sont fournis en Annexe 10.1 deux exemples de questionnaires vierges tels que ceux administrés aux répondants dans le cadre de cette expérience et sur lesquels il est possible de visualiser ces pages écrans (en page 4 du questionnaire).

logique expérimentale. Plusieurs pré-tests ont ainsi été nécessaires pour aboutir à un scénario crédible et réaliste visant à mettre les répondants dans une situation de sollicitation de données personnelles, sans leur révéler pour autant ni le sujet précis de l'étude[191], ni les facteurs manipulés, ni les hypothèses sous-jacentes.

2.1.2 Facteurs manipulés et opérationnalisation

Nous présentons tour à tour les trois facteurs ayant fait l'objet d'une manipulation dans le cadre de cette expérience. Même si plusieurs pré-tests ont été nécessaires, nous insistons ici plus particulièrement sur l'opérationnalisation « finale » à laquelle nous avons abouti, tout en indiquant les difficultés auxquelles nous avons pu être confrontés pour chacun d'eux.

Facteur 1 : les mentions de RVP

Le premier facteur manipulé dans le cadre de cette expérience concerne la politique de respect de la vie privée mise en place par l'entreprise, en l'occurrence apposée sur son site Internet, par le biais de mentions. Cette politique fait référence à des règles qui, en France, sont encadrées par la Loi[192]. Celle-ci, tout en donnant un cadre général, s'abstient toutefois, le plus souvent, de préciser de quelle façon ces principes doivent être mis en œuvre. Les entreprises françaises, tout en se conformant à la législation, disposent donc de marges de manœuvre quant à la manière de les appliquer. Il suffit ainsi de surfer sur Internet pour constater à quel point il existe un écart dans les politiques de RVP mises en place par les différents sites.

Parmi les principes à respecter en matière de vie privée, trois d'entre eux sont particulièrement importants. Il s'agit des principes d'information, de choix et de sécurité. Cette importance, reconnue publiquement, est également confirmée par les recherches déjà menées sur le sujet, même si la plupart d'entre elles concernent le contexte américain, dont on a vu qu'il était différent du contexte européen et français.

Culnan (1995) souligne ainsi l'importance de la connaissance (ou de l'information) du consommateur. Pour l'augmenter, elle suggère que les entreprises mettent en place des notices (ou chartes) d'informations qui soient claires et accessibles. Le principe de choix (ou de

[191] Dans l'introduction, nous leur avons simplement indiqué qu'il s'agissait d'une étude portant sur les pratiques des consommateurs français sur Internet.
[192] Le chapitre 2 revient sur ces règles et sur la différence existant entre la vision européenne et américaine. A noter que ces lois s'appliquent aussi sur Internet, des spécificités inhérentes à ce média étant d'ailleurs à signaler.

consentement) concerne, quant à lui, la possibilité donnée au consommateur de décider de ce qui sera fait de ses données, une fois celles-ci divulguées. Comme le rappelle Jones (1991), pour que le consentement soit complet, il doit être « informé, intentionnel et volontaire ». Il ne peut donc y avoir consentement, au sens propre du terme, s'il n'y a pas d'information. Rappelons que, dans ce domaine comme dans beaucoup d'autres, le consentement n'est pas toujours explicite. Dans le champ commercial, les exemples sont légion où le consentement est plutôt implicite, par exemple sous la forme d'opt-out[193]. A noter que, sans réellement interdire cette pratique, la Directive Européenne souligne sa préférence pour le principe d'opt-in, qui implique un consentement explicite du consommateur. Là encore, une simple visite sur des sites européens montre que cette option est cependant loin de faire l'unanimité.

Ces deux dimensions d'information et de choix ont été utilisées dans de nombreuses recherches empiriques sur le respect de la vie privée. Nous pouvons citer, entre autres, les travaux de Milne et Boza (1999) ; Phelps, Nowak et Ferrell (2000) et ceux de Sheehan et Hoy (2000). Il y a donc une solide fondation dans la littérature pour l'usage de ce cadre afin d'investiguer l'état de la politique de respect de la vie privée de l'entreprise et/ou du site.

Nous y ajoutons un troisième aspect (la sécurité) qui, bien que présenté comme fondamental, n'en est pas moins souvent occulté. Une des raisons à cela est le fait que certains chercheurs considèrent la sécurité comme une dimension du respect de la vie privée tandis que d'autres prennent les 2 notions en compte séparément. Les études menées par Culnan pour le FTC[194] indiquent qu'on peut considérer la sécurité comme une partie de la politique de RVP, même si la sécurité liée au RVP n'est pas le seul aspect de la sécurité d'un site. La sécurité est, en outre, clairement considérée comme un des principes fondamentaux de la Directive Européenne de 1995. Miyazaki et Fernandez (2000) énumèrent 3 aspects majeurs concernant la sécurité d'un site : 1) la sécurisation des informations et des transactions ; 2) la garantie de sécurité contre la fraude aux cartes de crédit ; 3) l'existence d'options de paiement alternatives. Bien que notre sujet concerne majoritairement le premier aspect, d'autant que nous excluons d'étudier la fourniture de données personnelles dans le cadre d'un achat en ligne, nous ne pouvons totalement occulter le second. Nos entretiens qualitatifs prouvent en effet que les internautes français ont une version largement holistique du phénomène (le site est ou n'est pas sécurisé), sans se focaliser sur l'aspect informationnel ou financier concerné.

[193] Terme déjà défini au chapitre 2
[194] *Federal Trade Commission*, organisme chargé de réfléchir à la question de la protection de la vie privée des citoyens américains.

Un dernier point important à rappeler concerne la mise en œuvre concrète de la politique de RVP par un site Internet. Le plus souvent, celle-ci prend la forme d'une notice (ou charte)[195] qui est accessible par le biais d'un lien figurant sur la page d'accueil du site et/ou sur l'ensemble des pages sollicitant des données personnelles. Cette notice peut être distincte ou non de la page annonçant les conditions générales de vente et peut prendre des noms très divers comme « Nos engagements » (cf. Encadré 10.1 ci-après) ou « Politique de confidentialité ». Toutes les recherches académiques menées sur le sujet sont unanimes : peu de consommateurs font l'effort de consulter ces chartes qu'ils trouvent, le plus souvent, peu accessibles et peu lisibles.

Une autre manifestation de la politique de RVP d'un site se trouve dans les mentions de RVP[196] figurant généralement en bas des formulaires sollicitant des données personnelles, et dont un exemple est reproduit ci-après (cf. Encadré 10.2). Celles-ci reprennent d'une part le principe d'information (sous la forme de phrases expliquant le but de la collecte et l'utilisation qui sera faite des données) et d'autre part le principe de choix (question demandant à l'internaute s'il accepte que ses données soient transférées). Nos entretiens qualitatifs montrent que, si les internautes ne lisent pas les chartes de RVP, ils sont par contre particulièrement attentifs aux mentions de RVP figurant en bas du formulaire. Pour cette raison, nous faisons le choix de nous concentrer sur ces mentions. Parmi les mentions de RVP apparaissant sur un site, on trouve aussi des logos certifiant que l'entreprise respecte la vie privée de ses prospects/clients et/ou assure la sécurité des informations et des transactions. De nombreux logos existent[197], qu'ils soient de nature financière (tels Fia-Net) ou plus généraux et reprenant alors le nom d'organismes « certificateurs » (comme la Fevad en France[198]).

Vous trouverez à la page suivante, les encadrés 10.2 et 10.3 présentant les copies écrans mentionnées ci-dessus et tirées du site de l'opérateur de téléphonie mobile « SFR ».

[195] En anglais, *privacy notice*
[196] En anglais, *privacy statements*
[197] Nous ne débattrons pas ici sur les aspects spécifiques à chacun d'eux tout comme nous ne présenterons pas une typologie aboutie de ces logos. Nous renvoyons les lecteurs intéressés aux recherches spécifiques sur ce sujet dont celle de Bressolles (2003)
[198] Ce second type de logos est globalement peu implanté en France alors qu'ils sont très influents aux Etats-Unis (exemple : Truste, BBBonline, …)

Encadré 10.1 Copie écran de la page « Nos engagements » sur le site de SFR

Encadré 10.2 Copie écran du bas d'une page formulaire d'inscription sur le site de SFR

Dans le cadre de cette expérience, le facteur « Mentions de RVP » sera donc manipulé à travers les mentions relatives aux dimensions d'information et de choix placées en bas du formulaire et à travers l'absence ou la présence d'un logo de sécurité. De plus, 2 niveaux seront opposés : un niveau « faible », correspondant à des mentions de RVP limitées (absence d'informations, choix par défaut, absence de logo) et un niveau « élevé », correspondant à de nombreuses mentions (informations, choix par opt-in, présence d'un logo).
Comme expliqué précédemment, nous nous sommes largement inspirés des sites réels pour choisir concrètement les mentions et le logo à faire figurer sur la page du formulaire que nous avons nous-mêmes créée. Nous avons ainsi consulté plus de 35 sites dans le domaine du voyage en ligne[199] (aussi bien des entreprises à la notoriété très établie que de petites agences de voyages locales ayant créé leur propre site) sur lesquels nous avons systématiquement répertorié l'ensemble des éléments relatifs à la politique de respect de la vie privée (existence ou non d'une charte, mentions et logos figurant sur le bas des formulaires)[200]. De plus, de nombreux pré-tests (menés auprès d'experts et de consommateurs) ont été nécessaires pour valider ces choix et aboutir à une manipulation « satisfaisante » de ce facteur.

Ainsi, concernant le logo de « sécurité », nous avons volontairement choisi un logo « général » plutôt que de nature financière (comme Fia-Net) dans la mesure où la situation ne concernait aucunement un achat en ligne[201]. De même, pour la dimension « choix », nous avons opté pour un format oui/non plutôt que pour une simple case à cocher. L'opérationnalisation définitive de ce facteur est détaillée dans le tableau suivant et visualisable à travers les encadrés figurant ci-après et représentant les copies écran fournies aux répondants se situant respectivement dans les cas de mentions de RVP limitées ou nombreuses.

Tableau 10.1 Opérationnalisation des Mentions de Respect de la Vie Privée (Facteur 1)

Niveau	Opérationnalisation
Faible (limitées)	- pas d'information sur le but de la collecte et l'utilisation des données - option de choix par opt-out (cases oui / non, avec oui pré-coché) - absence de logo de « sécurité » sur la page du formulaire

[199] Notre idée initiale était de placer les répondants dans une situation de sollicitation de données personnelles sur un site de voyages. Les raisons ayant conduit à l'abandon de cette solution et au choix d'un autre secteur d'activité seront abordées dans le paragraphe présentant l'opérationnalisation du facteur « familiarité ». Ceci ne nous empêchait cependant pas de nous inspirer des pratiques des sites de voyages en matière de RVP, celles-ci étant, dans l'ensemble, peu différentes d'un secteur à l'autre.
[200] Un récapitulatif des résultats obtenus est disponible sur simple demande auprès de l'auteur.
[201] Des pré-tests qualitatifs réalisés avec le logo Fia-Net montrent que les répondants étaient déconcertés par la présence de ce logo alors qu'on ne leur demandait pas de fournir un numéro de carte de crédit.

Elevé (nombreuses)	- informations sur le but de la collecte et l'utilisation ultérieure des données - option de choix par opt-in (cases oui / non, pas pré-cochées) - présence du logo « site sécurisé » sur la page du formulaire

Encadré 10.3 Extrait du formulaire pour le traitement « Mentions de RVP limitées »

Encadré 10.4 Extrait du formulaire pour le traitement « Mentions de RVP nombreuses »

Facteur 2 : La familiarité avec l'entreprise sollicitant les données

Le second facteur manipulé dans le cadre de cette expérience concerne le niveau de familiarité existant entre le consommateur et l'entreprise le sollicitant. La familiarité se définit généralement en conjonction avec le degré d'expertise pour former le niveau de connaissance d'un individu envers une marque, un produit ou une entreprise. Dans ce contexte, la familiarité correspond au « nombre d'expériences liées au produit (à la marque ou à l'entreprise) accumulées par le consommateur » (Alba et Hutchinson 1987, p 411). Appliquée à l'entreprise, elle correspond ainsi à tous les contacts que le consommateur a déjà eus avec elle, que ce soit directement (transactions passées) ou indirectement (par le biais de la publicité ou du bouche à oreille par exemple). On peut aussi considérer, avec Siriex et Dubois (1999), qu'il s'agit du « poids de l'expérience passée avec la marque (l'entreprise) ».

De manière à obtenir un maximum de variance, nous choisissons alors d'opposer une entreprise avec laquelle l'internaute est très familier[202] (familiarité élevée) et une entreprise avec laquelle il n'est pas familier du tout (familiarité nulle), soit deux niveaux de familiarité. Demander au répondant de donner le nom d'une entreprise avec laquelle il est familier (même en limitant le choix à un secteur spécifique) fait courir le risque d'obtenir une grande diversité de réponses possibles. Pour limiter ce risque, nous choisissons alors de sélectionner un secteur oligopolistique, avec un nombre très limité de fournisseurs (si possible pas plus de trois)[203]. Rapidement, le choix du secteur de la téléphonie mobile s'est imposé. Il offrait en effet l'avantage d'être limité à trois « gros » opérateurs et d'être très impliquant pour la cible de répondants visée (des étudiants). Nous limitions ainsi le risque de devoir éliminer des répondants n'ayant pas de fournisseur dans ce domaine. De plus, il s'agit d'un secteur très important économiquement et pour lequel les intervenants sont particulièrement soucieux d'entretenir des relations privilégiées avec leurs clients, ce qui implique alors de multiples sollicitations de données personnelles à diverses occasions possibles (changement de forfaits, organisation de jeux-concours, adhésion à un club de fidélité, …).

En ce qui concerne la familiarité élevée, il s'agissait donc de mettre le répondant dans une situation de sollicitation de données personnelles par le biais d'un jeu-concours organisé par son opérateur actuel de téléphonie mobile (son opérateur principal s'il en a plusieurs).

Dans le cas de la familiarité nulle, nous avons simulé l'arrivée sur le marché d'un nouvel opérateur, inconnu jusqu'ici, et ayant donc un nom fictif (« Mobile Telecom »). Nous avons alors demandé au répondant d'imaginer qu'en cherchant sur Internet des informations sur les tarifs et les forfaits de téléphonie mobile, il se soit retrouvé sur le site de ce nouvel opérateur, lequel mettait en évidence, sur sa page d'accueil, la possibilité de participer à un grand jeu-concours permettant de gagner de nombreux lots (dont un séjour au Kenya[204]).

La manipulation par instructions (utilisée pour ce facteur) s'est avérée délicate à mettre en œuvre. Les pré-tests montrent en effet que ce type de manipulation n'est pas toujours très impliquant. De plus, les répondants qui ne lisent pas attentivement le scénario peuvent ne pas intégrer complètement la situation présentée. Pour réduire ce biais (et ce, suite à plusieurs pré-tests), nous avons alors décidé de multiplier les occasions d'opérationnaliser le facteur, au sein du questionnaire. Outre le scénario indiquant si les répondants se trouvaient sur le site de leur opérateur ou sur celui d'un nouvel opérateur (familiarité élevée vs nulle)[205], nous avons

[202] Nous parlons bien ici de familiarité avec l'entreprise et pas forcément de familiarité avec le site.
[203] Nous renoncions alors à étudier le cas d'un site de voyage, étant donné le nombre élevé de sites de ce type.
[204] L'explication du concours et les lots en jeu étaient identiques d'un cas à l'autre (familiarité nulle vs élevée).
[205] Les phrases indiquant le type d'opérateur étaient mises en évidence par des majuscules et le caractère gras et souligné.

apposé le logo de l'opérateur en question au niveau du formulaire, afin d'ajouter au texte un caractère visuel[206], censé être plus « marquant ».

L'opérationnalisation définitive de ce facteur est détaillée dans le tableau 10.2 et visualisable à travers les Encadrés 10.5 et 10.6 représentant la page formulaire fournie aux répondants se situant respectivement dans les cas de familiarité nulle ou élevée. Au final, les analyses réalisées *a posteriori* (présentées au point 5.2) prouvent que la manipulation a bien été effective et que les répondants affectés à chaque cas ont bien intégré la situation proposée.

Tableau 10.2 Opérationnalisation de la Familiarité avec l'entreprise sollicitante (Facteur 2)

Niveau	Opérationnalisation
Nulle	- phrase du scénario indiquant une sollicitation sur le site d'un nouvel opérateur - répétition du terme « nouvel opérateur » et de son nom « Mobile Telecom (5 fois) - présentation des principales caractéristiques de ce nouvel opérateur : « vous n'en avez jamais entendu parler auparavant » ; « aucun client en France » et présentation du logo « Mobile Telecom » - présence du logo « Mobile Telecom » sur la page formulaire
Elevée	- phrase du scénario indiquant une sollicitation sur le site de son opérateur actuel - répétition du terme « votre opérateur » (5 fois dans le questionnaire) - obligation de cocher le logo de son opérateur au-dessus de la page formulaire

Encadré 10.5 Extrait du formulaire pour le traitement « Familiarité nulle »

Encadré 10.6 Extrait du formulaire pour le traitement « Familiarité élevée »

Cochez le logo correspondant à VOTRE OPERATEUR **Puis COMPLETEZ le formulaire qui suit**	orange ☐	SFR ☐	Bouygues Telecom ☐

[206] Dans le cas de la familiarité élevée, les logos correspondant aux trois principaux opérateurs étaient présentés, le répondant devant cocher celui correspondant au sien. Les questionnaires pour lesquels ce logo n'était pas coché ont d'ailleurs été supprimés par crainte d'une manipulation non effective. Pour la familiarité nulle, un logo simple (correspondant à l'opérateur fictif « Mobile Télécom ») créé par nos soins et présenté dans le texte était repris au niveau du formulaire.

Facteur 3 : La quantité de données sollicitées

Le troisième et dernier facteur manipulé dans le cadre de cette expérience concerne la quantité de données sollicitées. Cette dernière correspond alors au nombre de questions posées (ou de champs à compléter) sur le formulaire. Deux niveaux seront là aussi opposés : une quantité faible et une quantité élevée. Pour choisir le nombre et le type de données à faire figurer dans chaque cas, nous nous sommes inspirés à la fois de la littérature existante et de la réalité (ce que font couramment les sites à ce sujet). Ainsi, Hui et al. (2004) manipulent la quantité de données en modifiant la taille du formulaire à remplir (entre 4 et 23 questions).

D'après notre étude réalisée auprès de 35 sites de voyages, trois grands types de données sont couramment sollicités sur Internet, et notamment dans le cadre de jeux-concours :

- des données d'identification (données permettant d'identifier la personne comme son nom, son adresse postale, son email)
- des données sociodémographiques (situation familiale, profession, niveau d'études, ..)
- des données portant sur les habitudes et préférences comme les hobbies, le style de vie, le comportement d'achat, …

Concernant le nombre de données demandées, il ressort de nos analyses qu'il est de 8 en moyenne et de 10 dans le cadre des jeux-concours organisés en ligne, avec une fourchette allant de 3 ou 4 champs à compléter au minimum jusqu'à plus de 20.

De plus, nous nous sommes également inspirés des pratiques réelles des trois opérateurs de téléphonie mobile (Orange, SFR et Bouygues Télécom) présents sur le marché. Ainsi, à l'époque de nos réflexions sur l'opérationnalisation de ce facteur, deux de ces sites organisaient un jeu-concours. SFR organisait un jeu proposant de gagner un voyage et pour lequel un formulaire en 2 parties était proposé, la première portant sur des données d'identification (9 champs à compléter) et la seconde sur des données sociodémographiques et de préférence (7 questions posées). Nous avons alors emprunté certaines idées de ce formulaire, allant jusqu'à reprendre certaines des questions y figurant, afin de donner un caractère réel à notre manipulation.

Au final, nous avons choisi d'opposer un formulaire court (quantité faible) comprenant 5 données à fournir, toutes d'identification (nom, adresse, code postal, ville et email) à un formulaire long (quantité élevée) comprenant 20 champs à compléter et divisé en 3 parties (9 données d'identification, 5 données sociodémographiques et 5 données de préférence plus une

donnée financière correspondant au montant du revenu annuel). A noter que le formulaire court ne comprend aucune donnée facultative (ensemble des champs obligatoires, identifiés par une étoile *) tandis que le formulaire long comprend 3 données facultatives (une de chaque catégorie, soit la date de naissance, le nombre d'enfants et le revenu annuel). Ces manipulations ont été validées par de nombreux pré-tests menés auprès de professionnels et de consommateurs. Soulignons pour finir que, dans le cadre de cette première expérience, nous demandions aux répondants de remplir ce formulaire, et ceci, pour deux raisons principales. Tout d'abord, les pré-tests montrent que lorsque les répondants doivent simplement regarder le formulaire, certains parcourent très rapidement (voire le regardent à peine), ce qui réduit l'impact et l'efficacité de la manipulation. Ils sont en effet, malgré toutes les précautions que nous pouvons prendre, moins impliqués qu'ils ne le seraient s'ils étaient réellement en ligne face à ce type de formulaire pour lequel ils prendraient le temps de lire attentivement chaque champ à compléter. Ensuite, cela permet de les faire répondre aux questions qui suivent (et notamment à nos échelles de mesure) en meilleure connaissance de cause. Au final, nous présentons dans le tableau et dans les 2 encadrés qui suivent (cf. Encadrés 10.7 et 10.8) l'opérationnalisation du facteur quantité à laquelle nous avons abouti.

Tableau 10.3 Opérationnalisation de la Quantité de données demandées (Facteur 3)

Niveau	Opérationnalisation
Faible	5 champs à remplir : nom, adresse, code postal, ville, email
Elevé	20 champs à remplir répartis en 3 catégories : - « vos coordonnées » (correspondant aux données d'identification) - « votre profil » (correspondant aux données sociodémographiques) - « vos habitudes et préférences »

Encadré 10.7 Extrait du formulaire pour le traitement « Quantité faible »

Encadré 10.8 Extrait du formulaire pour le traitement « Quantité élevée »

Pour résumer, les huit traitements auxquels on aboutit sont présentés dans le tableau suivant :

Tableau 10.4 Le design expérimental en plan factoriel complet

Familiarité				Nulle		Elevée	
Quantité	Faible	Mentions RVP	Limitées	①	« F1Q1M1 »[207]	⑤	« F2Q1M1 »
			Nombreuses	②	« F1Q1M2 »	⑥	« F2Q1M2 »
	Elevée		Limitées	③	« F1Q2M1 »	⑦	« F2Q2M1 »
			Nombreuses	④	« F1Q2M2 »	⑧	« F2Q2M2 »

Nous fournissons en Annexe 10.1 et 10.1 bis[208] deux exemples de questionnaires correspondant respectivement aux traitements 1 et 8 et permettant de visualiser, dans leur ensemble, toutes les manipulations opérées sur ces facteurs.

[207] Se lit : F1 pour Familiarité niveau 1 (nulle), Q1 pour Quantité niveau 1 (faible), M1 pour Mentions RVP niveau 1 (limitées)

[208] L'Annexe 10.1 bis ne comporte que des extraits, correspondant aux pages où les facteurs sont manipulés.

Nous précisons, pour finir, que chacun des formulaires dispose du même design et de la même charte graphique, de manière à créer une certaine harmonie et à éviter tout biais lié à une perception différente de ces pages par les individus interrogés. Un pré-test mené à cet effet confirme l'absence de différence selon le type de traitement affecté.

Nous exposons à présent les variables qui seront contrôlées dans le cadre de cette expérience.

2.2 Les variables contrôlées

Nous détaillons ici le type de variable à contrôler (2.2.1) et le type de contrôle réalisé (2.2.2).

2.2.1 Le type de variable à contrôler

Il est important, dans une logique expérimentale, d'identifier toute variable externe susceptible d'interférer et donc de biaiser les résultats. Plusieurs variables externes, identifiées à travers l'étude de la littérature et nos entretiens qualitatifs seront ainsi contrôlées dans le cadre de cette première expérience. Il s'agit de variables, autres que les facteurs manipulés, susceptibles d'avoir une influence directe ou indirecte sur les variables médiatrices ou dépendantes étudiées ici, en l'occurrence sur l'évaluation de la sollicitation et/ou sur l'attitude face à la communication des données.

Dans le cadre de cette première expérience, ces variables de contrôle concernent ainsi l'origine sociodémographique du répondant, son expérience et utilisation du média Internet et son niveau d'implication vis-à-vis de la catégorie de produits/services concernée, en l'occurrence la téléphonie mobile. Toutefois, dans la mesure où nous interrogerons une population homogène sur certaines données sociodémographiques (des étudiants), notamment concernant l'âge (20 à 25 ans), la CSP et le niveau d'études (Master), nous ne contrôlerons que l'effet du sexe. Nous ne contrôlerons pas non plus, ni le comportement habituel de réponse ni la qualité perçue de la relation, leur influence se faisant plutôt ressentir au niveau des intentions comportementales et du comportement, variables non étudiées ici (mais l'étant dans les expériences 2 et 3).

Nous présentons maintenant le type de contrôle opéré sur ces variables.

2.2.2 Le type de contrôle

Les variables exogènes peuvent être contrôlées de différentes façons (Evrard et al., 2003). Classiquement, deux types de contrôles seront utilisés dans le cadre de cette expérience :

- une affectation aléatoire des individus interrogés aux différents traitements, qui permet une répartition *a priori* homogène des influences externes sur les unités-tests ;
- un contrôle statistique *a posteriori* sur l'ensemble des variables identifiées ci-dessus (sexe, expérience Internet et implication) par introduction dans les analyses en tant que co-variables.

Nous détaillons, dans le tableau suivant, les instruments de mesure adoptés pour chacune des variables exogènes contrôlées, après pré-tests auprès d'experts et de consommateurs.

Tableau 10.5 Instruments de mesure utilisés pour les variables externes contrôlées

Variables	Items utilisés	Types de mesure
Implication envers la catégorie de produits/services (la téléphonie mobile)	Echelle PIA adaptée de Strazzieri (1994) (3 items)	Likert en 7 points (de 1 : pas du tout d'accord à 7 : tout à fait d'accord)[209]
Sexe	H / F (1 item)	Variable nominale (binaire)
Expérience d'Internet	Ancienneté d'utilisation (1 item)	Ordinale à 3 modalités[210] (1 : moins de 2 ans, 2 : entre 2 et 5 ans ; 3 : plus de 5 ans)[211]
Utilisation d'Internet	- fréquence de consultation de la messagerie[212] (1 item) - fréquence de surf sur le web (1 item)	Ordinale à 4 modalités (de 1 : plusieurs fois par jour à 4 : moins souvent)[213]

De plus, nous rappelons que nous « contrôlons » également le quatrième facteur non manipulé dans cette expérience (le contexte de sollicitation) par fixation à un niveau défini *a priori* et identique à tous les répondants : le contexte d'une sollicitation par le biais d'un jeu-concours.

[209] La formulation de ces items et la vérification de la validité de cette échelle est présentée au chapitre précédent
[210] Une quatrième modalité intitulée « jamais » servait uniquement à repérer les répondants n'ayant jamais utilisé Internet et dont les questionnaires seraient ensuite éliminés
[211] Bien que nous perdions en précision par rapport au fait de mettre une question ouverte, les pré-tests montrent que les répondants préfèrent une question fermée. De plus, ce type d'échelle a déjà été utilisé dans la littérature puisqu'elle correspond à une classification « classique » des internautes en « débutants, familiers et experts ».
[212] Nous conservons, par précaution, la mesure de la consultation de la messagerie, malgré les critiques dont elle fait parfois l'objet dans la littérature, présentées dans le chapitre 8.
[213] Là encore, les pré-tests montrent que les répondants répondent plus facilement à ce type de format, qu'à une question portant sur le nombre de fois où ils consultent en un mois, par exemple.

Présentons maintenant le modèle ainsi que les hypothèses que nous serons amenés à tester.

3 Présentation du modèle à tester et des hypothèses de recherche

3.1 Présentation du modèle à tester

Rappelons que le modèle « complet » présenté au chapitre 8 (cf. figure 8.2) est en fait testé en trois parties, correspondant aux trois expériences menées successivement.

Dans cette première expérience, l'objectif est de mesurer l'impact de l'exposition à la sollicitation - par le biais de trois facteurs manipulés (Mentions de RVP, Familiarité avec l'entreprise et Quantité de données demandées) - sur l'évaluation de celle-ci (réduite à 3 des 4 perceptions identifiées au départ[214]). L'effet de cette évaluation et de la préoccupation pour le RVP sur l'attitude face à la communication des données est également étudié.

Nous précisons que chaque facteur manipulé au sein du design expérimental sera identifié par un numéro correspondant à une déclinaison du chiffre affecté au bloc « Exposition à la sollicitation » (i.e. le chiffre « 1 ») dans le modèle général de la recherche (cf. figure 7.6). Ainsi, le premier facteur manipulé dans le cadre de cette expérience (les Mentions de RVP) sera nommé « 1.1 », le second (la Familiarité avec l'entreprise) « 1.2 » et le troisième (la Quantité de données demandées) « 1.3 ». Le dernier facteur, non manipulé dans le cadre de l'expérience 1 (i.e. le Contexte de sollicitation), portera, quant à lui, le numéro « 1.4 ».

Le modèle à tester dans l'expérience 1 se présente donc comme suit (cf. figure 10.3).

Ce modèle comprend plusieurs hypothèses à tester qui ont, soit déjà été formulées dans le chapitre 7 (cas des hypothèses H2a, b et c et H3e), soit seront formulées dans la section qui suit (cas des hypothèses H1.1, H1.2 et H1.3, correspondant à la proposition P1 du chapitre 7).

[214] Deux raisons majeures expliquent ce choix. Une première échelle de valeur perçue (correspondant à la quatrième perception) avait été proposée initialement, en se fondant sur les échelles d'équité. Devant la faible validité de cette échelle (due notamment à l'anormalité des distributions), nous avons dû renoncer à l'utiliser. La nouvelle échelle de valeur perçue développée par nos soins comprenant 10 items, nous avons choisi de l'introduire prioritairement dans la seconde expérience afin de limiter la longueur du questionnaire de l'expérience 1, qui demandait déjà presque 15 minutes pour y répondre.

Figure 10.3 Modèle à tester dans le cadre de l'expérience 1

Nous présentons ci-après les hypothèses de recherche spécifiques à l'expérience 1.

3.2 Formulation des hypothèses de recherche

Nous formulons ici les hypothèses H1.1, H1.2 et H1.3, correspondant à la proposition P1 du chapitre 7, relative à l'effet de l'exposition à la sollicitation (facteurs situationnels manipulés) sur l'évaluation de celle-ci (confidentialité, sensibilité, pertinence perçues). Nous ne revenons pas sur les hypothèses H2a, b et c et H3e qui ont déjà été formulées et justifiées au chapitre 7.
Nous présentons tour à tour les hypothèses relatives à :
- l'influence des mentions RVP sur la confidentialité et la sensibilité perçues (H1.1 a et b[215])
- l'influence de la familiarité sur la confidentialité et la pertinence perçues (H1.2 a et c)
- l'influence de la quantité sur la sensibilité et la pertinence perçues (H1.3 b et c)

Nous avons volontairement limité les hypothèses relatives à l'influence de l'exposition à la sollicitation (3 facteurs) sur l'évaluation de celle-ci (3 perceptions) aux 6 proposées (sur les 9 possibles) pour deux raisons majeures. D'une part, nous ne pensons pas que les autres hypothèses soient « réalistes » et, d'autre part, nous manquons de matériau pour les justifier. Enfin, bien que nous ne formulions pas d'hypothèse spécifique à ce sujet, nous soupçonnons l'existence possible d'effets d'interaction entre les facteurs manipulés. Ainsi, d'après les résultats issus de la littérature, la familiarité pourrait agir en interaction avec les deux autres facteurs (mentions de RVP et quantité). Chellappa (2001) indique par exemple que les sites réputés pourraient être mieux perçus que les sites à la réputation limitée, même s'ils ne mettent pas en œuvre une politique de RVP très protectrice (mentions limitées), du fait de la confiance déjà établie avec le prospect/client.

3.2.1 L'influence des mentions de RVP sur la confidentialité et la sensibilité perçues

Les politiques de respect de la vie privée adressent la partie valeur attendue (*Expectancy*) de la théorie de la motivation c'est-à-dire l'attente selon laquelle une action (telle que le fait de fournir ses données) sera suivie du résultat attendu (que les données restent confidentielles). Selon Vroom (1964), le degré avec lequel les pratiques des entreprises sont formalisées dans

[215] Pour rappel, le nom de ces hypothèses se justifie d'une part par le numéro attribué à chaque facteur manipulé (par exemple 1.1 pour le premier facteur nommé « Mentions de RVP ») et d'autre part par la lettre affectée à chaque variable médiatrice ou dépendante du modèle (par exemple la lettre a pour la « confidentialité perçue »).

des politiques écrites a un impact sur les perceptions des individus. En particulier, cela accroît la confiance en leur permettant de prendre des décisions « informées » (Culnan 1999). En postant une politique de RVP sur leur site, les e-commerçants indiquent donc aux consommateurs quel résultat ils peuvent attendre de leur dévoilement. De plus, en augmentant la probabilité d'un résultat bénéfique, cela les encourage alors à participer.

L'absence de politique de respect de la vie privée est le facteur ayant le plus d'influence sur la confiance/méfiance du consommateur envers l'e-commerçant. En particulier, l'absence d'information et de choix engendre chez l'individu une sensation de perte de contrôle quant à l'utilisation ultérieure de ses données (Culnan 1995). A l'opposé, le fait pour une entreprise d'annoncer ses pratiques en matière de traitement de données personnelles et de laisser au consommateur le choix quant au contrôle ultérieur de ses informations permet de développer la confiance et d'encourager l'individu à divulguer les renseignements demandés (Dinev et Hart 2002). Plus précisément, Culnan et Armstrong (1999) montrent que le recours à des pratiques éthiques[216] permet de réduire les craintes des consommateurs en matière de confidentialité en leur donnant davantage de pouvoir et de contrôle. Or, les individus percevant avoir du contrôle ont une moins forte perception de vulnérabilité, surtout concernant l'utilisation secondaire de leurs données (Fusilier et Hoyer 1980). Ce constat est confirmé par Childers et Skinner (1996) qui concluent que les facteurs signalant que les données sont collectées avec intégrité et professionnalisme servent à accroître la coopération en établissant un sentiment de confiance selon lequel la confidentialité serait respectée.

L'ensemble de ces considérations nous amènent à formuler l'hypothèse suivante :

Hypothèse 1.1 (a) : *Plus les mentions de respect de la vie privée (RVP) sont nombreuses (vs limitées), plus la confidentialité des données sollicitées(a) est perçue comme élevée (vs faible)*

Lors de la fourniture de données personnelles, le consommateur est particulièrement attentif aux conséquences possibles pour lui (Gandy 1993). Ce constat, identifié dans la littérature, ressort également de nos entretiens individuels menés auprès de consommateurs français. Dès lors, l'absence d'information relative au but de la collecte et à ses conséquences (i.e. des mentions RVP limitées) conduit généralement les individus à produire leur propre version des faits, qui est alors, le plus souvent, défavorable, sous la forme de perceptions négatives. Ainsi, Stone et Stone (1990) indiquent que le fait de ne pas connaître l'utilisation ultérieure de ses données (et de ne pas pouvoir la maîtriser) augmente la vulnérabilité de l'individu en diminuant sa capacité à formuler des cognitions réalistes et appropriées. Pour ces auteurs,

[216] Par exemple sous la forme de *Fair Information Pinciples*, principes « éthiques » mis en place aux Etats-Unis par le FTC (cf. Chapitre 2)

plusieurs facteurs sont susceptibles d'influencer le fait qu'un individu considère que les informations qui lui sont demandées sont privées et ne concernent que lui-même (définition stricte du concept de sensibilité perçue des données). Ces facteurs concerneraient notamment la transparence des procédures. Ainsi, plus la sollicitation est jugée transparente (grâce à des mentions de RVP nombreuses), moins la sollicitation est jugée envahissante et donc moins les données sollicitées sont considérées comme sensibles.

Ceci nous amène à formuler l'hypothèse suivante :

Hypothèse 1.1 (b) : *Plus les mentions de respect de la vie privée (RVP) sont nombreuses (vs limitées), moins (vs plus) la sensibilité des données sollicitées(b) est perçue comme élevée.*

3.2.2 L'influence de la familiarité sur la confidentialité et la pertinence perçues

L'expérience passée avec l'entreprise sollicitant des données serait un facteur déterminant dans l'acceptation de fournir les informations. Ainsi, Hine et Eve (1998) indiquent que, lors d'une sollicitation, la perception d'invasion de sa vie privée dépendrait des relations passées entre l'entreprise et le consommateur. De même, Culnan et Armstrong (1999) montrent que le fait de s'appuyer sur une relation existante permet d'éviter que la sollicitation soit perçue comme envahissante. Enfin, Jourard (1966) tout comme Stone et al. (1983) soulignent que la cible du dévoilement et la relation entretenue avec elle influencent les perceptions de violation de la vie privée. Les consommateurs incluraient donc la réputation du commerçant (facteur non étudié ici) et ses expériences passées avec lui (la familiarité) pour évaluer les risques en matière de respect de la vie privée, surtout en termes de confidentialité. Milne et Boza (1999) indiquent que l'expérience passée avec l'entreprise est une des raisons amenant l'individu sollicité à lui faire confiance quant à sa manière d'utiliser les données qu'il aura bien voulu lui délivrer. Ceci nous amène à formuler l'hypothèse suivante :

Hypothèse 1.2 (a) : *Plus la familiarité avec l'entreprise sollicitant les données est élevée (vs nulle), plus la confidentialité des données sollicitées (a) est perçue comme élevée (vs faible).*

La familiarité aurait également une influence sur la pertinence perçue des données sollicitées. Ainsi, Wang et Petrison (1993) soulignent que les consommateurs sont susceptibles d'accepter certaines sollicitations, par ailleurs jugées envahissantes et peu pertinentes, quand elles viennent d'entreprises avec lesquelles ils sont familiers. Ce constat a été maintes fois démontré dans le domaine du marketing direct. Rogers (1996), par exemple, note que les consommateurs sont moins hostiles aux opérations de marketing direct organisées par des

entreprises avec lesquelles ils ont déjà correspondu par le passé que par des entreprises avec lesquelles ils ne sont pas familiers. Appliqué au niveau de la sollicitation des données (plutôt qu'à leur utilisation), ce constat amène alors à formuler l'hypothèse suivante :

Hypothèse 1.2 (c) : *Plus la familiarité avec l'entreprise sollicitant les données est élevée (vs nulle), plus la pertinence des données sollicitées (c) est perçue comme élevée (vs faible).*

3.2.3 L'influence de la quantité sur la sensibilité et la pertinence perçues

Le type et le nombre d'informations demandées influenceraient les croyances des individus relatives aux pratiques des entreprises en matière de sollicitation et d'utilisation ultérieure des données. L'accroissement de la quantité d'informations demandées engendrerait ainsi une perception élevée de vulnérabilité. La raison pour laquelle cela aurait un effet important sur les cognitions serait liée au fait que cela influence, à la fois, le type d'inférences pouvant être réalisées par la suite sur l'individu, l'étendue avec laquelle il serait stigmatisé par le dévoilement de telles informations et le type d'actions qui pourront alors être dirigées contre lui (Stone et Stone 1990). Une quantité élevée de données demandées aurait donc un effet sur la sensibilité perçue de ces données, telle que formulée à travers l'hypothèse suivante :

Hypothèse 1.3 (b) : *Plus la quantité de données demandées est élevée (vs faible), plus (vs moins) leur niveau de sensibilité (b) est perçu comme élevé.*

Woodman et al. (1982) prouvent que les consommateurs utilisent un critère de pertinence pour juger si telle information peut être utilisée dans un but particulier. De la même manière, on peut penser que les individus auront recours à ce critère au moment de la sollicitation de leurs données personnelles. Dès lors, plus le nombre de données demandées augmente, plus le consommateur aura tendance à se poser la question des raisons amenant l'entreprise à solliciter de telles informations, ce qui n'est pas forcément le cas si elle se contente de collecter un nombre réduit de renseignements. Plus le nombre de données demandées augmente, plus l'entreprise court donc le risque que le consommateur juge leur collecte non pertinente par rapport au but recherché. Ceci nous amène à formuler l'hypothèse suivante :

Hypothèse 1.3 (c) : *Plus la quantité de données demandées est élevée (vs faible), moins (vs plus) la pertinence des données sollicitées (c) est perçue comme élevée.*

En résumé, voici l'ensemble des hypothèses à tester dans le cadre de l'expérience 1 :

Tableau 10.6 Récapitulatif des hypothèses à tester dans l'expérience 1

H 1.1 (a) (1)*	Plus les mentions de respect de la vie privée (RVP) sont nombreuses (vs limitées), plus la confidentialité des données sollicitées est perçue comme élevée (vs. faible)
H 1.1 (b) (2)	Plus les mentions de respect de la vie privée (RVP) sont nombreuses (vs limitées), moins (vs plus) la sensibilité des données sollicitées est perçue comme élevée
H 1.2 (a) (3)	Plus la familiarité avec l'entreprise sollicitant les données est élevée (vs. nulle), plus la confidentialité des données sollicitées est perçue comme élevée (vs. faible)
H 1.2 (c) (4)	Plus la familiarité avec l'entreprise sollicitant les données est élevée (vs. nulle), plus la pertinence des données sollicitées est perçue comme élevée (vs faible)
H 1.3 (b) (5)	Plus la quantité de données demandées est élevée (vs. faible), plus (vs moins) leur niveau de sensibilité est perçu comme élevé
H 1.3 (c) (6)	Plus la quantité de données demandées est élevée (vs. faible), moins (vs plus) la pertinence des données sollicitées est perçue comme élevée
H 2 (a) (7)	Plus la confidentialité des données sollicitées est perçue comme élevée (vs. faible), plus l'attitude face à leur communication est favorable (vs défavorable)
H 2 (b) (8)	Plus la sensibilité des données sollicitées est perçue comme élevée (vs. faible), moins l'attitude face à leur communication est favorable (vs défavorable)
H 2 (c) (9)	Plus la pertinence des données sollicitées est perçue comme élevée (vs faible), plus l'attitude face à leur communication est favorable (vs défavorable)
H 3 (e) (10)	Plus la préoccupation pour le RVP est élevée (vs faible), moins l'attitude face à la communication des données sollicitées est favorable (vs défavorable)

(*) Pour faciliter la reconnaissance des hypothèses, notamment dans la Discussion, nous les numérotons parallèlement de façon « classique »

4 Recueil des données et caractéristiques des répondants

4.1 L'instrument de collecte et le recueil des données

L'instrument de mesure utilisé (cf. Annexe 10.1) comprend 7 parties, intitulées comme suit :

- « bonjour à toutes et à tous » : introduction visant à présenter l'objectif de l'étude, à en expliquer le déroulement et à garantir au répondant la confidentialité de ses réponses
- « vous et la téléphonie mobile » : questions visant à identifier l'utilisation de la téléphonie mobile par le répondant (nom de l'opérateur, niveau d'implication, …)
- « imaginez la situation suivante » : scénario visant à immerger le répondant dans la situation de sollicitation de données personnelles lui ayant été affectée (parmi les huit possibles) et copie-écran du formulaire de sollicitation que le répondant devait compléter

- « vos réactions face à ce formulaire » : questions visant à mesurer les réactions de l'individu face au formulaire venant d'être complété (attitude, intentions comportementales, …)
- « votre impression face à ce formulaire » : questions visant à évaluer les perceptions relatives à la sollicitation des données (confidentialité, sensibilité et pertinence perçues)
- « quelques questions d'ordre général » : questions visant à collecter des données générales sur le répondant (préoccupation pour le respect de la vie privée notamment)
- « avant de terminer » : questions visant à évaluer l'effectivité des manipulations réalisées (*manipulation checks*)
- « pour mieux vous connaître » : questions visant à identifier le profil du répondant d'un point de vue sociodémographique (sexe) et expérientiel (ancienneté et utilisation d'Internet)

L'organisation des différentes parties se veut la plus cohérente possible, en accord avec les hypothèses à tester et les commentaires récoltés lors des pré-tests. Il s'agissait aussi d'éviter au maximum les biais inhérents à ce type d'instrument (effet de halo et de contamination notamment). Nous avons ainsi veillé à mesurer systématiquement les variables dépendantes avant les variables médiatrices, de manière à ne pas contaminer les résultats. De même, suite aux recommandations de spécialistes de la logique expérimentale[217], nous avons vérifié l'effectivité de nos manipulations à la fin, pour ne pas induire de biais dans les réponses.

Ce questionnaire de type « papier crayon » a été rédigé pour répondre à une logique d'auto-administration. La page d'introduction est ainsi censée permettre de renseigner au maximum le répondant sur ce qui est attendu de lui. Il a été administré auprès de 5 classes d'étudiants en Master (première et deuxième année) de l'Université d'Angers, par un enquêteur unique. L'administration a duré moins de 10 jours, correspondant aux dates auxquelles les cours magistraux et les TD étaient prévus avec ces différentes classes[218]. Les questionnaires ont été affectés de manière entièrement aléatoire aux répondants, qui disposaient en outre de la possibilité de décliner toute participation[219]. Le questionnaire était systématiquement rempli en classe, au début ou à la fin du cours (un créneau de 15 minutes était réservé à cet usage), sans que les étudiants en soient pour autant informés à l'avance. 270 questionnaires complétés ont ainsi été obtenus, auquel il a fallu soustraire 27 questionnaires « invalides » pour les raisons suivantes : non utilisation d'un téléphone portable, opérateur autre que les trois

[217] Notamment celles de Priya Raghubir en visite à l'Université Paris Dauphine en 2004 et 2005.
[218] S'agissant à chaque fois de Masters différents (i.e. gestion, économie, droit, …), les étudiants ne pouvaient pas communiquer entre eux, sauf effet du hasard (deux étudiants de Masters différents se connaissant).
[219] En réalité, aucun étudiant sollicité n'a refusé de répondre. Les étudiants étrangers ayant répondu ont toutefois été supprimés de l'échantillon final, de manière à ne pas créer de biais dans les réponses (mauvaise compréhension des réponses, incidence de la nationalité sur la mesure de la préoccupation pour le RVP, ..).

« majeurs », non utilisation d'Internet, questionnaire très incomplet, réponses en halo, … Enfin, sur les 243 questionnaires restant, 11 ont été supprimés aléatoirement pour égaliser la taille des effectifs correspondant aux différents traitements. On aboutit ainsi à 29 individus par traitement, chiffre convenable si l'on s'en tient aux recommandations habituelles[220].
Les caractéristiques de cet échantillon global de 232 individus sont à présent commentées.

4.2 Les caractéristiques de l'échantillon

L'échantillon sur lequel a été menée cette étude est composé d'étudiants. Nous revenons sur la justification et les conséquences d'un tel choix avant de présenter le profil des répondants.

D'après Calder, Phillips et Tybout (1981), un échantillon d'étudiants est acceptable quand on teste des modèles théoriques pour lesquels il est important d'avoir un échantillon homogène. Ils expliquent ainsi que lorsque le chercheur veut tester une théorie, il doit créer le test le plus fort possible. Un échantillon de convenance composé d'individus assez homogènes est alors recommandé car l'hétérogénéité augmente les termes d'erreur et réduit la chance de détecter la violation de la théorie quand celle-ci est fausse. Dans ce cas, il convient cependant de recourir à des niveaux de facteurs pertinents pour cette population. Le choix du secteur d'activité (la téléphonie mobile) et du contexte de sollicitation de données (i.e. le jeu-concours) dans le cadre de cette expérience répond ainsi, en grande partie, à cette exigence.

Cette question du type d'échantillon interrogé est importante car elle conditionne la validité externe de l'expérience. On estime généralement que le fait de recourir à un échantillon d'étudiants limite, en grande partie, ce type de validité. Toutefois, pour Cook et Campbell (1979), un échantillon de convenance est lui-même représentatif d'une certaine population, même s'il est souvent difficile d'en définir les caractéristiques et donc de pouvoir appliquer les résultats à une population plus large. De plus, des méta-analyses réalisées dans le domaine du marketing montrent qu'il n'y aurait pas de différence en termes d'effet de taille entre des échantillons d'étudiants et de non étudiants (Sawyer et Ball 1981). De même, plusieurs auteurs suggèrent que les étudiants pourraient ne pas être si différents que cela des non-étudiants, en particulier en ce qui concerne les recherches relatives aux comportements sur Internet (Brackett et Carr 2001 ; Gallager, Parsons et Foster 2001).

[220] On recommande généralement un minimum de 30 individus par traitement, chiffre à partir duquel on peut estimer que la distribution des données est normale.

Nous indiquons à présent, dans le tableau suivant, le profil des 232 répondants interrogés dans le cadre de cette première expérience :

Tableau 10.7 Caractéristiques de l'échantillon interrogé dans le cadre de l'expérience 1

Variables	Valeurs	Fréquences	Pourcentage
Variables sociodémographiques et expérientielles			
Sexe	Masculin Féminin	85 147	36,6% 63,4%
Expérience d'Internet	Moins de 2 ans Entre 2 et 5 ans Plus de 5 ans	17 120 95	7,3% 51,7% 40,9%
Fréquence d'utilisation de la messagerie	Plusieurs fois / jour 1 à 2 fois / jour Plusieurs fois par semaine Moins souvent	27 54 87 40	13,0% 26,0% 41,8% 19,2%
Fréquence de surf	Plusieurs fois / jour 1 à 2 fois / jour Plusieurs fois par semaine Moins souvent	29 35 110 58	12,5% 15,1% 47,4% 25,0%
Nombre d'achats déjà réalisés en ligne depuis qu'ils utilisent Internet	Aucun Moins de 5 De 5 à 20 Plus de 20	111 63 45 13	47,8% 27,2% 19,4% 5,6%
Variables liées à la téléphonie mobile			
Nom de l'opérateur	Orange SFR Bouygues télécom	110 82 37	48% 35,8 16,2%
Type d'abonnement	Carte prépayée Forfait Autre	39 184 3	17,3% 81,4% 1,3%

L'échantillon d'étudiants interrogés est donc majoritairement féminin[221], expérimenté sur Internet (la majorité ont entre 2 et 5 ans d'ancienneté) mais ne l'utilisant pas forcément quotidiennement (la majorité l'utilise plusieurs fois par semaine). Enfin, près de la moitié n'ont jamais acheté sur Internet, ce qui s'explique aussi en grande partie par leur statut et les moyens financiers qui y sont liés. En ce qui concerne l'utilisation de la téléphonie mobile, la répartition entre les trois opérateurs est étrangement similaire aux parts de marché recensées par l'ART[222] fin août 2005 (respectivement 47,5% ; 35,6% et 17,1%[223]).

[221] Cela s'expliquerait surtout par le recrutement dans des filières majoritairement fréquentées par des filles.
[222] ART pour Autorité de Régulation des Télécoms
[223] Source : Le Journal du Net, www.journaldunet.com/cc/05_mobile

Après avoir présenté les caractéristiques de l'instrument de collecte et de l'échantillon interrogé, nous abordons maintenant la partie concernant l'analyse des données.

5 L'analyse des données

Nous verrons, dans un premier temps, les conditions relatives à la validité du plan d'expérience (5.1), avant d'aborder la question de l'effectivité des manipulations réalisées (5.2). Nous indiquerons ensuite très rapidement la qualité psychométrique des instruments de mesure utilisés (5.3) pour terminer par la présentation du plan de traitement (5.4).

5.1 La validité interne du plan d'expérience

La vérification de la validité du plan d'expérience comporte 6 étapes majeures. Celles-ci correspondent, en grande partie, aux conditions d'application des analyses de variance, couramment utilisées pour tester les résultats d'une expérimentation (Howell, 1998) :

- le contrôle des variables exogènes
- l'égalité des traitements
- l'homogénéité des variances
- la normalité des données
- l'indépendance des observations
- l'orthogonalité des variables manipulées

Nous abordons tour à tour chacune des ces conditions.

5.1.1 Le contrôle des variables exogènes

Le contrôle des variables exogènes, la manière de le réaliser et la justification d'un tel contrôle ont déjà été abordés précédemment dans ce chapitre. Nous ne détaillerons donc pas plus ce point et renvoyons les lecteurs au paragraphe 2.2 de ce chapitre.

5.1.2 L'égalité des traitements

Bien qu'il existe des tests permettant de s'affranchir d'une stricte égalité des échantillons, il est généralement recommandé de parvenir à égaliser le nombre d'individus par traitement,

afin de pouvoir réaliser une analyse de variance. Certains questionnaires ayant été supprimés aléatoirement pour parvenir à une égalité du nombre d'individus interrogés, les huit cellules expérimentales sont donc rigoureusement égales, soit 29 individus par traitement.

5.1.3 L'homogénéité des variances

Pour pouvoir procéder à une analyse de variance, les variances des populations doivent nécessairement être égales. Nous procédons alors au test d'égalité des variances de Levene pour chacune des conditions expérimentales (mentions RVP, familiarité et quantité à deux niveaux chacun), eu égard aux variables médiatrices et dépendantes mesurées dans l'expérience 1, à savoir la confidentialité (CONFID), la sensibilité (SENS) et la pertinence (PERT) perçues ainsi que l'attitude face à la communication des données sollicitées (ATT).

Les résultats du test de Levene présentés en Annexe 10.2 (point A) ne conduisent pas à rejeter l'hypothèse d'égalité des variances entre les traitements expérimentaux. Cette condition est donc respectée.

5.1.4 La normalité des variables

Nous examinons la normalité des variables à travers les indices d'asymétrie et d'aplatissement et l'histogramme des distributions. Pour que les variables soient considérées comme normales, il faut que ces coefficients soient respectivement inférieurs à 1 et 1,5 et que les distributions suivent approximativement une courbe de Gauss.
La normalité des variables médiatrices et dépendantes du modèle à tester dans l'expérience 1 est examinée sur la totalité de l'échantillon puis pour chaque traitement expérimental.

Les résultats fournis en Annexe 10.2 (point B) conduisent à accepter la condition de normalité des variables du modèle.

5.1.5 L'indépendance des observations

A travers cette condition, il s'agit de vérifier que les différences liées aux conditions expérimentales ne sont pas biaisées par des différences liées à la répartition de l'échantillon. Nous disposons de 8 groupes expérimentaux ; ces groupes ne doivent donc pas se distinguer d'après leurs caractéristiques individuelles, surtout d'ordre sociodémographique (sexe) et

expérientiel (ancienneté et utilisation d'Internet). Cette condition d'application de l'analyse de variance est censée être assurée *a priori* par l'affectation aléatoire des unités expérimentales aux différents traitements. Un contrôle *a posteriori* est cependant nécessaire. Pour vérifier cette condition, nous effectuons un test de chi-deux. Celui-ci ne doit pas être significatif pour témoigner de l'indépendance des cellules au niveau de ces caractéristiques individuelles.
Les résultats de ce test sont fournis dans le tableau suivant :

Tableau 10.8 Indépendance des observations en fonction des traitements expérimentaux

Caractéristiques	Mentions RVP limitées vs nombreuses	Familiarité Nulle vs élevée	Quantité Faible vs élevée
Sexe	0,019 (p = 0,892)	0,464 (p = 0,496)	*4,178 (p = 0,041)*
Expérience d'Internet	0,187 (p = 0,911)	0,622 (p = 0,733)	4,442 (p = 0,109)
Fréquence de consultation de la messagerie	2,887 (p = 0,577)	7,559 (p = 0,109)	7,166 (p = 0,127)
Fréquence de surf	1,105 (p = 0,776)	3,262 (p = 0,353)	3,262 (p = 0,353)

L'analyse des résultats de ce tableau montre que les échantillons ne se distinguent pas d'après leurs caractéristiques individuelles, à l'exception du sexe qui se distingue sur les échantillons à la quantité faible et élevée. Il y a ainsi plus de femmes qui ont été affectées au traitement « quantité élevée ». Le fait que l'échantillon soit très majoritairement féminin pourrait, en partie, expliquer ce résultat. L'influence de la variable sexe sur les résultats de l'expérimentation sera de toute façon testée par insertion en tant que co-variable.

5.1.6 L'orthogonalité des variables manipulées

La condition *sine qua non* du bon déroulement d'une expérimentation et de la validité d'un plan factoriel quel qu'il soit, est que l'ensemble des variables indépendantes soient réellement indépendantes, c'est-à-dire qu'elles n'aient pas d'influence les unes sur les autres afin de pouvoir isoler les effets de chacune d'entre elles sur la variable dépendante. En effet, si les variables indépendantes sont corrélées, il est impossible d'envisager de les manipuler dans un même plan expérimental (Keppel 1991). L'orthogonalité des facteurs manipulés est donc une condition importante que tout chercheur doit vérifier avant de construire un plan factoriel.

Nous avons ici 3 facteurs manipulés : les mentions de RVP affichées sur la page du formulaire, la familiarité avec l'entreprise sollicitant les données et la quantité de données demandées. Intuitivement, il n'y a *a priori* aucune raison que chacun de ces facteurs n'influence les autres. Nous vérifions cependant cette condition par une analyse des

corrélations, non pas entre les facteurs eux-mêmes (puisqu'il s'agit de variables binaires, codées 1-2) mais entre les perceptions de ces manipulations (récoltées par le biais de 3 ou 4 items postés en fin de questionnaire) appelées Manipulation Checks (MC) et présentées dans le paragraphe suivant.

Les résultats de ces analyses de corrélations sont présentés dans le tableau ci-dessous :

Tableau 10.9 Orthogonalité des facteurs manipulés (corrélations de Pearson)

Facteurs	Mentions RVP perçues	Familiarité perçue	Quantité perçue
Mentions RVP perçues	1	-	-
Familiarité perçue	0,066 (p = 0,319)	1	-
Quantité perçue	- 0,050 (p = 0,449)	0,015 (p = 0,822)	1

D'après ce tableau, il n'existe aucune corrélation significative entre les perceptions des différents facteurs. Il y a donc bien orthogonalité des variables manipulées.

5.2 La vérification des manipulations expérimentales

Avant de traiter les données, nous devons vérifier que les manipulations expérimentales ont été réalisées avec succès. Cette vérification est fortement recommandée par Perdue et Summers (1986). Pour ce faire, plusieurs questions ont été posées en fin de questionnaire afin de mesurer les perceptions des répondants en fonction du traitement expérimental auquel ils étaient affectés. Plus spécifiquement, 4 questions étaient posées pour chaque facteur manipulé, afin de vérifier, par exemple, que les individus affectés au cas « faible » se sentaient bien dans une situation « différente » de ceux affectés au cas « élevé »[224]. De plus, nous posions également une question visant à évaluer le réalisme perçu du scénario proposé. Ces deux points seront abordés successivement ci-dessous.

L'effectivité des manipulations expérimentales

Les 12 questions visant à vérifier les manipulations expérimentales (*Manipulation Checks*) se trouvent dans le questionnaire (page 8) fourni en Annexe 10.1. Parce que nous disposons de 4 questions par facteur manipulé, nous commençons par réaliser une analyse factorielle pour

[224] Les individus affectés au cas « quantité faible » doivent ainsi estimer le formulaire plus court que ceux affectés au cas « quantité élevée ».

chaque groupe de 4 items (suivie du calcul de l'Alpha de Cronbach) afin de vérifier que chaque item mesure bien ce qu'il est censé mesurer et que les « échelles » sont fiables.
Les résultats de ces analyses sont fournis en Annexe 10.3 et résumés ci-dessous :

Tableau 10.10 Analyse factorielle exploratoire sur les « *Manipulation Checks* »

Concept	Nombre d'items	KMO / Test de Bartlett	% de variance	Qualité de la représentation (min – max)	α de Cronbach
Mentions RVP perçues	4	0,805 / 0,000	69,5%	0,61 – 0,77	0,85
Familiarité perçue	4	0,852 / 0,000	91,2%	0,87 – 0,93	0,97[225]
Quantité perçue	4	0,789 / 0,000	74,8%	0,54 – 0,80	0,84

Les résultats de ces analyses confirment la possibilité d'évaluer l'effectivité des manipulations expérimentales sur la base de ces 12 questions. La vérification des manipulations fera alors l'objet de tests-t sur échantillons indépendants. L'objectif est de déduire qu'il existe des différences de moyenne significatives sur la perception des facteurs manipulés[226]. On teste l'hypothèse H0 qui postule l'absence de différence entre les moyennes. Si celle-ci n'est pas validée, alors on ne peut pas rejeter l'inégalité des moyennes (Evrard et al. 2003). Les résultats de ces analyses sont présentés dans le tableau suivant :

Tableau 10.11 Vérification des manipulations expérimentales

		N	Moyennes	Ecart-type	Test t	Sig.
Mentions RVP perçues	limitées	116	2,81	1,650	16,317	0,000
	nombreuses	116	5,80	1,089		
Familiarité perçue	nulle	116	1,34	0,691	30,975	0,000
	élevée	116	6,00	1,468		
Quantité perçue	faible	116	1,67	1,109	23,189	0,000
	élevée	116	5,27	1,253		

Ces résultats confirment l'effectivité des manipulations expérimentales puisque les différences de moyenne sont systématiquement significatives. Ainsi, par exemple, les formulaires conçus pour avoir un niveau de quantité de données demandées faible sont bien perçus comme plus courts que ceux à la quantité élevée et cette différence est significative. Il

[225] Cet alpha très élevé fait soupçonner une forte redondance entre certains items. Nous conservons néanmoins les 4 items, dans la mesure où ce « biais » n'est pas préjudiciable à la suite des analyses.
[226] La fiabilité des « échelles » de « *Manipulation Checks* » ayant été démontrée, nous réalisons le test-t sur chacun des items et sur la moyenne des items. Nous ne présentons dans le tableau 10.13 que le résultat pour la moyenne.

en est de même pour les deux autres facteurs manipulés (Mentions de RVP et Familiarité). Nous pouvons donc considérer que cette manipulation est réussie. Cette conclusion est confirmée par les résultats des analyses concernant le réalisme du scénario.

Le réalisme du scénario

Afin de vérifier que le scénario proposé aux répondants est bien crédible et de réduire ainsi le biais d'artefact de la demande, nous avons inclus, en fin de questionnaire, une question leur demandant d'évaluer le degré de réalisme de la situation proposée sur une échelle sémantique différentielle en 7 points (de pas réaliste du tout à très réaliste). Nous indiquons, dans les tableaux suivants, le niveau de réalisme moyen pour l'ensemble de l'échantillon (tableau 10.12) ainsi que les résultats de tests de chi-deux visant à évaluer la différence de réalisme selon les caractéristiques individuelles des répondants et le scénario auquel ils étaient affectés (tableau 10.13).

Tableau 10.12 Niveau de réalisme perçu du scénario proposé

Niveau de réalisme perçu	Fréquences	Pourcentages	Pourcentages cumulés
1 (pas réaliste du tout)	5	2,2%	2,2%
2	5	2,2%	4,4%
3	20	8,6%	12,9%
4	46	19,8%	32,8%
5	45	19,4%	52,2%
6	77	33,2%	85,3%
7 (très réaliste)	34	14,7%	100%
TOTAL	232	100%	-
Moyenne	5,10		
Médiane	5,00		
Mode	6,00		

Dans l'ensemble, le scénario proposé est donc perçu comme réaliste avec une moyenne de 5,1 sur 7. Moins de 5% des individus considèrent le scénario comme pas (1) ou peu (2) réaliste. De plus, comme l'indique le tableau ci-dessous, le niveau de réalisme perçu ne diffère pas selon les caractéristiques individuelles des répondants (tests de chi-deux non significatifs). Enfin, le niveau de réalisme ne diffère pas non plus selon le niveau de traitement affecté au répondant, hormis pour le facteur « Mentions de RVP » ($p = 0,056$). Un examen approfondi de ce test montre ainsi que les individus affectés au traitement « Mentions de RVP limitées » jugent le scénario moins réaliste que ceux affectés au traitement opposé.

Tableau10.13 Réalisme perçu selon les caractéristiques individuelles et les facteurs manipulés

Variables	Chi-deux	Significativité
Sexe	5,798	p = 0,446
Expérience d'Internet	8,534	p = 0,742
Fréquence de consultation de la messagerie	18,526	p = 0,777
Fréquence de surf	16,684	p = 0,545
Mentions de RVP	12,290	*p = 0,056*
Familiarité	8,982	p = 0,175
Quantité	7,755	p = 0,257

Voyons à présent, avant d'aborder la partie résultats (6), la qualité psychométrique des instruments de mesure utilisés (5.3) ainsi que le plan de traitement envisagé (5.4).

5.3 La qualité psychométrique des instruments de mesure

Les six variables mesurées dans le cadre de cette première expérience (les 3 variables de perceptions, l'attitude, la préoccupation RVP et le niveau d'implication) ont fait l'objet d'un développement d'échelles présenté au chapitre précédent. La qualité psychométrique de l'ensemble de ces instruments de mesure a été soigneusement vérifiée et a pu être établie. Nous ne reviendrons pas sur ces analyses déjà exposées au chapitre précédent et nous nous contentons, pour rappel, d'indiquer dans le tableau suivant, les principaux résultats obtenus :

Tableau 10.14 Qualités psychométriques des instruments de mesure utilisés

Concepts	Nombre d'items	Qualité de la représentation (min – max)	α de Cronbach	Rhô de Jöreskog	VC[227] (ρVC)	VD
Confidentialité perçue (CONFID)	2	0,60 – 0,78	0,77	0,795	0,660	OUI
Sensibilité perçue (SENS)	2	0,62 – 0,75	0,78	0,796	0,661	OUI
Pertinence perçue (PERT)	2	0,73 – 0,77	0,83	0,773	0,630	OUI
Attitude face à la communication des données (ATT)	3	0,59 - 0,83	0,81	0,836	0,634	OUI
Préoccupation RVP (PREOC)	3	0,77 – 0,81	0,87	0,876	0,702	OUI
Implication (IMPL)	3	0,67 – 0,77	0,82	-	-	-

[227] Correspond à la première « dimension » de la validité de trait c'est-à-dire à la validité convergente (VC), la seconde étant la validité discriminante (VD). La validité prédictive n'est pas reproduite dans ce tableau mais a été vérifiée au chapitre 9.

Ce tableau confirme la qualité psychométrique des instruments de mesure mobilisés dans le cadre de cette première expérience et nous autorise à les utiliser pour vérifier les hypothèses posées. Nous terminons cette section en exposant le plan de traitement prévu.

5.4 Le plan de traitement

L'approche classique pour analyser les résultats d'une expérimentation est de recourir soit à une analyse de variance soit à un modèle de régression. L'équivalence de ces 2 techniques est d'ailleurs reconnue par l'ensemble de la communauté scientifique. Si le modèle de régression est la seule méthode possible quand les variables indépendantes sont continues, tel n'est pas le cas ici puisqu'elles correspondent aux facteurs manipulés mesurés par des variables binaires codées (1 ; 2). L'influence des facteurs manipulés sur les trois variables de perception (correspondant respectivement aux hypothèses H1.1, H1.2 et H1.3) sera donc vérifiée par le biais d'une analyse de variance, plus spécifiquement par une analyse de variance multi-variée (puisqu'il y a trois variables dépendantes) avec introduction de variables de contrôle (analyse de covariance), c'est-à-dire par une MANCOVA. Nous venons de voir que les conditions d'application d'une analyse de variance étaient réunies (point 5.1 et 5.2). Nous pouvons donc mener à bien une telle analyse.
En ce qui concerne la partie droite du modèle (et donc les hypothèses H2 et H3e), il s'agit d'établir l'existence de liens de cause à effet entre des variables continues (les 3 variables de perceptions ainsi que la préoccupation RVP sur l'attitude). Nous procéderons donc par le biais d'une régression multi-variée, en introduisant successivement chacune de ces variables au sein d'un bloc, le dernier bloc étant constitué de l'ensemble des variables de contrôle.

Les conditions d'application d'un modèle de régression multiple sont les suivantes :

1. l'espérance des erreurs est nulle (ou au moins constante) : correspond au fait de ne pas oublier de variable explicative qui soit corrélée avec celles incluses dans le modèle
2. la dispersion des résidus doit être homogène (condition d'homoscédasticité)
3. les observations sont indépendantes
4. la distribution des résidus suit une loi normale

Les conditions 1 et 3 ont déjà été vérifiées précédemment. (cf. point 5.1) tandis que les conditions 2 et 4 seront vérifiées lors de l'analyse des résultats (cf. Section 6). Deux autres conditions non indiquées ci-dessus mais qui doivent aussi être examinées concernent respectivement l'absence de multi-colinéarité et la normalité des distributions de l'ensemble

des variables introduites dans le modèle. La première sera vérifiée au moment de l'analyse des résultats par l'examen des indicateurs de Tolérance et de VIF (pour *Variance Inflation Facteur*) qui devront être respectivement supérieur à 0,3 et inférieur à 3,3 pour signifier une absence de multi-colinéarité extrême[228]. La seconde constitue un postulat fondamental selon lequel la distribution des observations, pour chacune des variables métriques et assimilées métriques (cas des variables ordinales), doit suivre une loi normale. En cas d'écart par rapport à la normalité, les tests statistiques comme le t de Student deviennent en effet invalides.

Nous avons déjà vérifié la normalité des variables de perception et d'attitude (cf. paragraphe 5.1.4). Nous vérifions à présent la normalité des autres variables indépendantes du modèle (préoccupation RVP) et des variables de contrôle métriques (niveau d'implication), tous traitements confondus. Les résultats fournis en Annexe 10.4 confirment le respect de cette hypothèse, dans la mesure où les coefficients d'asymétrie et d'aplatissement sont conformes aux règles communément admises.

Au final, le tableau ci-dessous récapitule l'ensemble du plan de traitement envisagé :

Tableau 10.15 Plan de traitement des hypothèses correspondant à l'expérience 1

Hyp.	Variable indépendante	Variable dépendante	Tests réalisés	Résultats escomptés
H1.1 a	Mentions RVP	Confidentialité	MANCOVA (avec introduction des variables de contrôle en tant que co-variables)	Les coefficients liant les facteurs manipulés aux trois variables de perception doivent être significatifs. Nous contrôlons, dans le même temps, l'impact possible des variables exogènes identifiées.
H1.1 b	Mentions RVP	Sensibilité		
H1.2 a	Familiarité	Confidentialité		
H1.2 c	Familiarité	Pertinence		
H1.3 b	Quantité	Sensibilité		
H1.3 c	Quantité	Pertinence		
H2 a	Confidentialité	Attitude	Régression multiple par blocs (bloc 1 : variables de perception, bloc 2 : préoccupation RVP et bloc 3 : variables de contrôle)	Les coefficients correspondant aux variables indépendantes doivent être significatifs. Nous contrôlons, dans le même temps, l'impact possible des variables exogènes identifiées.
H2 b	Sensibilité	Attitude		
H2 c	Pertinence	Attitude		
H3 e	Préoccupation pour le RVP	Attitude		

[228] Ces chiffres constituent des seuils d'acceptation empiriques mais ne sauraient être considérés comme une référence absolue.

Nous présentons, dans la section suivante, l'ensemble des résultats obtenus sur les tests réalisés et faisons ainsi le point sur la validation (ou non) des hypothèses proposées.

6 Présentation des résultats

Nous présentons, dans cette section, les résultats des tests des hypothèses.
Nous débutons (6.1) par ceux qui concernent l'influence de l'exposition à la sollicitation (correspondant aux trois facteurs manipulés, i.e. Mentions RVP, Familiarité et Quantité) sur l'évaluation de celle-ci (correspondant à trois des quatre perceptions identifiées au sein du cadre conceptuel, i.e. confidentialité, sensibilité et pertinence perçue). Pour rappel, six hypothèses ont été développées en ce sens et justifiées au paragraphe 3.2. Elles seront testées conjointement par le biais d'une MANCOVA, dont les résultats sont exposés ci-après.

Nous poursuivons cette section (6.2) par les résultats concernant l'impact de l'évaluation et de la préoccupation pour le RVP sur l'attitude face à la communication des données. Quatre hypothèses sont concernées ici (respectivement H2 a, b et c et H3 e), qui ont été développées et justifiées dans le chapitre 7 et qui seront testées conjointement par le biais d'une régression multiple avec introduction de blocs successifs.
Nous testons ici le cœur de notre modèle à savoir l'impact respectif de la situation (évaluation) et des caractéristiques individuelles (préoccupation RVP) sur le degré d'agrément de l'individu eu égard à la divulgation de ses données (attitude). Notre objectif est alors de montrer que si ces deux « mécanismes » influencent conjointement l'attitude, l'effet des caractéristiques situationnelles est proportionnellement plus important que celui des caractéristiques individuelles. Autrement dit, face à la décision de fournir ou non des données à une entreprise (en particulier sur un site Internet), l'individu se déciderait davantage sur la base de considérations « pratiques » que de convictions personnelles.

6.1 L'impact de l'exposition à la sollicitation sur l'évaluation de celle-ci

Cet effet doit être testé par le biais d'une analyse de variance. L'ensemble des conditions nécessaires à la réalisation d'une analyse de variance ayant été vérifiées précédemment (cf. paragraphe 5.1), nous procédons à la mise en œuvre du plan de traitement.

Comme indiqué ci-dessus, pour tester l'ensemble des hypothèses concernées ici (H1.1 a et b, H1.2 a et c et H1.3 b et c) nous procédons par le biais d'une MANCOVA (procédure GLM, sous SPSS[229]). Nous introduisons, conjointement, les trois facteurs manipulés (Mentions RVP, Familiarité et Quantité) en tant que facteurs fixés, les trois variables de perception (confidentialité, sensibilité et pertinence perçues) en tant que variables dépendantes et l'ensemble des variables de contrôle (implication, sexe, expérience et utilisation d'Internet) en tant que co-variables. Nous spécifions ce modèle en tant que modèle factoriel complet dans la mesure où, bien que n'ayant pas formulé d'hypothèses en ce sens, nous soupçonnons l'existence possible d'effets d'interaction entre les différents facteurs.
Enfin, puisque nos hypothèses sont formulées dans le sens d'une supériorité d'un des paramètres comparés sur l'autre (H0 : m1 – m2 > 0), nous optons pour des tests unilatéraux à 10%. L'hypothèse sera donc validée si la significativité du test F pour le facteur concerné est inférieure ou égale à 10%. Nous conservons néanmoins le seuil habituel de 5% pour l'effet des co-variables.

Les résultats de l'analyse multivariée de covariance (MANCOVA) sont présentés ci-après.

Tout d'abord, les résultats des tests multivariés (Trace de Pillai notamment)[230] montrent que seuls les facteurs Mentions RVP et Quantité ont un impact global significatif sur l'ensemble des variables de perception (p = 0,048 et p = 0,000). L'interaction entre ces deux facteurs est également significative (p = 0,058), tout comme l'effet du niveau d'implication (p = 0,015).

Voyons à présent, dans le détail, l'influence des différents facteurs (et de leurs éventuels effets d'interaction) sur chacune des variables de perception, à travers le tableau suivant.
Par souci de lisibilité, nous indiquons, en grisé, les résultats correspondant au test des 6 hypothèses formulées et, en gras et italique, les coefficients correspondant à des effets significatifs au seuil de 10% (tests unilatéraux) ou de 5% (pour les variables de contrôle).

[229] Nous fournissons en Annexe 10.5, les fichiers de syntaxe ainsi que les principaux résultats (« sorties SPSS ») correspondants à cette analyse.
[230] Ceux-ci sont présentés en Annexe 10.5 (point B).

Tableau 10.16 Influence des facteurs manipulés sur les trois variables de perception

Source	Variable dépendante	Somme des carrés de type III	ddl	Moyenne des carrés	F	Sig.	Hypothèse
Modèle corrigé	ScorCONFID	162,025 (a)	12	13,502	1,729	0,062	
	ScorPERT	245,095 (b)	12	20,425	2,435	0,005	
	ScorSENS	381,904 (c)	12	31,825	3,273	0,000	
Constante	ScorCONFID	188,566	1	188,566	24,141	0,000	
	ScorPERT	128,846	1	128,846	15,362	0,000	
	ScorSENS	671,273	1	671,273	69,042	0,000	
ScorIMPL	ScorCONFID	22,768	1	22,768	2,915	0,089	
	ScorPERT	57,315	1	57,315	6,833	***0,010***	
	ScorSENS	88,729	1	88,729	9,126	***0,003***	
Sexe	ScorCONFID	0,104	1	0,104	0,013	0,908	
	ScorPERT	1,218	1	1,218	0,145	0,703	
	ScorSENS	3,261	1	3,261	0,335	0,563	
Exp_Internet	ScorCONFID	7,183	1	7,183	0,920	0,339	
	ScorPERT	7,444	1	7,444	0,887	0,347	
	ScorSENS	14,735	1	14,735	1,516	0,220	
Util_mess	ScorCONFID	1,592	1	1,592	0,204	0,652	
	ScorPERT	30,475	1	30,475	3,633	***0,058***	
	ScorSENS	24,713	1	24,713	2,542	0,112	
Util_surf	ScorCONFID	2,541	1	2,541	0,325	0,569	
	ScorPERT	0,974	1	0,974	0,116	0,734	
	ScorSENS	37,626	1	37,626	3,870	***0,050***	
Mentions de RVP	***ScorCONFID***	51,305	1	51,305	6,568	***0,011***	***H1.1 a (V[231])***
	ScorPERT	30,774	1	30,774	3,669	***0,057***	
	ScorSENS	33,771	1	33,771	3,473	***0,064***	***H1.1 b (V)***
Familiarité	ScorCONFID	14,230	1	14,230	1,822	0,179	***H1.2 a (NV)***
	ScorPERT	1,724	1	1,724	0,206	0,651	***H1.2 c (NV)***
	ScorSENS	19,976	1	19,976	2,055	0,153	
Quantité	ScorCONFID	0,006	1	0,006	0,001	0,978	
	ScorPERT	80,719	1	80,719	9,624	***0,002***	***H1.3 c (V)***
	ScorSENS	137,903	1	137,903	14,184	***0,000***	***H1.3 b (V)***
Familiarité * Quantité	ScorCONFID	23,455	1	23,455	3,003	***0,085***	
	ScorPERT	26,085	1	26,085	3,110	***0,079***	
	ScorSENS	22,247	1	22,247	2,288	0,132	
Familiarité * Mentions de RVP	ScorCONFID	0,007	1	0,007	0,001	0,977	
	ScorPERT	9,352	1	9,352	1,115	0,292	
	ScorSENS	3,895	1	3,895	0,401	0,527	
Quantité * Mentions de RVP	ScorCONFID	24,226	1	24,226	3,101	***0,080***	
	ScorPERT	0,033	1	0,033	0,004	0,950	
	ScorSENS	19,910	1	19,910	2,048	0,154	
Familiarité * Quantité * mentionsRVP	ScorCONFID	20,866	1	20,866	2,671	0,104	
	ScorPERT	8,474	1	8,474	1,010	0,316	
	ScorSENS	8,718	1	8,718	0,897	0,345	
Erreur	ScorCONFID	1702,814	218	7,811			
	ScorPERT	1828,496	218	8,388			
	ScorSENS	2119,533	218	9,723			
Total	ScorCONFID	17394,000	231				
	ScorPERT	14018,410	231				
	ScorSENS	19336,000	231				

(a) $R^2 = 0,087$ (R^2 ajusté = 0,037) ; (b) $R^2 = 0,118$ (R^2 ajusté = 0,070) ; (c) $R^2 = 0,153$ (R^2 ajusté = 0,106)

[231] V pour hypothèse validée, NV pour non validée

Nous discutons tout d'abord les résultats relatifs à l'impact de chaque facteur manipulé (6.1.1, 6.1.2 et 6.1.3) avant d'évoquer l'influence des co-variables (variables de contrôle) (6.1.4).

6.1.1 L'impact des mentions de RVP

Au regard des résultats obtenus et conformément à ce que nous avions prévu, le facteur « Mentions de RVP » aurait un effet significatif à la fois sur la confidentialité (p = 0,011) et sur la sensibilité perçue des données sollicitées (p = 0,064), **validant ainsi les hypothèses H1.1 a et H1.1 b**. Le fait pour un site d'apposer (ou non) des mentions relatives à la protection de la vie privée sur la page du formulaire à remplir aurait donc une incidence sur l'évaluation que l'individu porte à la sollicitation, évaluation susceptible (nous le verrons au paragraphe 6.2) d'influencer son attitude face à la communication de ses données.

Plus précisément, quand les mentions de RVP sont nombreuses, l'individu perçoit une plus forte confidentialité (8,6 vs. 7,8) et une moins forte sensibilité de ses données (8,2 vs 8,9) (cf. Figure 10.4). Apposer des mentions de RVP sur le formulaire contribue donc à rassurer la personne quant à l'utilisation ultérieure de ses données (confidentialité) et quant à la vulnérabilité que leur divulgation peut occasionner (sensibilité).

Figure 10.4 Impact des Mentions de RVP sur la confidentialité et la sensibilité perçues

Les « Mentions RVP » auraient aussi un impact significatif (que nous n'avions pas prévu) sur la pertinence perçue des données sollicitées (p = 0,057). Plus précisément, un nombre élevé de mentions de RVP amènerait l'individu à percevoir la sollicitation comme plus pertinente (7,5 vs 6,9). Autrement dit, le fait d'apposer des mentions de RVP contribuerait à renforcer la

légitimité de l'entreprise à solliciter de telles informations. Ce résultat se révèle donc particulièrement intéressant, puisqu'il n'avait jamais été suggéré par la littérature.

6.1.2 L'impact de la Familiarité

Au regard des résultats obtenus, et contrairement à ce que nous avions supposé, le facteur « Familiarité » n'a aucun effet (direct), ni sur la confidentialité (p = 0,179), ni sur la pertinence perçue des données sollicitées (p = 0,651), conduisant ainsi à **ne pas valider les hypothèses H1.2 a et H1.2 c**. Comme nous nous y attendions, aucun effet n'est décelé non plus sur la sensibilité perçue des données sollicitées (p = 0,153). Le fait pour un individu d'être sollicité par une entreprise familière plutôt que par une entreprise à laquelle il n'a jamais eu à faire auparavant n'aurait donc aucune incidence sur l'évaluation qu'il porterait à la sollicitation, tant du point de vue de la confidentialité que de la pertinence perçue des données sollicitées. Ce résultat peut paraître surprenant, eu égard à l'abondante littérature (notamment sur le marketing relationnel) préconisant le contraire. Il s'explique toutefois, au moins en partie, par l'existence d'effets d'interaction. La familiarité n'aurait ainsi pas d'effet direct sur la confidentialité et la pertinence mais un effet combiné avec la quantité de données demandées (p = 0,085 et p = 0,079 respectivement). Autrement dit, l'individu sollicité ne tiendrait pas compte de la seule familiarité pour évaluer la sollicitation mais plutôt de son interaction avec le facteur « Quantité de données demandées » (cf. Figure 10.5).

Il apparaît ainsi que, pour une familiarité élevée, une faible quantité de données demandées (formulaire court) engendrerait un plus haut niveau de confidentialité perçue qu'une quantité élevée (formulaire long). Cet effet s'explique assez facilement : un individu connaissant bien l'entreprise sollicitant les données (pour avoir déjà eu de nombreux contacts avec elle) pourrait voir, dans un formulaire long, la possibilité pour celle-ci de faire un usage externe de ces données (transfert ou revente à des tiers) et se sentirait alors plus rassuré si elle se contente de demander des données basiques visant uniquement à l'identifier (cas du formulaire court). Toutefois, et ce, de façon contre-intuitive, l'effet inverse se produirait en cas de familiarité nulle. En l'occurrence, une quantité élevée de données demandées engendrerait une plus forte confidentialité perçue que dans le cas d'une quantité réduite. Ce constat peut sembler difficile à expliquer de prime abord, sauf à imaginer le raisonnement suivant. Un individu ne connaissant pas l'entreprise sollicitant ses données et qui se voit sommé d'indiquer uniquement son nom, son adresse et son email (formulaire court) peut parfaitement se dire que l'unique objectif de cette dernière est de constituer un fichier qui sera ensuite revendu à

des tiers (faible confidentialité). Il ne peut donc pas imaginer, étant donné le faible nombre de renseignements demandés, qu'elle ait pour vocation de mieux le connaître afin d'établir une relation privilégiée avec lui. Au contraire, si elle lui demande de détailler davantage son profil et d'indiquer ses goûts et ses préférences (formulaire long), il peut se dire qu'il s'agit là d'une première rencontre utile visant à établir les bases d'une relation à long terme. Une telle entreprise aurait alors de grandes chances de garantir une confidentialité élevée de ses données, de manière à entamer la relation sur des bases saines et solides.

L'effet d'interaction de la familiarité et de la quantité sur la pertinence perçue des données sollicitées peut aussi s'expliquer facilement. En effet, si en cas de familiarité nulle, il est jugé plus pertinent de demander une quantité limitée d'informations, les résultats indiquent qu'il serait encore plus pertinent de le faire en cas de familiarité élevée. Autrement dit, si l'on raisonne en termes de connaissance client, il serait préférable de toujours limiter le nombre d'informations requises à chaque interaction. De plus, il vaudrait mieux placer ce type de sollicitation au début de la relation (phases de connaissance et d'exploration au sens de MacNeil 1980 et de Dwyer Schurr et Oh 1987) plutôt qu'au milieu ou à la fin, au risque de paraître trop intrusif. Enfin, les entreprises devraient aussi adapter leurs formulaires à ce qu'elles savent déjà du client à ce moment là. Pourquoi en effet lui demander pour la énième fois son adresse postale ou sa situation familiale s'il les a déjà fournies auparavant ? Il serait alors préférable d'insister sur les goûts, les préférences, les habitudes et les comportements de la personne plutôt que sur des données basiques déjà présentes dans la base. Ces questions auraient en tout cas plus de chance d'être jugées pertinentes, surtout en cas de relation déjà avancée. Ce résultat encourage donc les entreprises à réfléchir sérieusement à la quantité et au type de questions à poser sur ces formulaires et à les adapter au contexte de la relation.

Figure 10.5 Effets d'interaction Familiarité x Quantité

6.1.3 L'impact de la Quantité de données demandées

Au regard des résultats obtenus et conformément à ce que nous avions supposé, le facteur « Quantité » a un effet à la fois sur la sensibilité (p = 0,000) et sur la pertinence perçue des données sollicitées (p = 0,002), conduisant ainsi à **valider les hypothèses H1.3 b et H1.3 c**. Comme nous nous y attendions aussi, aucun effet de ce facteur sur la confidentialité perçue des données n'est décelé (p = 0,978). Il agirait toutefois sur ce type de perception en interaction avec le facteur « Mentions de RVP » (p = 0,080).

L'effet direct de la Quantité sur les perceptions de sensibilité et de pertinence serait conforme au sens supposé (cf. Figure 10.6). Ainsi, plus la quantité de données demandées est élevée (formulaire long), plus la sensibilité perçue augmente et plus la pertinence perçue diminue. De tels résultats s'expliquent aisément. Quand le nombre d'informations à fournir augmente, le niveau de vulnérabilité de l'individu s'accroît parallèlement (sensibilité élevée), tandis que la légitimité de l'entreprise à recueillir de telles données se réduit (faible pertinence). L'évaluation de la sollicitation serait donc d'autant plus favorable (en termes de sensibilité et de pertinence) que le nombre d'informations à fournir serait limité car cela rassurerait l'individu sur les intentions de l'entreprise sollicitant ces données.

Figure 10.6 Impact de la Quantité sur la sensibilité et la pertinence perçues

En résumé, **quatre des six hypothèses posées concernant l'impact des facteurs manipulés sur l'évaluation de la sollicitation sont validées : il s'agit des hypothèses H1.1 a, H1.1 b, H1.3 b et H1.3 c**, qui soulignent le rôle respectif des mentions de RVP et de la Quantité de données demandées. De plus, si **les deux autres hypothèses concernant l'impact de la Familiarité (H1.2a et H1.2c) ne sont pas validées**, on constate toutefois l'existence d'un effet d'interaction entre la Familiarité et la Quantité qui expliquerait, au moins en partie, ce résultat.

Au final, la confidentialité perçue des données sollicitées dépendrait des mentions de RVP figurant sur le formulaire et des interactions entre ces mentions et la quantité et entre la familiarité et la quantité. La sensibilité perçue des données sollicitées serait, quant à elle, influencée par l'effet direct des mentions de RVP et de la Quantité de données demandées. Enfin, la pertinence des données serait évaluée sur la base des mentions de RVP, du nombre de données demandées (Quantité) et de l'interaction entre ce dernier et la Familiarité avec l'entreprise sollicitant les informations.

Cette analyse de variance incluait des variables exogènes individuelles dont nous souhaitions contrôler l'impact. Nous commentons dans le paragraphe suivant (6.1.4) les résultats concernant l'effet de ces variables de contrôle.

6.1.4 L'impact des variables exogènes

Sur l'ensemble des variables exogènes introduites en tant que co-variables, seul le niveau d'implication a un impact global significatif ($p = 0,015$) sur les trois variables de perception (confidentialité, sensibilité et pertinence). Si on entre dans les détails, on s'aperçoit que cet effet concerne surtout la sensibilité perçue ($F = 9,126$, $p = 0,003$) et la pertinence perçue ($F = 6,833$, $p = 0,010$), et, dans une moindre mesure, la confidentialité ($F = 2,915$, $p = 0,089$).

Précisément, un niveau élevé d'implication diminuerait la sensibilité perçue et augmenterait la pertinence perçue (corrélations de Pearson = 0,142 et - 0,145 ; $p = 0,05$ respectivement). Ces deux résultats vont dans le sens de ce qui était attendu. Il semble en effet normal de constater qu'un individu fortement impliqué dans la téléphonie mobile trouve pertinent qu'on sollicite ses données et juge celles-ci moins sensibles qu'un individu faiblement impliqué.

Ce constat correspond en tout point aux conclusions des recherches sur l'implication (et notamment aux travaux de Amine 1990) selon lesquelles un individu impliqué est plus intéressé par l'achat d'un produit et sera donc plus disposé à fournir des renseignements à une entreprise susceptible de répondre à ses attentes et de proposer des offres correspondant à ses besoins.

A l'opposé, deux des variables de contrôle n'auraient aucun effet sur l'évaluation de la sollicitation : il s'agit du sexe d'une part (p = 0,927) et de l'expérience d'Internet d'autre part (p = 0,598). Autrement dit, hommes et femmes d'un côté et experts et novices sur Internet de l'autre ne montreraient aucune différence quant à la manière d'évaluer une sollicitation de leurs données personnelles en ligne. Ce constat, bien que contraire à ce qu'on aurait pu supposer d'après l'analyse de la littérature, n'est cependant pas forcément surprenant. En effet, l'impact des variables individuelles est très largement controversé, certains auteurs démontrant l'existence d'un effet quand d'autres constatent l'effet inverse et d'autres encore l'absence d'effet. Dès lors, même si les conclusions de la littérature étaient plus assurées en ce qui concerne l'impact de ces deux variables (sexe et expérience d'Internet), ce résultat peut se comprendre. Il pourrait ainsi s'expliquer par le manque de variance de notre échantillon sur ces deux critères (pour mémoire l'échantillon est très majoritairement féminin et « aguerri » aux pratiques d'Internet.) La même étude menée sur un échantillon plus diversifié pourrait donc donner des résultats différents, ce qui conduit à prendre ceux-là avec précaution.

Les deux dernières variables de contrôle auraient, quant à elles, un effet très ciblé, sur la pertinence perçue pour ce qui concerne la fréquence de consultation de la messagerie (p = 0,058) et sur la sensibilité perçue pour ce qui concerne la fréquence de surf (p = 0,050). Plus précisément, les individus consultant rarement leur messagerie perçoivent la sollicitation de leurs données comme moins pertinente que ceux qui la consultent quotidiennement. Ce résultat n'est pas très surprenant. Un individu ayant une faible utilisation d'Internet est peu habitué à être sollicité et trouvera donc ce genre de sollicitation très intrusive. De même, on constate que les individus surfant rarement jugent les données à fournir peu sensibles comparativement aux individus surfant quotidiennement. Là encore, on peut expliquer ce résultat par l'habitude conférée par la pratique du surf. Les personnes surfant tous les jours savent ainsi qu'il n'est pas forcément nécessaire de fournir beaucoup d'informations pour obtenir ce que l'on cherche tandis que les novices peuvent s'imaginer le contraire.

Dans l'ensemble, c'est l'implication qui a l'impact le plus fort sur l'évaluation de la sollicitation, les autres variables exogènes ayant un impact plus restreint, voire inexistant.
Ce critère (le niveau d'implication) doit donc également être pris en compte pour évaluer l'impact de la situation sur la communication de données personnelles.

Après avoir étudié la partie gauche du modèle (hypothèses H1.1, H1.2 et H1.3), passons à présent à la partie droite qui concerne l'impact respectif de l'évaluation de la sollicitation et de la préoccupation pour le respect de la vie privée sur l'attitude face à la communication des données. Cette partie sera testée par le biais d'une régression multiple par blocs successifs.

6.2 L'impact de l'évaluation de la sollicitation et de la préoccupation pour le RVP sur l'attitude face à la communication des données

L'ensemble des conditions préalables à la réalisation d'une régression ayant été vérifiées (cf. paragraphe 5.4)[232], nous procédons à la mise en œuvre du plan de traitement. Comme indiqué ci-dessus, pour tester les hypothèses concernées ici (H2 a, b et c et H3 e) nous procédons par le biais d'une régression multiple avec introduction de blocs successifs (méthode « entrée », sous SPSS). Nous introduisons, dans le bloc 1, les trois variables de perception (confidentialité, sensibilité et pertinence), dans le bloc 2, la préoccupation pour le RVP et, dans le bloc 3, toutes les variables de contrôle (implication, sexe, expérience et utilisation d'Internet). La variable dépendante est l'attitude face à la communication des données.

Les résultats concernant les conditions d'utilisation de la régression autres que celles vérifiées précédemment sont présentés en Annexe 10.6, sauf en ce qui concerne la multicolinéarité entre les variables indépendantes du modèle, qui est vérifiée ci-après. Les indicateurs de Tolérance et de VIF fournis au tableau 10.17 (respectivement supérieurs à 0,3 et inférieurs à 3,3) montrent alors qu'il n'existe pas de multicolinéarité extrême, et que nous pouvons donc commenter les résultats, sans crainte de biais possibles au niveau des tests t.

Les résultats de la régression multiple réalisée sont résumés dans le tableau ci-après. Nous y faisons figurer les indicateurs d'ajustement du modèle ainsi que les coefficients obtenus. Nous commentons ensuite les résultats concernant la validation (ou non) des hypothèses testées.

[232] Certaines conditions ne peuvent être vérifiées qu'*a posteriori* et seront donc détaillées dans cette section.

Tableau 10.17 Régression visant à expliquer l'attitude face à la communication des données

	Bloc 1		Bloc 2		Bloc 3		Colinéarité	
	B	Sig.	B	Sig.	B	Sig.	Tol.	VIF
Constante	8,918	0,000	13,175	0,000	13,745	0,000		
Confidentialité perçue	0,291	**0,003**	0,202	**0,042**	0,192	**0,054**	0,761	1,314
Sensibilité perçue	- 0,361	**0,000**	- 0,313	**0,001**	- 0,308	**0,001**	0,607	1,647
Pertinence perçue	0,485	**0,000**	0,377	**0,001**	0,346	**0,002**	0,546	1,830
Préoccupation RVP			-0,220	**0,001**	- 0,233	**0,000**	0,656	1,523
Implication					0,111	0,086	0,923	1,084
Sexe					- 0,127	0,808	0,962	1,040
EXP d'Internet					- 0,078	0,852	0,939	1,064
UTIL (messagerie)					- 0,113	0,598	0,922	1,085
UTIL (surf web)					- 0,225	0,433	0,838	1,194
R^2	0,354		0,386		0,399		-	
R^2 ajusté	0,346		0,375		0,374		-	

Ces résultats montrent que les trois variables de perception (confidentialité, sensibilité et pertinence) ont une influence significative sur l'attitude face à la communication des données, y compris après introduction du deuxième et du troisième bloc (p = 0,054 ; p = 0,001 et p = 0,002 respectivement), **validant ainsi les hypothèses H2 a, H2 b et H2 c**. Comme pressenti donc, l'évaluation de la sollicitation (sous la forme de perceptions) influence de manière significative l'attitude de l'individu face à la communication de ses données personnelles.

Cette influence « situationnelle »[233] s'avère d'ailleurs prépondérante puisqu'elle explique près de 35% de la variance du modèle. Parmi ces trois variables perceptuelles, deux d'entre elles semblent d'ailleurs plus déterminantes (la sensibilité et la pertinence perçue) et expliqueraient donc davantage l'attitude que la confidentialité perçue[234].

L'influence de la préoccupation pour le respect de la vie privée (variable individuelle) s'avère également significative (p = 0,000), sans que son introduction n'améliore toutefois de manière importante le pouvoir explicatif du modèle (37,4% avec vs. 34,6% sans, soit moins de 3% de gain). **Nous validons ainsi également l'hypothèse H3 e** qui préconisait une influence de cette variable sur l'attitude face à la communication des données. De plus, l'examen des coefficients standardisés montre que la préoccupation pour le RVP influence plus l'attitude que la confidentialité perçue et que son influence est quasi équivalente à celle de la sensibilité et de la pertinence perçue (bêta = - 0,229).

[233] Par opposition à l'influence individuelle liée aux caractéristiques personnelles des individus.

[234] Nous tirons cette information de la lecture des coefficients standardisés (bêta) qui sont respectivement de 0,116 (pour la confidentialité) ; 0,220 (pour la pertinence) et – 0,215 (pour la sensibilité).

Enfin, l'introduction des variables de contrôle n'améliore pas le pouvoir explicatif du modèle, aucune de ces variables n'étant déterminante au seuil de 5%. Seule l'implication aurait une influence sur l'attitude, le coefficient n'étant toutefois significatif qu'au seuil de 10%.

Pour conclure cette seconde partie, nous pouvons dire que **nous validons l'ensemble des 4 hypothèses concernées (H2 a, H2 b, H2 c et H3 e)**, soulignant ainsi l'impact des 3 variables de perception et de la préoccupation pour le RVP sur l'attitude face à la communication des données.

Il apparaît toutefois que l'influence de la situation (à travers les trois perceptions) serait significativement plus importante que l'influence des caractéristiques individuelles (à travers la préoccupation pour le RVP), ces dernières expliquant moins de 3% du modèle contre 35% pour les caractéristiques situationnelles.

Ce résultat est très important puisqu'il montre que, face à une sollicitation de ses données personnelles, l'individu se déciderait davantage sur la base de la situation dans laquelle il se trouve et des conditions qui lui sont proposées que d'après ses convictions personnelles. Ceci confirme les conclusions de Acquisti (2004) et Acquisti et Grossklags (2004) qui montraient l'existence d'un écart entre l'attitude générale de l'individu face au respect de sa vie privée et sa décision, en cas de sollicitation effective. Ainsi, un individu fortement préoccupé par le respect de sa vie privée pourrait accepter de délivrer des informations personnelles si le contexte dans lequel il se trouve lui apparaît sécurisant (forte confidentialité perçue, pertinence perçue et faible sensibilité des données sollicitées), et inversement.

7 Récapitulatif des résultats et limites inhérentes à cette expérience

7.1 Récapitulatif des résultats et commentaires

Nous présentons ici le récapitulatif des hypothèses testées dans le cadre de cette première expérience puis commentons brièvement ci-après les principaux résultats obtenus.

Tableau 10.18 Récapitulatif des hypothèses à tester et des résultats obtenus (Expérience 1)

Hypothèses testées dans l'Expérience 1		Résultats
H 1.1 (a) (1)	Plus les mentions de respect de la vie privée (RVP) sont nombreuses (vs limitées), plus la confidentialité des données sollicitées est perçue comme élevée (vs. faible)	**Validée (p = 0,011)**
H 1.1 (b) (2)	Plus les mentions de respect de la vie privée (RVP) sont nombreuses (vs limitées), moins (vs plus) la sensibilité des données sollicitées est perçue comme élevée	**Validée (p = 0,064)**
H 1.2 (a) (3)	Plus la familiarité avec l'entreprise sollicitant les données est élevée (vs. nulle), plus la confidentialité des données sollicitées est perçue comme élevée (vs. faible)	Non validée (p = 0,179)
H 1.2 (c) (4)	Plus la familiarité avec l'entreprise sollicitant les données est élevée (vs. nulle), plus la pertinence des données sollicitées est perçue comme élevée (vs faible)	Non validée (p = 0,651)
H 1.3 (b) (5)	Plus la quantité de données demandées est élevée (vs. faible), plus (vs moins) leur niveau de sensibilité est perçu comme élevé	**Validée (p = 0,001)**
H 1.3 (c) (6)	Plus la quantité de données demandées est élevée (vs. faible), moins (vs plus) la pertinence des données sollicitées est perçue comme élevée	**Validée (p = 0,002)**
H 2 (a) (7)	Plus la confidentialité des données sollicitées est perçue comme élevée (vs. faible), plus l'attitude face à leur communication est favorable (vs défavorable)	**Validée (p = 0,054)**
H 2 (b) (8)	Plus la sensibilité des données sollicitées est perçue comme élevée (vs. faible), moins l'attitude face à leur communication est favorable (vs défavorable)	**Validée (p = 0,001)**
H 2 (c) (9)	Plus la pertinence des données sollicitées est perçue comme élevée (vs faible), plus l'attitude face à leur communication est favorable (vs défavorable)	**Validée (p = 0,002)**
H 3 (e) (10)	Plus la préoccupation pour le RVP est élevée (vs faible), moins l'attitude face à la communication des données sollicitées est favorable (vs défavorable)	**Validée (p = 0,000)**

Les résultats de cette première expérience montrent que les mentions de RVP, la familiarité avec l'entreprise sollicitante[235] et la quantité de données demandées pourraient avoir une influence significative sur le comportement de divulgation en ligne, à travers leur influence sur l'évaluation que l'individu fait de la sollicitation.

La présence de mentions de RVP apparaît ainsi susceptible d'accroître la prédisposition des individus à fournir des données personnelles, en accentuant la confidentialité perçue des données et en réduisant leur sensibilité perçue. Si la familiarité n'a aucun effet direct, elle agirait toutefois en interaction avec la quantité de données demandées pour influencer à la fois la confidentialité perçue mais aussi la pertinence perçue des données sollicitées. Enfin, le fait d'accroître le nombre de données demandées pourrait détériorer la probabilité que la personne accepte de les fournir, en accentuant leur niveau de sensibilité et en réduisant la pertinence perçue d'une telle sollicitation. Des variables exogènes, principalement le niveau d'implication vis-à-vis de la catégorie de produit/services concernée, semblent également susceptibles d'influencer cette phase d'évaluation.

La seconde partie des résultats concerne l'influence relative de la situation de collecte (à travers son évaluation) et des caractéristiques individuelles (notamment de la préoccupation pour le respect de la vie privée) sur l'attitude de l'individu face à la communication de ses données. Il apparaît ainsi que les caractéristiques situationnelles seraient prépondérantes et pourraient dépasser les simples convictions personnelles.

Nous résumons dans la figure suivante l'ensemble des hypothèses validées et invalidées :

[235] La familiarité n'aurait pas un effet direct mais en interaction avec la quantité de données demandées.

Figure 10.7 Récapitulatif des hypothèses validées dans le cadre de l'expérience 1

Comme dans toute logique expérimentale, celle menée ici n'est pas exempte de limites qu'il convient de présenter maintenant. Sans remettre en cause totalement les résultats obtenus, elles amènent toutefois à prendre ces derniers avec précaution et à réfléchir aux solutions à mettre en œuvre dans les expérimentations suivantes pour en réduire l'effet.

7.2 Les limites de l'expérience 1

Trois limites majeures ont pu être détectées dans le cadre de cette première expérience. Elles concernent l'opérationnalisation des facteurs manipulés, la validité interne et externe.

Par limite relative à l'opérationnalisation des facteurs manipulés, nous entendons les conséquences liées aux choix effectués dans ce domaine. Ainsi, par exemple, nous avons choisi d'opérationnaliser la politique de RVP par le biais de mentions présentes sur le formulaire de données. Nous aurions pu, tout aussi bien, nous en tenir à la présence ou à l'absence d'une charte au sein du site, comme le font une majorité de chercheurs. De plus, comme le soulignent Sprott, Hardesty et Miyazaki (1998), le format choisi pour opérationnaliser les mentions de RVP (telles que celles choisies ici) peut également avoir des conséquences sur les perceptions et les comportements des consommateurs et donc sur les résultats obtenus. Des recherches seraient donc nécessaires pour tester l'impact d'autres formes de mentions et voir ainsi l'impact réel de celles-ci, quel que soit le format utilisé.
La manière dont nous avons choisi d'opérationnaliser la familiarité (cf. 2.1.2) pourrait aussi avoir eu des conséquences sur les résultats et, en particulier, sur l'absence de lien direct entre celle-ci et les variables de perception.

La deuxième limite que nous avons pu identifier est propre à réduire la validité interne de cette expérience. Elle concerne en effet le fait d'interroger des individus à propos d'une situation de sollicitation de données personnelles sur Internet, par le biais d'un questionnaire papier. Il est probable que cette mise en situation ne constitue pas une méthode appropriée pour former des jugements dans ce domaine et qu'il soit alors nécessaire d'envisager un design plus réaliste. Interroger les répondants par le biais d'un questionnaire en ligne constituerait une première approche possible qui pourrait par la suite être complétée par une mise en situation réelle.

La troisième et dernière limite concerne les choix de l'échantillon interrogé et du contexte de mise en situation et la faible validité externe que ces choix engendrent. Dans cette première expérience nous avons en effet interrogé un échantillon homogène d'individus (jeunes, étudiants au niveau d'études élevé) que nous avons soumis à un scénario spécifique de sollicitation de données personnelles (formulaire à remplir dans le cadre d'un jeu-concours organisé par un opérateur de téléphonie mobile). Le choix du contexte est une première limite à la généralisation des résultats obtenus. Il s'agit en effet d'un cas très particulier de sollicitation de données personnelles et il est fort à parier que ces mêmes individus auraient réagi tout autrement dans un autre cadre de sollicitation (autre secteur, autre contexte). De plus, comme nous l'avons déjà souligné au paragraphe 4.2, le choix d'un échantillon composé d'étudiants n'est pas exempt de reproches possibles. Ainsi, les échantillons de convenance (en particulier ceux formés d'étudiants) ne permettent pas (toujours) de fournir un contexte adéquat pour obtenir des attitudes et des comportements représentatifs de l'ensemble de la population, de par les spécificités auxquelles ils sont confrontés. Il est donc nécessaire de prévoir une nouvelle étude sur un échantillon plus représentatif de la population française pour pouvoir dire si les résultats obtenus ici sont valables plus généralement.

Conclusion du chapitre 10

Ce chapitre a présenté les résultats d'une expérimentation visant à tester, d'une part, l'effet de l'exposition à une sollicitation de données personnelles sur l'évaluation de celle-ci et, d'autre part, l'impact de cette évaluation et de la préoccupation de l'individu pour le respect de sa vie privée sur son attitude face à la communication des données.

Après avoir indiqué les objectifs de l'expérimentation, nous avons présenté les facteurs manipulés, les variables contrôlées et la manière dont l'expérience s'est déroulée. Des hypothèses spécifiques à l'étude ont ensuite été développées, conduisant à la formulation du modèle à tester. La présentation des caractéristiques des répondants a ensuite permis d'insister sur les biais engendrés par ce type d'expérience.

Les analyses préalables ont permis de souligner la validité interne du plan d'expérience, du fait du contrôle des variables exogènes, de l'égalité des échantillons, de la normalité des variables testées, de l'indépendance des observations et de l'orthogonalité des facteurs manipulés. Nous avons également été soucieux de vérifier l'effectivité des manipulations réalisées, afin de nous assurer que celles-ci avaient été comprises par les répondants. Après avoir établi la qualité psychométrique de l'ensemble des instruments de mesure, nous avons présenté le plan de traitement pressenti pour mener à bien nos analyses. Ces analyses ayant démontré la possibilité de tester nos hypothèses dans le respect des conditions exigées, nous avons procédé à l'étude des résultats.

Les résultats obtenus confirment l'hypothèse générale d'un effet de l'exposition à la sollicitation sur l'évaluation de celle-ci par le consommateur (proposition P1). En particulier, les mentions de RVP et la Quantité de données demandées auraient une influence manifeste sur la perception de la sollicitation, tandis que la Familiarité avec l'entreprise sollicitante n'aurait qu'un effet indirect en interaction avec la Quantité. Nous confirmons également l'influence à la fois situationnelle (par le biais de l'évaluation de la sollicitation) et individuelle (par l'intermédiaire de la préoccupation pour le RVP) sur l'attitude face à la communication des données demandées, la première se révélant d'ailleurs largement prépondérante. Autrement dit, d'après les résultats de cette première expérience, en cas de sollicitation, les considérations « pratiques » l'emporteraient sur les convictions personnelles.

Partie III – EXPLIQUER

Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne

Analyse empirique des résultats

CHAPITRE 11

SOLLICITATION DE DONNÉES PERSONNELLES ET INTENTIONS COMPORTEMENTALES DE RÉPONSE (ÉTUDE 4)

« We believe that online access panels are the future for consumer data collection around the world and that quality data can only come from quality panels ».

Bloomerce (http : //www.bloomerce.com)

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

Introduction du chapitre 11

Dans le cadre de ce travail doctoral, nous cherchons à expliquer les réactions attitudinales et comportementales des consommateurs face à une sollicitation de leurs données personnelles, plus particulièrement, dans le cas d'un formulaire à remplir sur Internet. Comme nous l'avons indiqué précédemment, le modèle conceptuel que nous souhaitons tester (cf. figure 8.2 du chapitre 8) le sera en trois temps, à travers trois expériences successives.

Dans le chapitre précédent (i.e. chapitre 10), nous avons exposé les résultats de la première expérience réalisée dans le cadre du plan expérimental. Cette expérience visait, dans un premier temps, à étudier l'impact de l'exposition à la sollicitation (à travers trois facteurs manipulés) sur l'évaluation de celle-ci (reprenant trois des quatre perceptions identifiées dans le modèle conceptuel). Dans un second temps, nous souhaitions étudier l'impact respectif de la situation (à travers l'évaluation de celle-ci) et des convictions personnelles (résumées par la préoccupation pour le respect de la vie privée) sur l'attitude de l'individu face à la communication des données demandées. Nos résultats tendraient d'ailleurs à prouver la supériorité des facteurs situationnels sur les variables individuelles.

Malgré les conclusions intéressantes que cette première expérience permet de tirer, celle-ci n'est pas exempte de limites, auxquelles nous souhaitons remédier dans les expérimentations suivantes. Ainsi, dans la deuxième expérience, exposée dans le présent chapitre, nous sondons une population différente que nous interrogeons en ligne.

La même démarche que celle suivie pour mener à bien la première expérience sera utilisée dans celle-ci (cf. figure 10.1, chapitre 10). De même, nous suivrons approximativement le même plan pour en exposer les résultats. Nous évoquerons ainsi successivement les objectifs et apports de l'expérience 2 (1), avant d'exposer le design expérimental (2). Nous présenterons ensuite le modèle à tester et formulerons les hypothèses de recherche spécifiques à cette deuxième expérience (3). Nous exposerons également brièvement les caractéristiques du mode de recueil des données et des répondants interrogés (4). Enfin, après avoir indiqué les étapes préalables à l'analyse de données (5), nous présenterons les résultats (6) que nous récapitulerons et commenterons rapidement avant d'en présenter les principales limites (7).

1 Objectifs et apports de l'expérience 2

Comme prévu au chapitre 8, dans cette deuxième expérience, nous nous focaliserons sur la partie centrale du modèle à tester (cf. figure 11.1 ci-dessous), à savoir de l'exposition à la sollicitation (facteurs situationnels) jusqu'aux intentions comportementales.

Figure 11.1 Modèle simplifié testé dans l'expérience 2

(*) Variables déjà mesurées dans l'expérience 1 et qui ne seront pas étudiées dans l'expérience 2

Le choix des facteurs manipulés et des variables perceptuelles étudiées répond alors à une volonté de compléter l'analyse réalisée dans le cadre de la première expérience, tout en favorisant les recoupements possibles. Nous manipulerons ainsi deux facteurs (dont un non étudié dans l'expérience 1) et analyserons leur impact sur l'évaluation de la sollicitation, résumée par la valeur perçue (i.e. le gain net perçu à répondre), variable non mesurée précédemment. De plus, nous ne nous contenterons pas ici d'étudier l'attitude du consommateur face à la communication de ses données puisque nous irons jusqu'à mesurer les intentions comportementales de réponse. En cela, nous souhaitons vérifier que l'attitude est bien un médiateur de l'effet des croyances sur les intentions comportementales, résultat prédit par la théorie de l'action raisonnée (TRA) mais démenti par l'étude menée par Davis, Bagozzi et Warshaw (1989). Enfin, nous mesurerons aussi l'impact de la sollicitation sur l'intention du consommateur de poursuivre la relation avec l'entreprise l'ayant sollicité, vérifiant ainsi les hypothèses prédites par la théorie de l'engagement.

L'objectif de l'expérience 2 sera aussi de prendre en compte les limites identifiées dans le cadre de la première expérience et de faire en sorte d'y remédier. Deux limites sont plus particulièrement visées dans la mesure où elles tendent à limiter à la fois la validité interne et

externe de nos premiers résultats. Il s'agit, d'une part, du mode de recueil des données (questionnaire papier-crayon) et, d'autre part, du type d'échantillon interrogé (étudiants) et du contexte de mise en situation (sollicitation de données personnelles par le biais d'un jeu-concours organisé par un opérateur de téléphonie mobile).

Concernant le premier, Belk (1975) suggère que le meilleur moyen d'élaborer un scénario de nature expérimentale est de mêler la description écrite de la situation (pour ce qui est de la perspective temporelle et des tâches à accomplir) à un stimulus visuel et/ou audio pour les dimensions sociales et physiques (environnement dans lequel l'individu se trouve). Pour cet auteur, quand on cherche à simuler une sollicitation de données personnelles sur Internet, interroger les répondants sur poste informatique (et non par un questionnaire papier) permettrait de mieux tester l'effet de la situation. C'est donc le dispositif que nous mettrons en place dans le cadre de cette deuxième expérience. De plus, les répondants seront interrogés chez eux ou à leur travail, c'est-à-dire dans un environnement qui leur est familier. Cette solution doit permettre d'augmenter leur niveau d'implication et donc la qualité de leurs réponses, tout en introduisant de nouveaux biais, liés à la difficulté de contrôler les caractéristiques de cet environnement (type de matériel et de connexion utilisés, bruit dans la pièce, etc, …) et à la plus grande possibilité de refus de participer (biais de non-réponse). Pour le premier biais (manque de contrôle), nous faisons l'hypothèse (toutefois réaliste) que l'éventuelle hétérogénéité des répondants concernant ces critères se répartira équitablement entre les traitements. Celui-ci devrait donc être, au moins en partie, maîtrisé. En ce qui concerne le second biais (biais de non réponse), nous avons déjà indiqué au chapitre 8, qu'il serait inévitablement présent dans le cadre de l'expérience 2. Nous pensons toutefois qu'il sera réduit, de par la nature même des individus interrogés (panélistes), car ils se sont inscrits pour participer à des enquêtes[236].

Le choix d'interroger des panélistes répond, en outre, à une volonté de s'affranchir des contraintes (et des limites) liées à la sélection d'un échantillon de convenance, comme c'était le cas dans l'expérience 1. Selon Ferber (1977), ce type d'échantillon est en effet dommageable pour deux raisons majeures. Tout d'abord, les gens sélectionnés ne conviennent pas forcément pour étudier le comportement en question (en l'occurrence l'acceptation de fournir ou non les données demandées). Les étudiants interrogés précédemment représentent en effet la première génération de jeunes ayant grandi avec Internet. Il est donc possible qu'ils ne voient pas la sollicitation de données en ligne comme un problème majeur et que cela ait

[236] Cette raison (habitude de répondre à des enquêtes) constitue toutefois un nouveau biais que nous évoquerons à la fin de ce chapitre dans les limites inhérentes à cette deuxième expérience.

eu des répercutions sur nos résultats. Leur niveau de préoccupation dans ce domaine pourrait, en outre, être différent de celui ressenti par une population plus âgée ayant parfois une plus faible expérience du web. La seconde raison est d'ordre « technique » : on est en effet face à une absence de fond théorique et statistique pour pouvoir généraliser les résultats à une population plus conséquente[237]. Le fait d'interroger des individus, choisis aléatoirement auprès d'un panel d'internautes, permet alors de pallier, en partie, ces limites.

Enfin, la dernière limite à laquelle nous souhaitions remédier concerne le contexte de sollicitation étudié dans le cadre de la première expérience. En ce qui concerne le choix du secteur d'activité (i.e. la téléphonie mobile), nous le conservons pour l'expérience 2[238], afin de pouvoir comparer (si cela est possible) les résultats avec ceux trouvés dans l'expérience 1.
Il s'agit, d'ailleurs, d'un secteur relativement « impliquant » pour l'ensemble de la population (la majorité des adultes ont un téléphone portable et l'utilisent quotidiennement), ce qui permet de créer un scénario à la fois crédible et réaliste[239]. Le contexte dans lequel les données sont sollicitées constituera, pour sa part, un facteur manipulé (à trois niveaux) plutôt que d'être « fixé », dès le départ, à un niveau spécifique (i.e. celui d'un jeu-concours).
L'ensemble des facteurs manipulés dans le cadre de cette deuxième expérience ainsi que les variables qui seront contrôlées est, à présent, évoqué.

2 Le design expérimental

Le design proposé dans le cadre de cette deuxième expérience est un plan factoriel complet de type 2x3[240] (deux facteurs manipulés, l'un à deux niveaux, l'autre à trois) en inter-sujet (chaque répondant n'est confronté qu'à un seul des 6 traitements possibles). De plus, comme nous l'avons déjà fait dans l'expérience 1, les deux autres facteurs présentés dans le modèle conceptuel et non manipulés ici, seront fixés à un niveau identique pour tous les répondants. La familiarité sera ainsi fixée à un niveau élevé (les répondants seront ainsi sollicités par leur opérateur actuel de téléphonie mobile) tandis que les mentions de RVP seront limitées à leur strict minimum (consentement concernant l'utilisation ultérieure des données par le biais de cases pré-cochées, absence d'informations et de logo garantissant la sécurité des transactions).

[237] La pratique qui consiste à interroger des étudiants est cependant courante dans les recherches en général et en marketing en particulier, celle-ci pouvant se défendre, notamment, par des questions de coût et de logistique.
[238] Malgré les limites à la validité externe que ce choix occasionne.
[239] Il y a alors peu de chances de devoir éliminer des répondants parce qu'ils ne s'identifieraient pas à la situation proposée.
[240] Nous y inclurons aussi un facteur bloqué (le sexe) dont la justification sera présentée ci-après.

Nous exposerons, dans un premier temps, la manière dont les facteurs choisis ont été concrètement manipulés (2.1) avant d'évoquer les variables qui seront contrôlées (2.2).

2.1 Les facteurs manipulés

Comme indiqué précédemment, 2 facteurs seront manipulés dans le cadre de l'expérience 2 :

1) la quantité de données demandées (déjà étudié dans l'expérience 1)
2) le contexte de sollicitation (non étudié jusqu'ici)

Nous avons déjà expliqué, dans le chapitre présentant le plan expérimental (i.e. chapitre 8), pourquoi ces deux facteurs étaient manipulés conjointement. Nous rappelons que la littérature sur le dévoilement de soi (exposée au chapitre 3), tout comme la théorie de l'échange social (étudiée au chapitre 7), indiquent que l'acceptation des consommateurs de fournir des données personnelles est, notamment, fondée sur un calcul des coûts et bénéfices à le faire.
Les entreprises peuvent donc « jouer » sur ce rapport pour encourager leurs clients à se dévoiler. Elles peuvent ainsi, soit réduire les coûts (par exemple en réduisant le nombre de données à fournir), soit réfléchir au type de contexte favorisant le dévoilement et/ou aux avantages à proposer en échange. La manipulation de ces 2 facteurs (Quantité et Contexte) et le test de leur effet sur la valeur perçue s'avèrent donc particulièrement prometteurs.

Nous ne reviendrons pas ici sur la nécessité de choisir des niveaux de traitement les plus différents possibles (pour obtenir un maximum de variance), puisque nous en avons déjà parlé dans le chapitre précédent. Nous rappelons cependant que notre objectif est de parvenir à un scénario crédible et réaliste avec des manipulations de facteurs intégrées par les répondants.

Nous précisons que la manière de présenter le scénario aux répondants sera très similaire à celle utilisée dans le cadre de l'expérience 1, même si l'interrogation a lieu cette fois-ci en ligne (vs. questionnaire papier crayon dans l'expérience précédente). Les répondants seront ainsi confrontés à un cas de sollicitation de données personnelles par leur opérateur actuel de téléphonie mobile (dans un des différents contextes possibles, selon le niveau auquel ils auront été affectés). Ils seront ensuite soumis à un formulaire de données identique à celui auquel ils auraient été confrontés s'ils s'étaient trouvés réellement dans cette situation.

Comme dans l'expérience 1, les six formulaires possibles[241] (correspondant aux différents traitements) auront été créés au préalable par nos soins, sous la forme de pages Internet. Selon le scénario auquel les répondants auront été affectés et le nom de l'opérateur auprès duquel ils ont leur compte (une des premières questions posées visait en effet à connaître le nom de cet opérateur[242]), ils se verront alors proposer un des formulaires possibles, sous la forme d'une image insérée dans le questionnaire. Pour exemple, nous présentons, en Annexe 11.1, les six pages possibles correspondant à l'opérateur Orange. Toutefois, à la différence de l'expérience 1, ils ne pourront pas remplir ce formulaire puisqu'il s'agit d'une image. Ce choix offre l'avantage de réduire le temps de passation du questionnaire déjà assez élevé (10 à 15 minutes), tout en ayant l'inconvénient de réduire le réalisme de la situation. De plus, nous avions expliqué dans le chapitre précédent que le fait de ne pas faire remplir le formulaire risquait d'être moins impliquant et donc de réduire l'effectivité des manipulations expérimentales, notamment en ce qui concerne le facteur « Quantité de données demandées ». En effet, même si visuellement, on fait bien la différence entre un formulaire court (5 données) et un formulaire long (20 données), cet effet est réduit quand il s'agit d'une image (sous la forme d'une copie écran)[243]. De plus, un aperçu de cette image est loin d'être équivalent (notamment d'un point de vue cognitif) au fait de lire chaque champ un par un et de les remplir (ce qui est le cas quand on remplit réellement un formulaire sur Internet), ce qui peut avoir une incidence sur nos résultats. Pour pallier, en partie, ce biais, nous avons alors décidé de demander aux répondants d'indiquer le nombre de champs auxquels ils auraient répondu s'ils avaient réellement eu ce formulaire à remplir. Ceci offrait l'avantage de les obliger à compter le nombre de champs figurant sur le formulaire (et donc d'assurer une « meilleure » manipulation du facteur quantité), tout en donnant une mesure (même approximative) de la variable « Réponse » (notée i dans notre cadre conceptuel).

Dans ce qui suit, nous revenons sur l'opérationnalisation des deux facteurs manipulés dans le cadre de cette expérience, brièvement pour ce qui est de la quantité de données (puisque la manipulation est identique à celle effectuée dans le cadre de l'expérience 1) et de manière plus approfondie en ce qui concerne le contexte de sollicitation jamais manipulé jusqu'ici.
Là encore, même si plusieurs pré-tests ont été nécessaires avant d'aboutir à l'opérationnalisation finale, nous insisterons sur celle à laquelle nous avons abouti.

[241] En réalité, 18 pages ont été créées correspondant aux 3 types d'opérateurs possibles en matière de téléphonie mobile (Orange, SFR et Bouygues Telecom), le logo de l'opérateur apparaissant en haut du formulaire.
[242] La réponse à cette question était alors enregistrée, de manière à affecter le formulaire avec le bon logo.
[243] Il est cependant renforcé par le fait que l'image du questionnaire court tient sur un écran, alors que dans le questionnaire long, les répondants devaient actionner la souris pour voir le bas du formulaire.

2.1.1 Opérationnalisation du facteur « Quantité »

Pour rappel, le premier facteur manipulé dans le cadre de l'expérience 2 (i.e. la quantité de données demandées) est opérationnalisé à travers le nombre de champs à compléter dans le formulaire présenté (cf. tableau 11.1). Comme dans l'expérience 1, deux niveaux seront opposés : une quantité faible (5 données demandées) et une quantité élevée (20 données demandées). Il s'agit donc d'une manipulation par événement puisque visualisable par la présence (quantité élevée) ou l'absence (quantité faible) de certains champs sur la page écran fournie (cf. encadrés 11.1 et 11.2). Nous ne revenons pas sur les raisons de ces choix qui ont déjà été exposés au chapitre précédent. Nous précisons toutefois que deux modifications mineures sont intervenues par rapport à la manipulation de ce facteur dans l'expérience 1. La première concerne le passage du champ « date de naissance » dans la catégorie « votre profil » et la visualisation d'un champ de réponse approprié (de format jj/mm/aaaa)[244]. La seconde concerne le champ « pour inviter un de vos proches à participer au jeu », remplacé par « pour proposer cette offre à un de vos proches »[245].

Tableau 11.1 Opérationnalisation de la Quantité de données demandées (Facteur 1)

Niveau	Opérationnalisation
Faible	5 champs à remplir : nom, adresse, code postal, ville, email
Elevé	20 champs à remplir répartis en 3 catégories : - « vos coordonnées » (correspondant aux données d'identification) - « votre profil » (correspondant aux données sociodémographiques) - « vos habitudes et préférences »

Encadré 11.1 Extrait du formulaire pour le traitement « Quantité faible »

[244] Cela rendait la manipulation plus visuelle, d'autant que les répondants ne faisaient qu'observer le formulaire.
[245] La justification de ce changement est évidente étant donnée les différents contextes testés dans l'expérience 2.

Encadré 11.2 Extrait du formulaire pour le traitement « Quantité élevée »

orange

Vos coordonnées

les champs suivis d'un* sont obligatoires, vous devez les compléter pour valider votre inscription

Civilité* ○ Mme ○ Mlle ○ M
Nom*
Prénom*
Adresse*
Code postal*
Ville*
Email*
Téléphone*

Votre profil

Date de naissance / /
Nationalité*
Niveau d'études*
Profession*
Situation familiale*
Nombre d'enfants (si vous n'en avez pas, mettez 0)

Vos habitudes et préférences

Hobbies/ centres d'intérêt*
Combien de temps passez vous au téléphone par mois ?*
A quand remonte l'achat de votre téléphone actuel ?*
Utilisez-vous des SMS ?* ○ Oui ○ Non
Pour proposer cette offre à un de vos proches, merci d'indiquer son email*
Revenus annuels en 2004 (si vous n'en avez pas, mettez 0)

Indiquez nous vos souhaits

Je désire recevoir vos offres par email ◉ Oui ○ Non
Je désire recevoir les bons plans de vos partenaires ◉ Oui ○ Non

2.1.2 Opérationnalisation du facteur « Contexte de sollicitation »

Comme nous l'avons suggéré à travers les résultats de notre étude descriptive préalable (exposés au chapitre 6), le contexte dans lequel les données sont sollicitées pourrait jouer un rôle majeur dans la décision de l'individu de fournir ou non les renseignements. C'est d'ailleurs une des raisons pour lesquelles nous avons décidé de manipuler ce facteur dans le cadre de notre dispositif expérimental. Dans cette étude exploratoire (étude 2), plusieurs contextes de sollicitation étaient ainsi étudiés : l'enregistrement préalable à l'entrée sur le site, l'inscription en vue d'obtenir un document ou une information (par exemple, pour recevoir un catalogue, dans le cas d'un site de voyages), la discussion sur un forum, l'abonnement à une newsletter, l'achat en ligne, la participation à un jeu-concours. Nous avons déjà indiqué que nous excluions de ce travail le contexte de l'achat en ligne. De plus, celui de l'enregistrement préalable à l'entrée sur le site, bien que courant outre-Atlantique est quasi inexistant en France, ce qui nous amène aussi à l'exclure. La discussion sur un forum constitue un cas très particulier de dévoilement, qui allie un aspect interpersonnel au côté commercial (qui nous intéresse ici). Nous choisissons donc également de ne pas l'étudier. Enfin, en ce qui concerne

l'inscription en vue d'obtenir un document/une information, s'il s'agit d'un contexte très courant de sollicitation de données personnelles dans certains secteurs (notamment dans celui du voyage), il est beaucoup moins employé dans d'autres.

Pour des raisons de validité externe, nous ne souhaitions pas étudier un contexte particulier de collecte mais plutôt un ensemble de contextes couramment rencontrés en cours de navigation. Notre observation des pratiques de 35 sites de voyages (déjà évoquée au chapitre précédent) donne, à cet égard, les résultats suivants :

- l'inscription à une newsletter est proposée dans 30 cas sur les 35 étudiés
- s'inscrire à un programme de fidélité ou sur un espace membre est possible sur 9 sites
- 8 sites proposent de s'inscrire pour obtenir des renseignements et/ou des brochures
- seuls 5 sites proposaient à l'époque de participer à un jeu-concours.

Pour des questions de réalisme, nous exigions également que ce soit des contextes courants sur les sites d'opérateurs de téléphonie mobile et correspondants à des cas de familiarité élevée. L'observation des pratiques des trois opérateurs majeurs montre que le jeu-concours, la newsletter et la création d'un espace personnel (cf. Encadré 11.3 ci-dessous) sont les situations de sollicitations de données personnelles qui reviennent le plus souvent.

Encadré 11.3 Exemple de procédure de sollicitation de données dans le cadre de l'inscription à un Espace Personnel sur le site de l'opérateur Orange

Ce sont donc les trois « niveaux » du facteur « Contexte » que nous retiendrons dans le cadre de cette deuxième expérience, ce choix ayant en outre été validé auprès de professionnels et suite à plusieurs pré-tests auprès de consommateurs. Ces derniers confirment qu'il s'agit de contextes à la fois courants et réalistes, dans le cadre d'une sollicitation de données personnelles sur le site de leur opérateur de téléphonie mobile. Enfin, nous estimons *a priori* qu'il s'agit là de cas suffisamment différents pour donner un maximum de variance[246].

Comme pour la familiarité dans l'expérience 1, la manipulation du contexte dans le cadre de l'expérience 2 se fait par le biais d'instructions fournies aux répondants. Une page entière du questionnaire est ainsi vouée à la présentation de la situation (et donc du contexte de sollicitation). De plus, comme dans l'expérience 1, nous renforçons cette manipulation « écrite », par une manipulation visualisable, à travers le rappel du contexte en haut de la page du formulaire à remplir (présentée sous la forme d'une image insérée dans le questionnaire).

L'opérationnalisation définitive du facteur « Contexte », qui comprend 3 « niveaux »[247] (i.e. le Jeu-concours, la Newsletter et la création d'un Espace Personnel) est résumée dans le tableau ci-dessous. Nous proposons en Annexe 11.2 les copies écrans des questionnaires fournis aux répondants pour chacun des six traitements possibles. La manipulation du contexte ayant lieu à deux niveaux (en page 6 du questionnaire, qui présente la situation et à travers le formulaire à remplir en page 7 du questionnaire), deux encadrés seraient nécessaires pour chaque niveau. Nous ne présentons alors ci-dessous que celle réalisée dans le cadre du « Jeu-Concours » et renvoyons à l'Annexe 11.2 pour les deux autres niveaux (i.e. Newsletter et Espace Personnel).

Tableau 11.2 Opérationnalisation du Contexte de sollicitation (Facteur 2)

Niveau	Opérationnalisation
Jeu-concours (1)	<u>En page 6 du questionnaire :</u> - scénario décrivant une sollicitation de données par le biais d'un jeu-concours - répétition des termes « Jeu-Concours » (3 fois) ou « Jeu » (2 fois) - énoncé des avantages liés à la fourniture de données dans ce contexte (nombreux lots, 45000 euros, Safari au Kenya, …) <u>En page 7 du questionnaire :</u> - rappel de la situation : « dans le cadre de l'inscription au Jeu-Concours » - rappel du contexte sur la page écran du formulaire

[246] Cette hypothèse est cependant contredite par les résultats de notre étude descriptive préalable, les internautes interrogés se comportant de manière quasi similaire face à une newsletter et un jeu-concours (cf. tableau 6.9, chapitre 6). Nous choisissons cependant de ne pas tenir compte de ce constat suite aux pré-tests menés.

[247] Nous mettons le terme « niveau » entre guillemets car s'il est approprié pour désigner un cas faible et un cas élevé, il l'est moins en ce qui concerne les différents contextes possibles de sollicitation proposés ici.

Newsletter (2)	Hormis les avantages liés à la fourniture de données dans ce contexte (i.e. « informations sur la téléphonie mobile, conseils et astuces pour mieux utiliser [votre] mobile) », les autres moyens d'opérationnaliser ce niveau sont identiques au cas précédent. Seuls les termes changent : « Newsletter » remplace ici « Jeu-Concours »
Espace Personnel (3)	Idem cas précédent. Les avantages liés à la fourniture de données dans ce contexte sont les suivants : « informations sur [votre] compte, [..], offres personnalisées ». Le terme « Espace Personnel » remplace ici celui de « Newsletter »

Encadré 11.4 Page 6 du questionnaire correspondant au traitement « Contexte Jeu-concours »

IMAGINEZ LA SITUATION SUIVANTE !

Vous vous rendez sur le site Internet de VOTRE OPERATEUR de téléphonie mobile (celui de votre opérateur principal si vous en avez plusieurs) afin d'obtenir des informations sur les nouveaux services qu'il propose et les tarifs en vigueur.

Arrivé(e) sur la page d'accueil du site de VOTRE OPERATEUR, vous remarquez un bandeau qui défile vous proposant de participer à un JEU CONCOURS afin de gagner de NOMBREUX LOTS d'une valeur totale de 45 000 euros (295 000 francs), dont un SAFARI au Kenya, des téléphones portables et des abonnements.

Comme vous avez un peu de temps devant vous, vous décidez de voir en quoi consiste ce jeu.

Vous cliquez alors sur " JEU-CONCOURS " et tombez sur une nouvelle page comprenant un formulaire à remplir pour participer au jeu.

Ce formulaire se trouve page suivante.

Regardez TRES ATTENTIVEMENT ce formulaire
et IMAGINEZ que vous deviez le COMPLETER pour participer au Jeu-Concours.

Encadré 11.5 Page 7 du questionnaire correspondant au traitement « Contexte Jeu-concours »

Panel On The Web

Enquête

Taux d'avancement

FORMULAIRE A COMPLETER

Voici le FORMULAIRE que vous auriez eu à remplir dans le cadre de l'inscription au JEU-CONCOURS
Plusieurs questions concernant ce FORMULAIRE vont vous être posées par la suite.
REGARDEZ donc TRES ATTENTIVEMENT ce FORMULAIRE avant de répondre à la question qui suit.

Fichier Edition Affichage Favoris Outils ?

Pour participer à notre grand JEU-CONCOURS
Et peut être gagner un des nombreux LOTS en jeu
Dont un SAFARI au Kenya
COMPLETEZ le formulaire ci-dessous

SFR

Vos coordonnées

les champs suivis d'un* sont obligatoires, vous devez les compléter pour valider votre inscription

Nom*
Adresse*
Code postal*
Ville*
Email *

Indiquez nous vos souhaits

Je désire recevoir vos offres par email — Oui Non
Je désire recevoir les bons plans de vos partenaires — Oui Non

Terminé — Intranet local

Q05 - Dans la réalité, à combien de questions de ce FORMULAIRE auriez-vous répondu ?

Comme nous l'avons déjà indiqué précédemment, en plus de ces deux facteurs manipulés, nous y ajoutons un facteur bloqué (le sexe du répondant) à deux modalités (homme/femme). Un design expérimental avec facteurs bloqués est théoriquement utilisé quand on soupçonne que des variables exogènes identifiables et contrôlables peuvent modifier la relation étudiée (Keppel 1991). Conformément à cette position théorique, si l'on bloque des facteurs, c'est pour contrôler leur impact et non pas pour mesurer leur effet. Cependant, Iacobucci (2001) indique que pour tester l'impact de variables individuelles comme l'âge ou le genre, les chercheurs peuvent être amenés à bloquer ces facteurs. En effet, la sélection aléatoire des répondants ne garantit pas forcément une bonne répartition sur ces critères. Dès lors, afin de nous prémunir contre un risque de sous ou de sur représentation de la population féminine au sein de l'échantillon (comme ce fût le cas dans l'expérience 1), nous avons décidé de bloquer le sexe dans le design expérimental de cette deuxième expérience.

Si nécessaire, les plans expérimentaux peuvent être superposés pour élaborer le design de la recherche (Evrard, Pras et Roux 2003). En l'occurrence, le design complet de l'expérience 2 repose sur un plan factoriel 2x3 avec un facteur bloqué (le sexe). Pour résumer, les six traitements prévus, auxquels s'ajoute l'impact du sexe, sont présentés ci-dessous :

Tableau 11.3 Design expérimental complet de l'expérience 2

Facteur bloqué (sexe)	Facteurs manipulés	Contexte		
	Quantité	Jeu-Concours (C1)	Newsletter (C2)	Espace Personnel (C3)
Hommes	Faible (Q1)	Traitement Q1C1	Traitement Q1C2	Traitement Q1C3
	Elevée (Q2)	Traitement Q2C1	Traitement Q2C2	Traitement Q2C3
Femmes	Faible (Q1)	Traitement Q1C1	Traitement Q1C2	Traitement Q1C3
	Elevée (Q2)	Traitement Q2C1	Traitement Q2C2	Traitement Q2C3

Nous précisons pour finir que, comme pour l'expérience 1, chaque formulaire dispose du même design et de la même charte graphique, de manière à éviter tout biais inhérent à une perception différente de ces pages par les individus interrogés.

Voyons, à présent, les variables dont nous contrôlerons l'effet dans le cadre de l'expérience 2.

2.2 Les variables contrôlées

Nous détaillons ici le type de variable à contrôler (2.2.1) et le type de contrôle réalisé (2.2.2) sans revenir aux justifications théoriques d'un tel contrôle, déjà exposées auparavant.

2.2.1 Le type de variable à contrôler

Dans le cadre de cette deuxième expérience, quatre types de variables exogènes seront contrôlés. Elles concernent le niveau d'implication du répondant (IMPL) vis-à-vis de la catégorie de produits/services concernée (en l'occurrence la téléphonie mobile), son origine sociodémographique (âge, sexe, CSP, niveau d'études), son expérience (EXP) et son utilisation (UTIL) du média Internet et son comportement habituel de réponse (COMP_HAB). Par rapport à la première expérience où nous interrogions une population homogène, nous contrôlerons donc l'ensemble des variables sociodémographiques. De plus, nous contrôlerons aussi le comportement habituel de réponse parce que nous soupçonnons son influence sur les intentions comportementales de répondre et de mentir. Deux types de comportements habituels seront ainsi distingués : l'habitude de répondre et l'habitude de mentir.

2.2.2 Le type de contrôle

Trois types de contrôles seront utilisés dans le cadre de cette expérience :
- une affectation aléatoire des individus interrogés aux différents traitements, qui permet une répartition *a priori* homogène des influences externes sur les unités-tests ;
- un contrôle *a priori* pour les variables de genre (facteur bloqué) et d'âge[248]
- un contrôle statistique *a posteriori* sur l'ensemble des variables exogènes identifiées ci-dessus par introduction dans les analyses en tant que co-variables.

Nous détaillons ci-dessous les instruments de mesure adoptés pour chaque variable exogène contrôlée. Nous ne revenons pas sur les échelles déjà utilisées dans le cadre de l'expérience 1 et détaillons plutôt celles correspondant aux variables nouvellement contrôlées.

[248] Il a été fait en sorte que les individus des différentes classes d'âge se répartissent équitablement au sein des traitements. C'est d'ailleurs la raison pour laquelle les questions portant sur l'âge et le sexe du répondant se trouvent dès la page 2 du questionnaire. La répartition finale selon ces deux critères est fournie en Annexe 11.3.

Tableau 11.4 Instruments de mesure utilisés pour les variables externes contrôlées

Variables	Items utilisés	Types de mesure
Niveau d'implication (IMPL)	Echelle adaptée de Strazzieri (1994) (3 items)	Likert en 7 points
Age	15-24 ; 25-34 ; 35-44 ; 45-54 ; 55-64 ; 65 ans et plus (1 item)	Variable ordinale à 6 modalités
Sexe	H / F (1 item)	Variable nominale (binaire)
CSP	Agriculteur, profession intermédiaire, …[249] (1 item)	Variable nominale[250]
Niveau d'études	Inférieur au Bac, Bac à Bac+2, Bac+3 ou 4, Bac+5 et plus (1 item)	Variable ordinale à 4 modalités
Expérience d'Internet (EXP)	Ancienneté d'utilisation (1 item)	Variable ordinale à 3 modalités
Utilisation d'Internet (UTIL)	- fréquence de consultation de la messagerie (1 item) - fréquence de surf sur le web (1 item)	Variables ordinales à 4 modalités
Habitude de répondre (CHR)	- « quand une entreprise me demande des informations personnelles, je les lui donne facilement » (1 item)	Likert en 7 points (de 1, pas du tout d'accord, à 7, tout à fait d'accord)
Habitude de mentir (CHM)	« si je réponds, je dis toujours la vérité » (1 item)	Likert en 7 points

De plus, nous rappelons que nous « contrôlons » également les deux autres facteurs non manipulés dans cette expérience (les mentions de RVP et la familiarité) par fixation à un niveau défini *a priori* et identique à tous les répondants.

Nous présentons ci-après le modèle ainsi que les hypothèses que nous serons amenés à tester dans le cadre de cette deuxième expérience.

[249] Actifs occupés (selon la Classification de l'INSEE), auxquels s'ajoute la population inactive (étudiants, …)
[250] En réalité, cette variable sera recodée en variable ordinale à 4 niveaux (inactifs, CSP -, CSP moyen et CSP +).

3 Présentation du modèle à tester et des hypothèses de recherche

3.1 Présentation du modèle à tester

Rappelons que le modèle « complet » présenté au chapitre 8 (cf. figure 8.2) est testé en trois parties, correspondant aux trois expériences menées successivement.

Dans cette deuxième expérience, l'objectif est de mesurer l'effet de l'exposition à la sollicitation - par l'intermédiaire de deux facteurs manipulés (i.e. Quantité et Contexte) - sur l'évaluation de celle-ci (résumée à travers la valeur perçue à répondre). L'impact de cette évaluation et de la préoccupation pour le RVP sur l'attitude face à la communication des données est également quantifié. Enfin, nous évaluons l'influence de l'attitude sur les intentions comportementales de réponse et l'impact de ces dernières sur l'intention de poursuivre la relation avec l'entreprise sollicitante.

Pour rappel, les deux facteurs manipulés ici (Quantité de données demandées et Contexte de sollicitation) porteront les numéros « 1.3 » et « 1.4 » puisqu'ils correspondent respectivement au troisième et quatrième facteur manipulé dans le cadre de notre plan expérimental.

Le modèle à tester dans le cadre de cette expérience se présente donc de la manière suivante (cf. figure 11.2 page suivante) :

Ce modèle comprend plusieurs hypothèses à tester qui ont, soit déjà été formulées dans le chapitre 7 (cas des hypothèses H2d, H3e, H4 et H5), soit seront formulées dans la section qui suit (cas des hypothèses H1.3 d et H1.4 d, correspondant à la proposition P1 du chapitre 7).

Figure 11.2 Modèle à tester dans le cadre de l'expérience 2

Nous présentons maintenant les hypothèses spécifiques à cette deuxième expérience.

3.2 Formulation des hypothèses de recherche

Nous formulons ici les hypothèses H1.3d et H1.4d, correspondant à la proposition P1 du chapitre 7, relative à l'effet des deux derniers facteurs manipulés (Quantité et Contexte) sur la valeur perçue à répondre (évaluation de la sollicitation). Nous ne revenons pas ici sur les hypothèses H2d, H3e, H4 et H5 qui ont déjà été formulées et justifiées au chapitre 7.
Nous exposons tour à tour les hypothèses relatives à :
- l'influence de la quantité sur la valeur perçue à répondre (H1.3 d[251])
- l'influence du contexte sur la valeur perçue à répondre (H1.4 d)

3.2.1 L'influence de la quantité sur la valeur perçue à répondre

Implicitement incluse dans la notion de « *privacy calculus* »[252] est l'idée selon laquelle les individus se comportent de manière à maximiser la différence entre les bénéfices et les coûts liés au dévoilement de données personnelles. Cette idée est d'ailleurs largement reprise dans la théorie de l'échange social et dans la théorie de la motivation ou de la valeur attendue (*Expectancy theory*), toutes deux exposées au chapitre 7. Dans ce travail doctoral, nous avons choisi de mesurer cette évaluation coûts/bénéfices par le concept de « valeur perçue ». Dans le domaine du marketing relationnel et du « *permission marketing* », il a été montré que les consommateurs estiment généralement les coûts de transmission (dévoilement) et de mise à jour de leurs informations personnelles très onéreux (Blattberg et al. 1978, Oliva Oliver et MacMillan 1992). De plus, plus ce coût serait jugé onéreux, moins ils auraient intérêt à participer à un programme relationnel (ou de fidélité). Une action jugée coûteuse bénéficierait ainsi d'une évaluation moins favorable qu'une action peu onéreuse.

La charge cognitive liée au remplissage du formulaire est un des principaux coûts auxquels est confronté un individu qui se voit sollicité pour fournir des données personnelles[253].

[251] Pour rappel, le nom des hypothèses se justifie par le numéro attribué à chaque facteur manipulé (i.e. 1.3 pour le facteur « Quantité ») et par la lettre affectée à chaque variable médiatrice (i.e. d pour la « valeur perçue »).
[252] Abordée au chapitre 3.
[253] Les différents coûts liés au dévoilement de données personnelles ont été présentés au chapitre 9.

Il s'agit, de plus, d'un coût immédiat et non différé. Plusieurs recherches dans le domaine de la participation à un sondage (dont celle de Childers et Skinner 1996) soulignent qu'un questionnaire court réduit le montant d'effort nécessaire pour y répondre et donc décroît le coût de la participation. Cet effet est d'autant plus amplifié que le répondant compare souvent le temps dont il dispose avec le temps nécessaire pour accomplir la tâche proposée (remplir le questionnaire). A bénéfice égal, un consommateur préférerait donc un questionnaire court car cela réduirait le coût perçu à répondre et augmenterait ainsi la valeur perçue à répondre. Appliquées à la sollicitation de données personnelles par l'entremise d'un formulaire à remplir, l'ensemble de ces considérations nous amènent à formuler l'hypothèse suivante :

Hypothèse 1.3 (d) : *Plus la quantité de données demandées est élevée (vs. faible), moins (vs. plus) la valeur perçue à répondre (d) est élevée.*

3.2.2 L'influence du contexte sur la valeur perçue à répondre

Culnan et Bies (2003) estiment que les consommateurs acceptent de délivrer des données personnelles si la valeur des services qu'ils en retirent est supérieure à la perte de vie privée qui en découle. Childers et Skinner (1996) montrent aussi que les facteurs qui soulignent la présence d'une récompense (souvent monétaire) en échange de la participation à une enquête servent à accroître la coopération en établissant un sentiment de confiance basé sur l'idée d'un échange plus équilibré. Enfin, dans le domaine de la sollicitation de données personnelles, il a été montré que la fourniture de données en échange d'un service personnalisé serait moins associée à des connotations négatives que si cela se faisait contre de la publicité.

Tous ces résultats indiquent que les individus accordent une valeur différente au dévoilement selon le contexte dans lequel ils sont sollicités, parce qu'ils y associent des bénéfices et des risques différents. Une telle différence au niveau de l'attractivité du résultat devrait alors affecter leur volonté de partager (ou non) les données demandées.

Ceci nous amène à formuler l'hypothèse suivante :

Hypothèse 1.4 (d) : *La valeur perçue à répondre (d) diffère selon le contexte dans lequel les données sont sollicitées (i.e. Jeu-Concours / Newsletter / Espace Personnel)*

En résumé, voici l'ensemble des hypothèses à tester dans le cadre de l'expérience 2 :

Tableau 11.5 Récapitulatif des hypothèses à tester dans l'expérience 2

H 3 (e)[254] (10)	Plus la préoccupation pour le RVP est élevée (vs. faible), moins l'attitude face à la communication des données sollicitées (e) est favorable (vs. défavorable)
H 1.3 (d) (11)	Plus la quantité de données demandées est élevée (vs. faible), moins (vs. plus) la valeur perçue à répondre (d) est élevée
H 1.4 (d) (12)	La valeur perçue à répondre (d) diffère selon le contexte dans lequel les données sont sollicitées (i.e. Jeu-Concours / Newsletter / Espace Personnel)
H 2 (d) (13)	Plus la valeur perçue à répondre (d) est élevée (vs. faible), plus l'attitude face à la communication des données sollicitées est favorable (vs. défavorable)
H 4 (f, g) (14, 15)	Plus l'attitude face à la communication des données sollicitées est favorable (vs. défavorable), plus (vs moins) l'individu a l'intention de répondre (de communiquer les données demandées) (f) et moins (vs. plus) il a l'intention de mentir (g)
H 5 (16)	Plus (vs. moins) l'individu a l'intention de communiquer les données demandées, plus (vs. moins) il a l'intention de poursuivre la relation avec l'entreprise qui l'a sollicité

4 Recueil des données et caractéristiques des répondants

4.1 L'instrument de collecte et le recueil des données

L'instrument de mesure utilisé dans le cadre de l'expérience 2 est semblable à celui employé dans la première à deux exceptions près : d'une part, des questions spécifiques à ce design ont été introduites (tandis que celles non adaptées ici étaient supprimées) et, d'autre part, il a été programmé pour être administré en ligne, ce qui induit certaines spécificités.

Les répondants sollicités cliquaient ainsi sur le lien qui leur était fourni et se retrouvaient sur la page d'accueil du questionnaire. 16 écrans différents suivaient, avec pour objectif de se limiter à une ou deux questions par page afin de faciliter le remplissage[255]. Il s'agissait en effet d'éviter que l'internaute doive utiliser la souris pour répondre aux questions suivantes.

Nous indiquons brièvement ci-dessous l'organisation de ce questionnaire, dont nous rappelons qu'il peut être consulté en Annexe 11.2 :

- la page 2 contient les questions relatives à l'âge et au sexe du répondant afin d'affecter les individus aux différents traitements en toute connaissance de cause (contrôle *a priori*)

[254] Cette hypothèse ayant déjà été testée dans l'expérience 1, elle porte le même numéro (i.e. « 10 »). Les hypothèses suivantes sont numérotées à la suite (i.e. à partir de 11).

[255] Cette condition a été respectée, hormis en page 14 qui comprend 3 questions (assez courtes toutefois).

- en page 3, nous contrôlons que les répondants possèdent bien un téléphone portable (ceux pour lesquels ce n'est pas le cas sont alors stoppés à ce niveau et exclus de l'échantillon final)
- en page 4, se trouve la question portant sur l'opérateur, qui sera ensuite utilisée pour affecter le « bon » formulaire (celui contenant le logo approprié)
- les pages 6 et 7 contiennent respectivement les manipulations des deux facteurs (Contexte et Quantité), à travers, d'une part, la mise en situation du répondant (manipulation du Contexte) et, d'autre part, la copie écran du formulaire de données (manipulation de la Quantité). Au total, 18 pages « 7 » ont ainsi été programmées correspondant aux six traitements (design 2x3) et aux trois opérateurs possibles (Orange, SFR et Bouygues Télécom)
- en pages 8 et 9, se trouvent respectivement les mesures de l'attitude et de la valeur perçue. Afin de faciliter les réponses, une nouvelle copie du formulaire (en plus petit toutefois) est présentée systématiquement aux répondants, en haut de ces deux pages
- les pages 10 et 11 permettent de mesurer successivement les intentions comportementales, la préoccupation pour le RVP et le comportement habituel de réponse
- la page 12 contient les questions portant sur la vérification des manipulations expérimentales
- les pages 13 à 15 reprennent l'ensemble des variables de contrôle non mesurées jusque là
- les pages 16 et 17 contiennent respectivement une question ouverte (remarques à formuler) et des remerciements

Nous avons donc suivi les mêmes précautions que celles employées pour l'expérience précédente, à savoir, mesurer les variables dépendantes avant les variables médiatrices[256], de manière à ne pas contaminer les résultats, et vérifier l'effectivité de nos manipulations expérimentales à la fin, pour ne pas induire de biais dans les réponses.

L'administration du questionnaire a duré 10 jours, entre le 9 et le 19 mars 2006 (une relance ayant été programmée le 17 mars). Nous fournissons en Annexe 11.4 la distribution des questionnaires répondus par jour. On y voit que 43% des répondants ont répondu le premier jour, 20% ayant répondu entre le 13 et le 19 mars, soit dans les 7 derniers jours.
Pour programmer et administrer ce questionnaire, nous avons fait appel à une société d'études partenaire (Panel On The Web), qui a, en outre, fournit l'échantillon de répondants (sélectionnés aléatoirement parmi son panel d'internautes français). Au total, 468 participants ont été sollicités personnellement et par email pour répondre à ce questionnaire[257], auxquels il

[256] Le respect de cette exigence ne devait toutefois pas s'affranchir d'une certaine cohérence. Ainsi, les mesures des intentions comportementales sont intervenues après celles de l'attitude et de la valeur perçue.
[257] L'email de sollicitation ne fournissait toutefois aucune indication quant à la nature de l'enquête à laquelle ils étaient conviés. Nous éliminons ainsi le risque de biais de non-réponse dû aux problèmes de RVP.

faut retirer 17 adresses emails invalides. Sur les 451 individus restants, 157 n'ont pas répondu et 11 ont fourni des réponses incomplètes ou ont abandonné avant la fin. Enfin, 31 individus ont été éliminés suite aux questions de la page 2 car leur profil ne permettait pas d'homogénéiser la répartition des effectifs selon l'âge et le sexe du répondant. Au final donc, 252 questionnaires complets ont été conservés, soit un effectif de 42 individus par traitement. Ceci correspond alors à un taux de participation réelle de 62,7% (soit (252+31)/468), ce qui est tout à fait raisonnable et permet de limiter de façon substantielle le biais de non réponse.
Nous présentons, dans le paragraphe suivant, les caractéristiques de cet échantillon.

4.2 Les caractéristiques de l'échantillon

L'échantillon sur lequel a été menée cette étude est composé de panélistes. Nous revenons brièvement sur la justification d'un tel choix avant de présenter le profil des répondants.

Pour nous démarquer des contraintes d'un échantillon de convenance (et des biais que ce choix engendre), nous souhaitions, dans le cadre de cette deuxième expérience, interroger un échantillon qui se rapproche le plus possible de la population à laquelle nous voulons pouvoir généraliser nos résultats, à savoir l'ensemble des internautes français.
Les caractéristiques sociodémographiques de l'audience Internet sont encore très marquées, même si le phénomène s'amenuise, au fur et à mesure que le média se démocratise. Toutefois, l'obstacle peut être contourné en faisant appel à un panel d'internautes, ce qui permet un contrôle *a priori* de la composition de l'échantillon et en assure ainsi la représentativité selon les objectifs et la cible de l'enquête. Ceci convenait parfaitement à nos exigences car nous pouvions ainsi bénéficier d'un échantillon « représentatif » de la population, tout en contrôlant *a priori* sa composition selon certains critères spécifiques (âge et sexe).
La société d'études à laquelle nous avons fait appel a dès lors sélectionné (aléatoirement), au sein de son panel, un échantillon d'internautes susceptibles de participer à l'expérimentation. Cet échantillon, représentatif du panel, l'est donc aussi eu égard à la population totale. Connaissant *a priori* la taille voulue de l'échantillon final et le taux de participation moyen de ses panelistes, la société a ainsi sollicité 468 individus pour aboutir aux 252 répondants.

Comme nous l'avons exposé au chapitre 8, ce choix d'interroger des panélistes n'est pas sans poser certains problèmes. En effet, si cela réduit le biais de non réponse (on a vu que le taux de participation à cette étude était très élevé, ce qui limite sérieusement ce biais) et le biais de couverture (l'échantillon étant représentatif du panel, lui-même représentatif de la population

des internautes français), il n'a cependant aucune incidence sur le biais du répondant (façon dont il se comporte face au questionnaire). Ce dernier serait d'ailleurs d'autant plus élevé que les panélistes interrogés se sont inscrits auprès de cette société d'étude, acceptant par là de participer régulièrement à des enquêtes. On peut donc supposer que ce sont des individus qui donnent facilement des renseignements aux autres (mais donnent-ils pour autant des données personnelles ?) et qui ont donc un comportement particulier. La question sera alors de savoir si cela n'a pas trop d'incidence sur les résultats que nous pourrons obtenir. Nous reviendrons sur ce sujet quand nous aborderons les limites de l'expérience 2 (cf. section 7 de ce chapitre).

Nous indiquons ci-dessous le profil des répondants interrogés dans le cadre de l'expérience 2 :

Tableau 11.6 Caractéristiques de l'échantillon interrogé dans le cadre de l'expérience 2

Variables	Valeurs	Fréquences	Pourcentage
Variables sociodémographiques			
Sexe	Masculin	128	50,8%
	Féminin	124	49,2%
Age	15-24 ans	79	31,3%
	25-34 ans	72	28,6%
	35-44 ans	44	17,5%
	45-54 ans	33	13,1%
	55-64 ans	20	7,9%
	65 ans et plus	4	1,6%
CSP[258]	Employés / ouvriers	98	38,9%
	Elèves / étudiants	57	22,6%
	Autres inactifs	40	15,9%
Niveau d'études	< Bac	59	23,4%
	Bac à Bac + 2	122	48,4%
	Bac + 3 ou 4	46	18,3%
	Bac + 5 et plus	25	9,9%
Variables expérientielles (expérience et utilisation d'Internet)			
Expérience d'Internet	Moins de 2 ans	26	10,3%
	Entre 2 et 5 ans	86	34,1%
	Plus de 5 ans	140	55,6%
Fréquence d'utilisation de la messagerie	Plusieurs fois / jour	173	68,7%
	1 à 2 fois / jour	61	24,2%
	Plusieurs fois par semaine	18	7,1%
	Moins souvent	0	0,0%
Fréquence de surf	Plusieurs fois / jour	176	69,8%
	1 à 2 fois / jour	46	18,3%
	Plusieurs fois par semaine	27	10,7%
	Moins souvent	3	1,2%

[258] Nous ne faisons figurer ici que les catégories les plus représentées. Par ailleurs certaines catégories étant très peu représentées, nous avons dû les regrouper avec d'autres afin d'obtenir des effectifs suffisants. Ainsi, les « autres inactifs » correspondent aux personnes à la retraite, au foyer ou à la recherche d'un emploi.

Nombre d'achats déjà réalisés en ligne depuis qu'ils utilisent Internet	Aucun Moins de 5 De 5 à 20 Plus de 20	17 64 88 83	6,7% 25,4% 39,4% 32,9%
Variables liées à la téléphonie mobile			
Nom de l'opérateur	Orange SFR Bouygues télécom	92 98 62	36,5% 35,9% 24,6%

L'échantillon de panélistes interrogés est donc majoritairement jeune (60% ont moins de 34 ans) et de niveau d'études moyen (près de la moitié ont entre le Bac et Bac+2). A titre de comparaison, 57% des internautes français avaient moins de 35 ans en 2003[259]. En ce qui concerne le sexe, nous constatons une répartition quasi équilibrée entre hommes et femmes[260], qui correspond à la volonté de faire du genre un facteur bloqué au sein du design expérimental. Cette répartition se rapproche de celle observée sur la population internaute française dans son ensemble (53% d'hommes et 47% de femmes fin 2004[261]). De même, alors que les cadres, professions intermédiaires et étudiants représentaient 53% de la population internaute française en 2003, ils sont 42% dans notre échantillon. L'échantillon que nous avons interrogé se rapproche donc des caractéristiques des internautes français.

Cet échantillon de répondants est, en outre, très expérimenté sur Internet (55% ont plus de 5 ans d'ancienneté) et une grande majorité (70%) l'utilise quotidiennement. Enfin, plus de 65% y font régulièrement des achats (contre moins de 7% qui n'en n'ont jamais fait). Ce côté « expert » s'explique principalement par le statut des répondants (i.e. des panélistes) qui, malgré toute la volonté des sociétés d'études, sont plutôt des gens aguerris au média Internet.

En ce qui concerne l'utilisation de la téléphonie mobile, la répartition entre les trois opérateurs est un peu différente des parts de marché recensées par l'ART fin 2005 (respectivement 47,5% ; 35,6% et 17,1%), notre échantillon comptant moins d'abonnés à Orange et plus d'abonnés à Bouygues Télécom. Cette répartition ne nous semble toutefois pas de nature à créer un biais au niveau des résultats.

Après avoir présenté les caractéristiques de l'instrument de collecte et de l'échantillon interrogé, nous abordons à présent la partie concernant l'analyse des données.

[259] Source Reseune (2003), [en ligne] www.reseune.fr

[260] L'écart constaté (128 vs 124) correspond, d'une part, aux limites de l'aléatoire et, d'autre part, au contrôle *a priori* sur la variable âge. La volonté d'équilibrer les traitements sur ces 2 variables (âge et sexe) a pour contrepartie d'introduire des contraintes supplémentaires qui aboutissent à une répartition imparfaite. Cette répartition, bien que n'étant pas parfaitement homogène, peut néanmoins être considérée comme acceptable.

[261] Source Médiamétrie (2004), [en ligne]

5 L'analyse des données

Nous verrons, dans un premier temps, les conditions relatives à la validité du plan d'expérience (5.1), avant d'aborder la question de l'effectivité des manipulations réalisées (5.2). Nous indiquerons ensuite très rapidement la qualité psychométrique des instruments de mesure utilisés (5.3) pour terminer par la présentation du plan de traitement (5.4).

5.1 La validité interne du plan d'expérience

La vérification de la validité du plan d'expérience comporte 6 étapes majeures déjà présentées au chapitre précédent (i.e. chapitre 10). Nous abordons tour à tour chacune des ces conditions.

5.1.1 Le contrôle des variables exogènes

Le contrôle des variables exogènes, la manière de le réaliser et la justification d'un tel contrôle ont déjà été abordés précédemment dans ce chapitre. Nous ne détaillerons donc pas plus ce point et renvoyons les lecteurs au paragraphe 2.2 de ce chapitre. Nous précisons toutefois que le contrôle *a priori* des variables âge et sexe, vérifié *a posteriori*, est présenté en Annexe 11.3. Nous constatons que la répartition selon le sexe est proche de l'objectif fixé qui était d'obtenir un même nombre d'hommes et de femmes par traitement. Quelques différences apparaissent cependant (i.e. 23 vs. 19 obtenu contre 21 vs. 21 attendu), dues à la procédure de sélection aléatoire mise en place lors de la programmation du questionnaire. En ce qui concerne l'âge, de mêmes écarts apparaissent (i.e. 15 vs. 12 dans la classe 15-24 ans).
Ces différences ne nous semblent cependant pas en mesure de biaiser nos résultats.

5.1.2 L'égalité des traitements

Un design expérimental avec le même nombre d'individus par cellule est nécessaire puisque les traitements statistiques d'analyse de la variance reposent sur ce principe. Le test F de Ficher est, en effet, le rapport entre la somme des carrés du traitement (qui repose sur l'égalité des échantillons) et la somme des carrés de l'erreur (Keppel 1991).
Nous vérifions donc cette condition sur le design factoriel (2x3) et sur le design complet (avec le sexe en facteur bloqué).

Tout d'abord, la répartition de nos effectifs sur les 6 traitements possibles est bien identique.

Tableau 11.7 Répartition des effectifs par traitement

Traitements	Contexte		
Quantité	Jeu-Concours (C1)	Newsletter (C2)	Espace Personnel (C3)
Faible (Q1)	42	42	42
Elevée (Q2)	42	42	42
Total	84	84	84

Toutefois, le nombre d'individus par traitement et selon leur sexe est proche mais pas égal.

Tableau 11.8 Répartition des effectifs par traitement et par facteur bloqué (sexe)

Facteur bloqué (sexe)	Facteurs manipulés	Contexte		
	Quantité	Jeu-Concours (C1)	Newsletter (C2)	Espace Personnel (C3)
Hommes	Faible (Q1)	22	23	23
	Elevée (Q2)	21	20	20
Femmes	Faible (Q1)	20	19	19
	Elevée (Q2)	22	22	22
Total		84	84	84

Ainsi, malgré toutes les précautions prises, la procédure d'affectation des répondants aux différents traitements n'a, en effet, pas permis d'atteindre la répartition homogène prévue. Toutefois, puisque les écarts sont peu importants et que la répartition par traitement est, elle, identique, nous ne pouvons supprimer aléatoirement certains répondants ni même collecter de nouvelles données. Nous retenons alors la solution proposée par Iacobucci (2001) qui consiste à utiliser le mode « Type III » dans SPSS, qui calcule la somme des carrés du traitement sans la contrainte d'égalité des échantillons. De plus, cet indicateur étant donné par défaut par SPSS lors des ANCOVA, nous n'aurons pas de précaution particulière à prendre.

5.1.3 L'homogénéité des variances

Nous procédons au test d'égalité des variances de Levene pour chaque traitement expérimental, eu égard à l'ensemble des variables dépendantes concernées, à savoir la valeur perçue (VAL) et l'attitude face à la communication des données sollicitées (ATT).

Le résultat des tests d'égalité des variances de Levene, présenté en Annexe 11.5 (point A), ne conduit pas à rejeter l'hypothèse d'égalité des variances entre les traitements expérimentaux. Cette condition est donc respectée.

5.1.4 La normalité des variables

Nous examinons la normalité de l'ensemble des variables médiatrices et dépendantes du modèle à tester (valeur, attitude, intentions de répondre, de mentir et de poursuivre la relation), à travers les indices d'asymétrie et d'aplatissement. Pour qu'elles soient considérées comme normales, ces coefficients doivent être respectivement inférieurs à 1 et 1,5.

Les résultats fournis en Annexe 11.5 (point B) montrent la normalité de l'ensemble des variables médiatrices et dépendantes du modèle, exception faite de la variable intention de mentir (IM), qui viole à 3 reprises les critères cités précédemment. A deux reprises, il s'agit seulement d'un problème d'asymétrie, le coefficient correspondant étant légèrement supérieur à 1. Dans un cas (Quantité de données demandées faible), les coefficients d'asymétrie et d'aplatissement tous deux supérieurs aux seuils de 1 et 1,5 (respectivement 1,5 et 2,5). D'après Curran, West et Finch (1996), les données sont alors considérées comme modérément non normales puisque les coefficients sont respectivement inférieurs à 2 et à 7. Il n'y a donc pas anormalité au sens strict et nous ne pensons pas que cela puisse remettre en cause, outre mesure, l'introduction de cette variable au sein du modèle.

5.1.5 L'indépendance des observations

Nous disposons de 6 groupes expérimentaux ; ces groupes ne doivent pas se distinguer d'après leurs caractéristiques individuelles : implication, données sociodémographiques et expérientielles (ancienneté et utilisation d'Internet). Même si cette condition est censée être assurée *a priori* par l'affectation aléatoire des répondants aux différents traitements, un contrôle *a posteriori* est cependant nécessaire. Nous le vérifions par un test de chi-deux.

Les résultats de ce test sont fournis dans le tableau ci-après :

Tableau 11.9 Indépendance des observations en fonction des traitements expérimentaux

Caractéristiques	Quantité (faible vs élevée)	Contexte (jeu-concours /newsletter /espace personnel)
Niveau d'implication	8,669 (p = 0,967)	30,81 (p = 0,728)
Sexe	1,016 (p = 0,313)	0,032 (p = 0,984)
Age	2,380 (p = 0,794)	1,338 (p = 0,999)
CSP	3,112 (p = 0,874)	11,614 (p = 0,637)
Niveau d'études	1,148 (p = 0,765)	6,841 (p = 0,336)
Expérience d'Internet	1,155 (p = 0,561)	2,635 (p = 0,621)
Fréquence de consultation de la messagerie	2,161 (p = 0,339)	4,167 (p = 0,384)
Fréquence de surf	0,457 (p = 0,928)	8,983 (p = 0,175)

L'analyse des résultats de ce tableau montre que les échantillons ne se distinguent pas d'après leurs caractéristiques individuelles. Il y a donc bien indépendance des observations.

5.1.6 L'orthogonalité des variables manipulées

L'absence d'orthogonalité entre les facteurs manipulés est une condition importante que tout chercheur doit vérifier avant de construire un plan factoriel.

Dans le cadre de cette deuxième expérience, nous avons 2 facteurs manipulés : la quantité de données demandées et le contexte de sollicitation. Intuitivement, il n'y a *a priori* aucune raison que l'un de ces facteurs n'influence l'autre. Nous vérifions cependant cette condition par une analyse de corrélation, non pas entre les facteurs (puisqu'il s'agit de variables nominales) mais entre les perceptions de ces manipulations (Manipulation Checks) qui seront présentées dans le paragraphe suivant. Le coefficient de Pearson obtenu est de - 1,111 (p = 0,080). Puisqu'il n'y a pas de corrélation significative, il y a bien orthogonalité des facteurs.

5.2 La vérification des manipulations expérimentales

Comme dans l'expérience 1, nous souhaitons vérifier que les manipulations expérimentales ont été réalisées avec succès. Pour ce faire, plusieurs questions (« *Manipulation Checks* ») ont été posées en fin de questionnaire afin de mesurer les perceptions des répondants selon le traitement auquel ils étaient affectés.

Plus spécifiquement, 3 questions étaient posées pour chacun des facteurs[262]. Nous pouvions ainsi vérifier que le formulaire « quantité faible » semblait bien plus court que le formulaire « quantité élevée ».

Parce que nous disposons de 3 questions par facteur manipulé, nous commençons par réaliser une analyse factorielle pour chaque groupe d'items (suivie du calcul de l'Alpha de Cronbach) afin de vérifier que chaque item mesure bien ce qu'il est censé mesurer.

Les résultats de ces analyses sont fournis en Annexe 11.6 et résumés ci-dessous :

Tableau 11.10 Analyse factorielle exploratoire sur les « *Manipulation Checks* »

Concepts	Nombre d'items	KMO / Test de Bartlett	% de variance	Qualité de la représentation (min – max)	α de Cronbach
Quantité perçue	3	0,632 / 0,000	62,2%	0,47[263] – 0,70	0,69
Contexte perçu	3	0,553 / 0,000	63,5%	0,27[264] – 0,83	0,70

Même si les résultats de ces analyses sont décevants, ils restent malgré tout acceptables, au moins pour ce qui est de vérifier l'efficacité des manipulations expérimentales. Nous décidons donc de mener cette analyse sur la base de 5 questions sur les 6 proposées (après suppression de l'item MC_c3). La vérification des manipulations fera alors l'objet d'un test-t sur échantillons indépendants pour la Quantité et d'une ANOVA (test F) pour le Contexte (puisqu'il y a 3 niveaux). L'objectif est de déduire qu'il existe des différences de moyenne significatives sur la perception des facteurs manipulés[265].

Les résultats de ces analyses sont présentés dans le tableau suivant :

[262] Les trois premières questions utilisées pour le facteur quantité dans l'expérience 1 ont été reprises. La dernière, formulée « j'ai trouvé que le formulaire était long à remplir », a été supprimée faute de pouvoir s'appliquer au cas présent (dans cette expérience, les répondants ne remplissaient pas le formulaire). Pour égaliser le nombre d'items par facteur, nous avons également posé trois questions pour le facteur contexte.

[263] L'item correspondant (MC_q1) est malgré tout conservé puisque sa qualité de représentation est proche du niveau acceptable (0,47 contre 0,5) et qu'il permet de conserver la validité de contenu de cette « échelle ». La conservation de cet item réduit cependant quelque peu la fiabilité de l'échelle (alpha de 0,69 contre 0,72 sans), qui reste toutefois acceptable.

[264] L'item correspondant (MC_c3) est supprimé car sa qualité de représentation est trop éloignée du seuil acceptable (0,27 contre 0,5). De plus, la relecture de l'item montre qu'il porte en effet à confusion, ce qui n'avait cependant pas été détecté lors du pré-test des échelles. Pour information, l'alpha obtenu sur les 3 items (0,70) passe à 0,86 après suppression de cet item. Il n'est donné qu'à titre indicatif puisqu'il ne porte que sur 2 items.

[265] Nous menons les analyses sur chacun des items et sur la moyenne des items. Nous ne présentons dans le tableau 11.11 que le résultat pour la moyenne.

Tableau 11.11 Vérification des manipulations expérimentales

Facteur	Niveau	N	Moyennes	Ecart-type	Test t ou F	Sig.
Quantité perçue	faible	126	2,19	1,267	15,987	0,000
	élevée	126	4,98	1,494		
Contexte perçu	jeu	84	5,43	1,691	43,027	0,000
	newsletter	84	3,12	1,989		
	espace personnel	84	3,09[266]	1,932		

Ces résultats confirment l'effectivité des manipulations expérimentales puisque les différences de moyenne sont systématiquement significatives.

Voyons à présent, avant d'aborder la partie résultats (6), la qualité psychométrique des instruments de mesure utilisés (5.3) ainsi que le plan de traitement envisagé (5.4).

5.3 La qualité psychométrique des instruments de mesure

Le modèle à tester dans le cadre de cette deuxième expérience (cf. figure 11.2) comprend six variables médiatrices et/ou dépendantes et quatre catégories de variables de contrôle.
Parmi ces dernières, seule l'implication fait l'objet d'une mesure multi-item et se trouve donc concernée par la vérification des qualités psychométriques de son instrument de mesure[267].
En ce qui concerne le cœur du modèle, quatre variables sur les six répertoriées disposent d'une mesure multi-items. Les variables d'intentions de répondre et de mentir sont, pour leur part, évaluées chacune par un item unique, mesuré sur une échelle en 5 points, de « non certainement pas »[268] à « oui certainement » (avec en point neutre « peut-être, peut-être pas »). Ces items sont formulés respectivement : « j'aurais rempli puis validé ce formulaire » et « j'aurais menti à certaines des questions posées »[269].

[266] Le faible écart entre les moyennes pour les niveaux « newsletter » et « espace personnel » s'explique par la suppression de l'item MC_c3, les deux items conservés portant sur le jeu-concours. Nous vérifions donc que les gens affectés au jeu perçoivent bien qu'ils étaient sollicités pour participer à une loterie, tandis que ceux affectés aux deux autres cas perçoivent qu'il ne s'agissait pas d'un jeu-concours.
[267] La mesure des autres variables de contrôle a été exposée précédemment dans ce chapitre, au paragraphe 2.2
[268] Nous rappelons que ce format d'échelle est couramment employé pour mesurer des intentions comportementales, notamment des intentions d'achat ou de fidélité.
[269] Pour établir la « validité » de ces items, ils ont été préalablement soumis à une évaluation auprès d'experts et à un pré-test auprès de consommateurs. Ces analyses montrent qu'ils sont bien compris et qu'ils mesurent bien ce qu'ils sont censés mesurer.

Au total, 5 variables sont donc concernées par la vérification de la qualité psychométrique de leurs instruments de mesure dans cette expérience : l'implication (IMPL), l'attitude (ATT), la préoccupation pour le RVP (PREOC), la valeur perçue (VAL) et l'intention de poursuivre la relation (IPR).

Pour rappel, les trois premières ont déjà été mesurées dans l'expérience 1. Les qualités psychométriques de ces échelles ont, en outre, déjà été établies et présentées au chapitre 9. Nous nous contenterons donc ici de réaliser une nouvelle analyse factorielle pour chacune de ces échelles, afin de vérifier leur validité et leur fiabilité (par le calcul de l'alpha de Cronbach). Ces résultats sont présentés dans le tableau 11.12 ci-dessous.

Tableau 11.12 Analyse factorielle sur les variables Attitude, Préoccupation et Implication

Concepts	Nb d' items	KMO / Test de Bartlett	% de variance	Qualité de la représentation (min – max)	α de Cronbach
Attitude (ATT)	3	0,638 / 0,000	72,7%	0,54 – 0,84	0,80
Préoccupation RVP (PREOC)	3	0,602 / 0,000	69,4%	0,49[270] – 0,83	0,77
Implication (IMPL)	3	0,715 / 0,000	77,5%	0,71 – 0,81	0,85

En ce qui concerne les deux dernières variables évoquées ci-dessus (valeur et intention de poursuivre la relation), elles ont fait l'objet d'un développement d'échelles (présenté au chapitre 9) et notamment d'une réplication de l'analyse exploratoire et d'une analyse confirmatoire sur la base des résultats de cette expérience. La qualité psychométrique de ces deux instruments de mesure a ainsi été soigneusement vérifiée et a pu être établie.

Nous ne reviendrons pas sur ces analyses déjà exposées précédemment et nous nous contentons, pour rappel, d'indiquer dans le tableau 11.13, les principaux résultats que nous avions obtenus.

[270] Cet item (PREOC2) est conservé, compte tenu de sa proximité avec le niveau d'acceptation (0,49 contre 0,5)

Tableau 11.13 Qualités psychométriques des autres échelles

Concepts	Nombre d'items	Qualité de la représentation (min – max)	Alpha de Cronbach	Rhô de Jöreskog	VC[271] (ρVC)	VD
Valeur perçue (VAL)	6[272]	0,69 – 0,81	0,70 et 0,86	0,76 et 0,86	0,62	Oui
Intention de poursuivre la relation (IPR)	3[273]	0,52 – 0,63	0,73	0,715	0,503	Oui

Ces tableaux confirment la qualité psychométrique des instruments de mesure mobilisés pour l'expérience 2 et nous autorisent à les utiliser pour vérifier les hypothèses posées[274].

Nous faisons d'ailleurs remarquer que les résultats obtenus ici, notamment en ce qui concernent les échelles présentées au tableau 11.12, sont similaires à ceux obtenus pour l'expérience 1 alors qu'ils avaient été établis sur la base d'un questionnaire papier-crayon. Ceci tendrait donc à confirmer les conclusions de King et Miles (1995) quant à l'équivalence de mesures selon le mode d'administration (par exemple : courrier vs. Internet).

Nous terminons cette section en exposant le plan de traitement prévu.

5.4 Le plan de traitement

L'influence des facteurs manipulés sur la valeur perçue (correspondant aux hypothèses H1.3d et H1.4d) sera vérifiée par l'intermédiaire d'une analyse de variance, avec introduction de variables de contrôle, c'est-à-dire par une ANCOVA.

Nous venons de voir que les conditions d'application d'une analyse de variance étaient réunies (point 5.1 et 5.2). Nous pouvons donc mener à bien une telle analyse.

En ce qui concerne la partie droite du modèle (et donc les hypothèses H2d, H3e, H4 et H5), il s'agit d'établir l'existence de liens de cause à effet entre des variables continues. Nous procéderons donc par le biais de régressions multi-variées, en introduisant successivement les variables au sein de blocs, le dernier bloc étant constitué des variables de contrôle.

[271] Correspond à la première « dimension » de la validité de trait c'est-à-dire à la validité convergente (VC), la seconde étant la validité discriminante (VD). La validité prédictive n'est pas reproduite dans ce tableau mais a été vérifiée au chapitre 9 en ce qui concerne la valeur perçue.

[272] Pour rappel, l'échelle de valeur perçue comporte deux dimensions (intérêt et coût) comprenant respectivement 4 et 2 items. Nous fournissons alors les indices de fiabilité pour ces deux dimensions respectivement.

[273] L'échelle initiale comportait un autre item, que nous avons dû supprimer pour en améliorer la qualité globale.

[274] Nous calculerons, pour chaque variable dépendante, un score par addition des items correspondants. Ainsi, la valeur perçue est calculée en soustrayant le score de coût (calculé sur les 2 items) au score d'intérêt (4 items).

Nous avons déjà indiqué au chapitre précédent (i.e. chapitre 10), les conditions d'application d'un modèle de régression multiple. Nous n'y reviendrons donc pas, nous contentant de donner, le cas échéant, les résultats obtenus concernant les analyses menées pour les vérifier.

Pour rappel, l'absence de multi-colinéarité sera vérifiée au moment de l'analyse des résultats par l'examen des indicateurs de Tolérance et de VIF (pour *Variance Inflation Facteur*) qui devront être respectivement supérieur à 0,3 et inférieur à 3,3. La normalité des variables sera, pour sa part, testée par l'examen du coefficient d'asymétrie (*skewness*) et du coefficient d'aplatissement *(kurtosis)* qui doivent être respectivement inférieurs à 1 et à 1,5.
Nous avons déjà vérifié la normalité de l'ensemble des variables médiatrices et dépendantes du modèle pour chacun des facteurs manipulés (cf. paragraphe 5.1.4). Nous vérifions à présent la normalité de la préoccupation pour le RVP et des variables de contrôle métriques (niveau d'implication et comportement habituel de réponse), tous traitements confondus.
Les résultats fournis en Annexe 11.7 confirment le respect de cette hypothèse, dans la mesure où les coefficients d'asymétrie et d'aplatissement sont conformes aux règles énoncées.

Bien que n'ayant pas formulé d'hypothèses à ce sujet, nous signalons que plusieurs variables au sein du modèle à tester, sont positionnées en tant que médiatrices. L'attitude jouerait ainsi un rôle médiateur entre la valeur perçue et les intentions comportementales de réponse (IR et IM). De même, l'intention de répondre (IR) jouerait un rôle médiateur entre l'attitude et l'intention de poursuivre la relation (IPR).

De manière classique, nous testerons ce rôle médiateur, en regardant si les conditions données par Baron et Kenny (1986) et reprises par Chumpitaz et Vanhamme (2003) sont vérifiées :

1. les variations des variables indépendantes entraînent des variations significatives de la variable dépendante
2. les variations des variables indépendantes entraînent des variations significatives des variables supposées médiatrices
3. la relation entre la variable indépendante et la variable dépendante n'est pas significative lorsque le chemin précédent est contrôlé (effet de la variable médiatrice)

Nous symbolisons ces conditions, dans le cadre de l'expérience 2, dans la figure qui suit :

Figure 11.3 Les relations de médiation dans le cadre de l'expérience 2

Dans la mesure où toutes les variables sont continues, nous testerons à chaque fois les relations (1), (2) et (3) par le biais de régressions simples et/ou multiples.

Le tableau suivant récapitule le plan de traitement envisagé (sans tenir compte toutefois du test des effets médiateurs puisqu'ils n'ont pas été formulés sous forme d'hypothèses) :

Tableau 11.14 Plan de traitement des hypothèses correspondant à l'expérience 2

Hyp.	Variable indépendante	Variable dépendante	Tests réalisés	Résultats escomptés
H1.3 d	Quantité	Valeur perçue (VAL)	ANCOVA (avec introduction des variables de contrôle comme co-variables)	Les coefficients liant les facteurs manipulés à la valeur perçue doivent être significatifs
H1.4 d	Contexte	Valeur perçue (VAL)		
H2 d	Valeur perçue	Attitude (ATT)	Régression multiple (bloc 1 : valeur perçue, bloc 2 : préoccupation pour le RVP et bloc 3 : variables de contrôle)	Les coefficients correspondant aux variables indépendantes doivent être significatifs
H3 e	Préoccupation pour le RVP	Attitude (ATT)		
H4 (f,g)	Attitude	Intentions de répondre (IR) et de mentir (IM)	2 régressions multiples	
H5	Intention de répondre	Intention de poursuivre la relation (IPR)	Régression multiple	

Nous présentons, dans la section suivante, l'ensemble des résultats obtenus sur les tests réalisés et faisons ainsi le point sur la validation (ou non) des hypothèses proposées.

6 Présentation des résultats

Nous présentons, dans cette section, les résultats des tests des hypothèses.
Nous débutons (6.1) par ceux qui concernent l'influence de l'exposition à la sollicitation (correspondant aux deux facteurs manipulés, i.e. Quantité et Contexte) sur l'évaluation de celle-ci (correspondant à la valeur perçue). Pour rappel, deux hypothèses ont été développées en ce sens et justifiées au paragraphe 3.2. Elles seront testées conjointement par le biais d'une ANCOVA[275], dont les résultats sont exposés ci-après.
Nous poursuivons cette section (6.2) par les résultats concernant l'impact de l'évaluation (valeur perçue) et de la préoccupation pour le RVP sur l'attitude face à la communication des données. Deux hypothèses sont concernées ici (respectivement H2 d et H3 e), qui ont été développées et justifiées dans le chapitre 7 et qui seront testées conjointement par le biais d'une régression multiple avec introduction de blocs. Nous testons alors le cœur du modèle conceptuel à savoir l'impact respectif de la situation (évaluation) et des caractéristiques individuelles (préoccupation pour le RVP) sur le degré d'agrément de l'individu eu égard à la divulgation de ses données (attitude). Pour rappel, dans l'expérience 1, nous avions montré que l'effet des caractéristiques situationnelles était proportionnellement plus important que celui des variables individuelles. Notre objectif est ici de vérifier si ce résultat se confirme dans cette nouvelle expérience.

De plus, au-delà de l'attitude face à la communication des données (phase où nous nous étions arrêtés dans l'expérience 1), nous poursuivons ici l'analyse, allant jusqu'à mesurer les intentions comportementales de réponse. Nous analyserons ainsi l'impact de l'attitude sur l'intention de répondre et de mentir (6.3). Deux hypothèses sont concernées ici, respectivement H4 f (pour l'intention de répondre) et H4 g (pour l'intention de répondre). Nous étudierons également (6.4) l'impact de l'intention de répondre sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données (correspondant à l'hypothèse H5), vérifiant ainsi les principes liés à la théorie de l'engagement.
Enfin, nous terminons cette analyse (6.5) par l'étude des deux effets médiateurs, celui de l'attitude et celui de l'intention de répondre, présentés précédemment (notamment dans la figure 11.3). Nous rappelons toutefois que ces deux effets n'ont fait l'objet d'aucune hypothèse préalable.

[275] Nous fournissons, en Annexe 11.8, les fichiers de syntaxe ainsi que les principaux résultats de cette analyse.

6.1 L'impact de l'exposition à la sollicitation sur l'évaluation de celle-ci

Comme indiqué ci-dessus, pour tester les deux hypothèses concernées ici (H1.3 d et H1.4 d), nous procédons par l'intermédiaire d'une ANCOVA. Nous introduisons, conjointement, les 2 facteurs manipulés (Quantité et Contexte) en tant que facteurs fixés, la valeur perçue comme variable dépendante et les variables de contrôle (variables sociodémographiques, expérience et utilisation d'Internet, implication[276]) en tant que co-variables. Nous spécifions ce modèle en tant que modèle factoriel complet afin de vérifier l'existence éventuelle d'effets d'interaction. Comme dans l'expérience 1, nous regardons la significativité du test F au niveau 10%[277] (test unilatéral), tout en conservant le seuil habituel de 5% pour les co-variables.

Les résultats de l'analyse de covariance sont présentés dans le tableau suivant. Par souci de lisibilité, nous mettons en grisé les résultats correspondant au test des deux hypothèses et en gras et italique les coefficients correspondant à des effets significatifs au seuil de 10% (pour le facteur Quantité) ou de 5% (pour les variables de contrôle et le facteur Contexte).
Nous commenterons ci-après, l'influence de la situation (des facteurs manipulés et de leurs éventuels effets d'interaction) sur l'évaluation de la sollicitation (à travers la valeur perçue).

Tableau 11.15 Influence des facteurs manipulés sur la valeur perçue (VAL)

Source	Somme des carrés de type III	ddl	Moyenne des carrés	F	Sig.	Hypothèse
Modèle corrigé	1530,036 [(a)]	13	117,695	2,016	0,020	
Constante	378,368	1	378,368	6,483	0,012	
Implication (IMPL)	512,619	1	512,619	8,783	***0,003***	
Sexe	2,706	1	2,706	0,046	0,830	
Age	135,136	1	135,136	2,315	0,129	
CSP	13,384	1	13,384	0,229	0,632	
Niveau d'Etudes	206,058	1	206,058	3,530	0,061	
EXP d'Internet	8,884	1	8,884	0,152	0,697	
UTIL (messagerie)	17,503	1	17,503	0,300	0,584	
UTIL (surf web)	35,162	1	35,162	0,602	0,438	
Quantité	185,429	1	185,429	3,177	***0,076***	***H1.3 d (V)***
Contexte	86,120	2	43,060	0,738	0,479	***H1.4 d (NV)***
Quantité x Contexte	53,544	2	26,772	0,459	0,633	
Erreur	13891,234	238	58,367			
Total	32392,000	252				
Total corrigé	15421,270	251				

(a) R^2 = 0,099 (R^2 ajusté = 0,050)

[276] Nous excluons toutefois le comportement habituel de réponse, dont nous pensons qu'il influence surtout les intentions comportementales.
[277] L'impact du facteur Contexte reposant sur une hypothèse de moyennes différentes, nous le testerons à 5%.

Nous discutons tout d'abord les résultats relatifs à l'impact de chaque facteur manipulé (6.1.1, 6.1.2) avant d'évoquer l'influence des co-variables (variables de contrôle) (6.1.3).

6.1.1 L'impact de la Quantité de données demandées

Au regard des résultats et conformément à ce que nous avions supposé, le facteur « Quantité » a un effet sur la valeur perçue à fournir les données demandées, au seuil de 10% (p = 0,076). Cet effet va dans le sens supposé, conduisant alors à **valider l'hypothèse H1.3 d**.
Ainsi, comme le montre la figure 11.4, plus la quantité de données demandées est élevée (formulaire long), plus la valeur perçue (ou le gain net) à répondre est faible. Au contraire, quand le nombre de champs à remplir sur le formulaire est réduit (formulaire court), la valeur perçue à répondre augmente. Ce résultat vient confirmer les principes issus de la théorie de l'échange social et de la théorie de la motivation. Ainsi, quand le nombre d'informations à fournir augmente, le coût lié au remplissage s'accroît parallèlement, réduisant alors le gain net à répondre (à bénéfice constant) L'évaluation de la sollicitation serait donc d'autant plus favorable (en termes de valeur perçue) que le nombre d'informations à fournir serait limité car cela réduirait l'effort à consentir par l'individu et le mettrait donc dans de meilleures dispositions pour répondre. Ce résultat va dans le même sens que celui obtenu dans l'expérience 1, où nous avions montré que l'évaluation de la sollicitation (en termes de sensibilité et de pertinence) était d'autant plus favorable que le nombre de données à fournir était limité, car cela rassurait l'individu sur les intentions de l'entreprise le sollicitant.

Figure 11.4 Impact de la Quantité sur la valeur perçue

6.1.2 L'impact du Contexte de sollicitation

Contrairement à ce que nous avions supposé, le contexte dans lequel les données sont sollicitées, correspondant ici à trois cas possibles (i.e. participation à un jeu-concours, inscription à une newsletter ou création d'un espace personnalisé sur le site), n'influencerait pas la valeur perçue à répondre (p = 0,479), conduisant ainsi à **ne pas valider l'hypothèse H1.4 d**. Alors que la quantité de données demandées aurait un impact sur les coûts à répondre (et donc sur le gain net perçu, à bénéfice égal), nous pensions en effet que le contexte agirait, pour sa part, sur la partie « bénéfices perçus » puisque dans chaque cas proposé, les avantages reçus en échange des données sont différents. Pourtant, il ne semble pas que ce soit le cas ici. Plus précisément, il apparaît que s'il existe un écart entre le jeu-concours et la newsletter en termes de valeur perçue[278], cet écart serait très réduit (voire quasi inexistant) entre la newsletter et l'espace personnel[279] (cf. figure 11.5). Ce résultat n'est peut être pas si étonnant que cela. Nous opposions en effet un bénéfice utilitaire (précisément monétaire), espéré (au sens d'incertain puisque seuls quelques participants remporteront les lots mis en jeu) et différé (le gain éventuel ne sera effectivement perçu qu'à moyen terme) dans le cas du jeu-concours à un bénéfice hédonique et/ou de reconnaissance, certain et immédiat dans les deux autres contextes. Bien que les avantages soient *a priori* sensiblement différents dans le cas de la newsletter (informations non personnalisées) et de l'espace personnel (offres personnalisées), la catégorie à laquelle chacun d'eux peut se raccrocher est similaire. Les répondants ne semblent donc pas avoir fait la différence entre ces deux types de bénéfices, assimilant ainsi ces deux situations à des contextes « proches ». Cet effet est, en outre, renforcé par le caractère « artificiel »[280] de la manipulation, le contexte dans lequel ils sont censés être sollicités étant simplement décrit dans le scénario et repris en haut du formulaire.

Par mesure de précaution, nous vérifions si le contexte n'a pas un effet direct sur l'attitude face à la communication des données sollicitées. Il se pourrait en effet que la valeur perçue ne joue pas le rôle de médiateur attendu. Cette hypothèse est cependant contredite par les résultats de l'analyse de variance (p=0,650), la différence d'attitude se révélant peu importante d'un contexte à l'autre. Il apparaît toutefois que ce sont, cette fois-ci, les contextes de jeu-concours et d'espace personnel qui sont les plus proches, se différenciant nettement (mais de manière non significative) de celui de la newsletter (cf. Figure 11.6).

---

[278] Cet écart n'est cependant pas significatif, en atteste le test de Bonferroni (p = 0,615).

[279] Ce que confirme le test de Bonferroni (p = 0,962).

[280] Malgré toutes les précautions prises pour renforcer le caractère réaliste des manipulations, il n'en reste pas moins qu'il ne s'agit pas d'un contexte réel, pour lequel les répondants auraient peut-être perçu avec plus de « force » la différence entre les cas « newsletter » et « espace personnel ».

En termes d'attitude, les répondants sont ainsi plus favorables au dévoilement d'informations dans le cadre d'un jeu-concours ou de la création d'un espace personnel sur le site que pour simplement recevoir une newsletter, cet écart n'étant cependant pas significatif. Les raisons d'un tel résultat (proximité du contexte d'espace personnel, avec la newsletter face à la valeur perçue, et avec le jeu-concours face à l'attitude) devront être analysées à l'avenir.

Le facteur « Contexte » testé dans le cadre de cette expérience n'a également aucun impact, ni sur la valeur perçue ni sur l'attitude, en interaction avec le facteur « Quantité » (cf. figures 11.5 et 11.6). Ainsi, quel que soit le niveau de données demandées (quantité faible ou élevée), l'impact du contexte n'est pas significativement différent dans les trois cas testés.

Figure 11.5 Impact du Contexte et du Contexte x Quantité sur la valeur perçue

Figure 11.6 Impact du Contexte et du Contexte x Quantité sur l'attitude

Au final, nous estimons que, même si l'effet du contexte n'est pas significatif dans le cadre de cette expérience, nous ne pouvons en conclure que le contexte dans lequel les individus sont sollicités pour fournir des données à caractère personnel n'influence aucunement le processus de réponse. Les contextes testés ici, bien que montrant certains écarts, ne sont peut être pas suffisamment « opposés » pour engendrer une évaluation différente de la sollicitation. Leur manipulation n'est peut être pas non plus suffisamment effective (au sens de réaliste) pour engendrer une perception distincte. De plus, si le contexte n'a pas d'impact significatif sur la valeur perçue (ni sur l'attitude face à la communication des données), cela ne veut pas dire qu'il n'en a pas sur d'autres éléments du processus de réponse. Enfin, il est également possible que le contexte n'agisse pas seul pour influencer la réponse du consommateur. Il pourrait en effet, en interaction avec d'autres facteurs que ceux manipulés ici, agir de manière significative sur l'évaluation de la sollicitation ou sur la suite du processus de réponse. L'ensemble de ces pistes devra être envisagé à l'avenir afin d'avoir une idée plus précise quant à l'impact du facteur « Contexte » sur la réponse d'un individu face à une sollicitation de données personnelles. Certaines de ces propositions seront ainsi testées dans l'expérience 3 (i.e. autres contextes que ceux évalués ici, manipulation plus réaliste, analyse de l'impact sur d'autres variables du processus de réponse), afin de trouver les plus opportunes d'entre elles.

En résumé, concernant l'effet des facteurs manipulés sur l'évaluation de la sollicitation, **une hypothèse sur les deux formulées est validée : il s'agit de l'hypothèse H1.3 d** qui souligne l'impact (négatif) de la « Quantité de données demandées » sur la valeur perçue à répondre. Cette dernière serait ainsi plus élevée en cas de quantité faible car cela réduirait le montant d'effort à fournir et donc le coût perçu à répondre (à bénéfice égal).

Ce résultat vient conforter ceux établis au cours de l'expérience 1 concernant l'impact de la quantité sur l'évaluation de la sollicitation (en termes de pertinence et de sensibilité perçues). Le nombre de données à fournir aurait donc un impact majeur sur le processus de réponse d'un individu confronté à une sollicitation de ses données personnelles, et notamment sur l'évaluation de la sollicitation.

De plus, si **l'hypothèse concernant l'impact du « Contexte de sollicitation » sur la valeur perçue à répondre (i.e. H1.4 d) n'est pas validée ici**, nous ne voulons pas conclure, pour autant, en l'absence d'effet de ce facteur sur le processus de réponse. D'autres moyens d'évaluer l'impact du contexte sur le comportement de l'individu seront ainsi testés dans l'expérience 3, afin d'avoir une idée plus précise quant à la portée explicative de ce facteur.

Cette analyse de variance incluait des variables exogènes individuelles dont nous souhaitions contrôler l'impact. Nous commentons ces résultats dans le paragraphe suivant (6.1.3).

6.1.3 L'impact des variables exogènes

Sur l'ensemble des variables exogènes introduites en tant que co-variables, seul le niveau d'implication a un impact significatif (p = 0,003) sur la valeur perçue. Si on entre dans les détails, on s'aperçoit que cet effet est positif : un niveau élevé d'implication augmenterait la valeur perçue à répondre (corrélation de Pearson = 0,205 ; p = 0,01). Ce résultat est tout à fait logique. Il semble en effet normal de constater qu'un individu très intéressé par la téléphonie mobile trouve plus profitable de répondre à une sollicitation de données personnelles dans ce domaine qu'un individu faiblement impliqué.
Ce constat confirme à la fois les conclusions des recherches sur l'implication (dont celles de Amine 1990) et les résultats établis dans l'expérience 1, concernant l'impact de l'implication sur les trois autres variables de perception (confidentialité, sensibilité et pertinence perçue).
Dans l'ensemble, l'implication a donc un impact significatif sur l'évaluation par l'individu de la sollicitation (à travers les quatre variables de perception testées), les autres variables exogènes n'ayant, quant à elles, quasiment aucun impact. Ce critère (le niveau d'implication) doit donc être pris en compte pour évaluer l'impact de la situation sur la communication de données personnelles.

Après avoir étudié la partie gauche du modèle à tester (hypothèses H1.3 d et H1.4 d), passons à présent à la partie droite qui concerne la suite du processus de réponse. Cette partie sera testée par le biais de régressions multiples avec entrée de blocs successifs.

6.2 L'impact de l'évaluation de la sollicitation et de la préoccupation pour le RVP sur l'attitude face à la communication des données

Comme indiqué ci-dessus, pour tester les deux hypothèses concernées ici (H2 d et H3 e), nous procédons par le biais d'une régression multiple avec introduction de blocs successifs. Nous introduisons, dans le bloc 1, la valeur perçue, dans le bloc 2, la préoccupation pour le RVP et, dans le bloc 3, les variables de contrôle (implication, variables sociodémographiques et expérientielles). La variable dépendante est l'attitude face à la communication des données.

L'ensemble des conditions préalables à la réalisation d'une telle régression ont été soit contrôlées précédemment (cf. paragraphe 5.4), soit sont présentées en Annexe 11.9 (point A). La dernière condition à vérifier (i.e. la multi-colinéarité entre les variables indépendantes du modèle) sera, pour sa part, contrôlée ci-après, à travers les indicateurs de Tolérance et de VIF. Les résultats (cf. tableau 11.16) témoignent d'une absence de multi-colinéarité extrême entre les variables indépendantes du modèle.

Les résultats de la régression multiple réalisée sont résumés dans le tableau ci-après. Nous y faisons figurer les indicateurs d'ajustement du modèle ainsi que les coefficients obtenus. Nous commentons ensuite les résultats concernant la validation (ou non) des 2 hypothèses testées.

Tableau 11.16 Régression visant à expliquer l'attitude face à la communication des données

	Bloc 1		Bloc 2		Bloc 3		Colinéarité	
	B	Sig.	B	Sig.	B	Sig.	Tol.	VIF
Constante	10,850	0,000	12,980	0,000	15,318	0,000		
Valeur perçue	0,389	**0,000**	0,357	**0,000**	0,336	**0,000**	0,776	1,289
Préoccupation RVP			-0,128	**0,022**	- 0,138	**0,014**	0,829	1,206
Implication					0,051	0,383	0,924	1,083
Sexe					0,308	0,538	0,980	1,020
Age					- 0,182	0,362	0,854	1,171
CSP					- 0,520	0,176	0,809	1,237
Niveau d'Etudes					- 0,092	0,770	0,774	1,292
EXP d'Internet					- 0,283	0,453	0,947	1,056
UTIL (messagerie)					- 0,355	0,463	0,692	1,445
UTIL (surf web)					- 0,286	0,480	0,703	1,423
R^2	0,372		0,386		0,406		-	
R^2 ajusté	0,370		0,381		0,381		-	

Ces résultats montrent que la valeur perçue a une influence réelle sur l'attitude face à la communication des données, y compris après introduction du deuxième (préoccupation pour le RVP) et du troisième (variables de contrôle) blocs (p = 0,000 respectivement), **validant ainsi l'hypothèse H2 d**. Comme pressenti donc, l'évaluation de la sollicitation (sous la forme de perceptions) influence significativement l'attitude de l'individu face à la communication de ses données personnelles. En effet, nous avons montré, dans l'expérience 1, que l'attitude était influencée de manière significative par la confidentialité, la sensibilité et la pertinence perçues. De plus, nous déduisons ici que la valeur perçue exerce aussi un impact majeur.

L'influence de la préoccupation pour le respect de la vie privée (variable individuelle) s'avère également significative (p = 0,014 dans le modèle 3), sans que son introduction n'améliore toutefois de manière substantielle le pouvoir explicatif du modèle (38,6% avec vs. 37,2% sans, soit moins d'1,5% de gain). **Nous validons ainsi l'hypothèse H3e** qui prévoyait l'influence de cette variable sur l'attitude face à la communication des données et confortons dès lors les résultats obtenus au cours de l'expérience 1[281]. De plus, l'examen des coefficients standardisés (bêtas) prouve que l'attitude a environ quatre fois plus d'impact sur la valeur perçue que la préoccupation pour le RVP (bêtas = 0,527 et -0,135 respectivement). Nous confirmons donc le constat établi dans l'expérience 1, selon lequel l'effet de la situation sur l'évaluation de la sollicitation serait prépondérant comparativement aux convictions personnelles (mesurées par la préoccupation pour le respect de la vie privée). Enfin, l'introduction des différentes variables de contrôle n'améliore pas le pouvoir explicatif du modèle (toujours 38%), aucune de ces variables ne se révélant significative au seuil de 5%.

Pour conclure sur cette deuxième partie des résultats, nous pouvons dire que **nous validons les 2 hypothèses concernées ici (H2 d et H3 e)**, soulignant ainsi l'impact de la valeur perçue et de la préoccupation pour le RVP sur l'attitude face à la communication des données.
Il apparaît aussi, confirmant alors les résultats obtenus pour l'expérience 1, que l'influence de la situation (à travers les variables de perception) serait significativement plus importante que l'influence des variables individuelles (à travers la préoccupation pour le RVP), ces dernières expliquant moins de 2% du modèle contre 37% pour les caractéristiques situationnelles.
Ce résultat conforte donc l'idée selon laquelle, face à une sollicitation de ses données personnelles, l'individu se déciderait davantage sur la base de la situation dans laquelle il se trouve et des conditions qui lui sont proposées que d'après ses convictions personnelles.

Nous poursuivons l'étude de la partie droite du modèle à tester, en examinant l'impact de l'attitude face à la communication des données sur l'intention de répondre et de mentir.

6.3 L'impact de l'attitude sur l'intention de répondre et de mentir

Comme dans le cas précédent, nous indiquons que les conditions nécessaires à la réalisation de ces analyses ont été contrôlées auparavant (cf. paragraphe 5.4) ou sont présentées en

[281] Nous avions en effet déjà trouvé un impact significatif de la préoccupation sur l'attitude (cf. chapitre 10).

Annexe 11.9 (points B et C). De plus, les résultats obtenus (cf. tableaux 11.17 et 11.18) montrent une absence de multi-colinéarité extrême entre les variables indépendantes testées.

Deux régressions successives ont en réalité été menées, la première ayant pour variable dépendante l'intention de répondre (IR), la seconde, l'intention de mentir (IM), les variables indépendantes concernant l'attitude (bloc 1) et un ensemble de variables de contrôle (bloc 2). Nous faisons figurer ci-après les tableaux de résultats donnant les indicateurs d'ajustement des modèles ainsi que les coefficients obtenus. Nous commentons ensuite les résultats concernant la validation (ou non) des 2 hypothèses testées. Nous précisons toutefois que, contrairement aux analyses précédentes, nous introduisons ici le comportement habituel de réponse en tant que variable de contrôle[282], celui-ci étant mesuré différemment dans chacun des modèles puisque la variable dépendante l'est également. Nous contrôlerons ainsi l'habitude de répondre (CHR) dans le modèle 1 et l'habitude de mentir (CHM) dans le 2.

6.3.1 L'impact de l'attitude sur l'intention de répondre

Tableau 11.17 Régression visant à expliquer l'intention de fournir les données (IR)

	Bloc 1		Bloc 2		Colinéarité	
	B	Sig.	B	Sig.	Tol.	VIF
Constante	1,239	0,000	1,433	0,004		
Attitude	0,168	**0,000**	0,149	**0,000**	0,792	1,263
Implication			-0,008	0,551	0,907	1,103
Sexe			-0,129	0,274	0,973	1,028
Age			-0,040	0,397	0,851	1,175
CSP			0,026	0,770	0,803	1,245
Niveau d'Etudes			- 0,136	0,068	0,768	1,301
EXP d'Internet			0,149	0,094	0,944	1,060
UTIL (messagerie)			0,098	0,389	0,693	1,444
UTIL (surf web)			- 0,131	0,168	0,706	1,416
Habitude de répondre (CHR)			0,112	**0,002**	0,802	1,246
R^2	0,445		0,484		-	
R^2 ajusté	0,442		0,463		-	

Les résultats de ce premier modèle confirment l'impact significatif de l'attitude sur l'intention de fournir les données demandées (i.e. intention de répondre) (p= 0,000), en contrôlant l'effet du comportement habituel de réponse, à savoir l'habitude de répondre (CHR) (p= 0,002).

[282] La justification de ce contrôle à ce niveau (et de son absence dans les analyses précédentes) a déjà été fournie précédemment dans le chapitre 8 ou au paragraphe 2.2 du présent chapitre.

De plus, l'effet a bien lieu dans le sens attendu, une attitude favorable ayant une plus forte tendance à engendrer une réponse positive, **amenant ainsi à valider l'hypothèse H4f**. L'effet de l'habitude de répondre est également positif, les individus ayant tendance à répondre favorablement ayant une plus forte probabilité de répondre, à chaque nouvelle sollicitation.
L'attitude expliquerait à elle seule plus de 44% de l'intention de répondre, l'ensemble des variables de contrôle (parmi lesquelles seule l'habitude de répondre est significative) expliquant, pour leur part, à peine 2% (46,3% avec contre 44,2 avec). Ce constat se confirme si l'on observe les coefficients standardisés. On s'aperçoit en effet que l'impact de l'attitude (présent), est très supérieur à celui du comportement habituel de réponse (passé), dans un rapport de 1 à 4 (bêtas=0,594 et 0,164 respectivement). On peut donc conclure que concernant la probabilité de fournir des données personnelles, la décision de l'individu serait davantage liée à la situation concernée (présent) qu'à la manière dont il a l'habitude de répondre face à ce type de sollicitation (passé), cette dernière ne devant toutefois pas être négligée.
Voyons à présent l'impact de l'attitude sur l'intention de mentir.

6.3.2 L'impact de l'attitude sur l'intention de mentir

Tableau 11.18 Régression visant à expliquer l'intention de fournir des données erronées (IM)

	Bloc 1		Bloc 2		Colinéarité	
	B	Sig.	B	Sig.	Tol.	VIF
Constante	2,979	0,000	4,927	0,000		
Attitude	-0,083	**0,000**	-0,046	**0,000**	0,814	1,229
Implication			0,001	0,919	0,933	1,071
Sexe			-0,249	**0,007**	0,950	1,052
Age			0,062	0,088	0,827	1,209
CSP			0,05	0,941	0,792	1,263
Niveau d'Etudes			-0,008	0,885	0,774	1,292
EXP d'Internet			0,054	0,431	0,944	1,059
UTIL (messagerie)			-0,032	0,709	0,691	1,447
UTIL (surf web)			-0,169	**0,021**	0,704	1,421
Habitude de mentir (CHM)			-0,365	**0,000**	0,826	1,211
R^2	0,168		0,532		-	
R^2 ajusté	0,165		0,512		-	

Ces résultats confirment également l'impact significatif de l'attitude sur l'intention de fournir des données erronées (i.e. de mentir) (p= 0,000), en contrôlant à la fois le sexe, la fréquence d'utilisation du web et le comportement habituel de réponse (habitude de mentir, CHM). Plus précisément, l'attitude a un effet négatif sur l'intention de mentir, ce qui va dans le sens de ce

qui était attendu et permet ainsi de **valider l'hypothèse H4g**[283]. Un individu ayant une attitude favorable face à la communication de ses données aurait donc une plus forte tendance à dire la vérité qu'un individu défavorable. Nous confirmons ainsi l'hypothèse selon laquelle le mensonge serait vu comme un moyen de participer à l'échange sans se dévoiler complètement ou encore comme un moyen pour le consommateur d'obtenir les bénéfices associés au dévoilement, sans s'exposer aux risques (Lwin et Williams 2003). L'effet (négatif) de l'attitude sur l'intention de mentir serait toutefois soumis au contrôle de trois catégories de variables individuelles, d'ordre sociodémographique (sexe), expérientiel (fréquence de surf) et comportemental (habitude de mentir), dont nous allons à présent analyser l'influence.

En ce qui concerne le sexe, il apparaît que les hommes ont une plus forte tendance à fournir des données erronées que les femmes (B = - 0,249, et ce, de manière significative (p = 0,007). Ce résultat semble donc aller dans le sens des théories sur le dévoilement de soi qui prédisent que, dans la mesure où les femmes parlent plus facilement d'elles-mêmes que les hommes (Archer 1979, Cozby 1973), elles accepteraient plus volontiers de donner des informations personnelles à une entreprise. Dans la même logique, elles auraient alors une moins forte tendance à fournir des données erronées. Ce résultat confirmerait également le rôle des normes sociales (Taylor et Altman 1975) qui établit que le dévoilement de la part d'un homme est considéré comme moins approprié que s'il est réalisé par une femme. Cette dernière n'aurait donc pas besoin de mentir. Enfin, Collins et Miller (1994) constataient déjà que les femmes tendent davantage à considérer le dévoilement de soi comme une preuve d'ouverture et de récompense, ce qui les pousserait alors à moins mentir que les hommes.

L'effet de l'utilisation d'Internet (plus précisément de la fréquence de surf) est également significatif (p= 0,021). Il apparaît ainsi que plus l'individu passe du temps à surfer sur le web, plus il a tendance à fournir des réponses erronées (B = - 0,169)[284]. L'effet de la fréquence d'usage du Net sur l'intention de mentir serait donc négatif, les surfeurs intensifs voyant dans le mensonge un moyen de participer à l'échange, tout en limitant les représailles possibles.

---

[283] Ce résultat doit cependant être pris avec précaution, étant donné le caractère pour le moins incertain de la mesure de l'intention de mentir. Comme nous l'avons déjà souligné précédemment (notamment, au chapitre 7), comment en effet être sûr qu'un répondant qui dit avoir l'intention de mentir ne ment pas (et inversement) ? Pour mieux appréhender le phénomène, des variables supplémentaires devraient être intégrées au modèle afin de mieux contrôler la « véracité » des réponses. Nous pensons notamment à la mesure de l'obligation morale. Nous reviendrons sur les limites de ce résultat et sur les voies de recherche qui en découlent dans la Discussion finale.

[284] Pour rappel, nous mesurons la fréquence de surf sur une variable ordinale à 4 modalités (de 1 : plusieurs heures par jour à 4 : moins souvent).

Dans notre étude exploratoire descriptive (cf. chapitre 6), nous avions trouvé que les internautes consacrant une part importante de leur temps à surfer acceptent plus facilement de délivrer des informations sur un site. Ce résultat allait dans le sens des conclusions établies par Alexander (1998) selon lesquelles une fréquence élevée de surf serait associée à de plus bas niveaux de préoccupation[285]. Nous montrons ici que, s'ils remplissent plus volontiers le formulaire, ils ont aussi une plus forte tendance à fournir des informations erronées.
Ces deux résultats, loin d'être antinomiques, ont plus vocation à se compléter. Il ne semble d'ailleurs pas anormal de constater qu'un internaute plus expert que la moyenne (car surfant quotidiennement) aurait une plus forte tendance à mentir. On peut en effet supposer qu'un internaute habitué à surfer aura déjà fait l'expérience d'une utilisation abusive de ses données, ce qui le poussera davantage à rechercher des solutions de protection et/ou de repli. Ce constat va d'ailleurs dans le sens des conclusions établies par Nyshadham (2000) qui montre qu'une expertise en ligne élevée est associée à un plus haut niveau de préoccupation.

Enfin, concernant le comportement habituel de réponse (plus précisément, l'habitude de mentir), il apparaît influencer également l'intention de mentir (p=0,000). L'item correspondant étant formulé « négativement »[286] (i.e. « si je réponds, je dis toujours la vérité »), nous trouvons, logiquement, que l'habitude de mentir a un impact positif sur l'intention de mentir (B = - 0,365) : les gens ayant l'habitude de mentir ont une plus forte probabilité de fournir des données erronées que ceux qui disent habituellement la vérité.
Si l'on s'intéresse à présent à l'impact respectif des quatre variables influençant l'intention de mentir, on s'aperçoit que l'attitude face à la communication de données personnelles (l'antécédent de l'intention de mentir dans notre modèle à tester) explique moins de 17% du modèle, quand les variables de contrôle en expliquent près de 35%. L'attitude aurait donc un impact moins important que les variables individuelles. Ce constat se confirme si l'on observe les coefficients standardisés. La variable qui semble en effet avoir le plus fort impact serait l'habitude de mentir (bêta = - 0,591), suivie de l'attitude, qui aurait un poids trois fois moins élevé (bêta = - 0,229). Enfin, le sexe et la fréquence de surf viennent bien après, avec des bêtas de - 0,124 et - 0,122 respectivement.

[285] Pour information, nous ne retrouvons pas dans nos résultats, l'existence d'un tel lien entre la préoccupation pour le RVP et la fréquence de surf, le test de Chi-deux réalisé donnant un résultat non significatif (p = 0,473).
[286] La formulation est en réalité positive si on s'intéresse à l'habitude de dire la vérité. Comme nous nous intéressons au comportement habituel de mensonge, on peut considérer qu'il est formulé « négativement ».

Pour conclure sur cette troisième partie des résultats, nous pouvons dire que **nous validons les hypothèses H4 f et H4 g**, soulignant alors l'impact de l'attitude face à la communication des données sur les intentions comportementales de répondre et de mentir. Il apparaît ainsi que la manière pour un individu de se comporter face à un formulaire serait majoritairement dépendante des dispositions favorables (ou non) que celui-ci a développé à son égard.
On observe toutefois des différences entre ces deux types de comportements. Ainsi, contrairement à l'intention de réponse (IR) qui serait majoritairement expliquée par l'attitude (variable situationnelle, effet du présent), nous trouvons que l'intention de mentir serait surtout influencée par des variables individuelles (sexe et fréquence de surf) et par l'effet du passé (i.e. habitude de mentir).

6.4 L'impact de l'intention de répondre sur celle de poursuivre la relation

Comme précédemment, les conditions nécessaires à la réalisation de cette analyse ont été contrôlées au paragraphe 5.4 ou sont présentées en Annexe 11.9 (point D). De plus, les résultats obtenus (cf. tableau 11.19) indiquent une absence de multi-colinéarité extrême entre les variables indépendantes testées.

Nous testons l'impact de l'intention de répondre sur l'intention de poursuivre la relation au moyen d'une régression, dans laquelle nous introduisons successivement l'intention de répondre (bloc 1) et les variables de contrôle (bloc 2). Nous précisons que ces dernières comprennent les variables sociodémographiques et expérientielles ainsi que le niveau d'implication. Nous n'introduisons pas le comportement habituel de réponse puisque le lien avec l'intention de poursuivre la relation ne nous paraît pas fondé. Nous ajoutons cependant deux variables non explicitées jusqu'ici et qui concernent le type d'opérateur auquel le répondant est affilié (parmi les trois possibles, i.e. Orange, SFR et Bouygues) et son désir d'en changer[287]. Nous souhaitons en effet mesurer l'impact de la sollicitation sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données, toutes choses égales par ailleurs, ce qui nous oblige à contrôler l'intention de changer d'opérateur pour toute autre raison que celle testée ici (i.e. le fait d'être sollicité pour fournir des données).

[287] Cette variable correspond à la question 3 du questionnaire posée en page 5 (cf. Annexe 11.2) et est formulée de la manière suivante : « envisagez-vous de changer prochainement d'opérateur ? » (oui/non).

Nous faisons figurer ci-après le tableau de résultats donnant les indicateurs d'ajustement du modèle ainsi que les coefficients obtenus. Nous commentons ensuite les résultats concernant la validation (ou non) de l'hypothèse testée (i.e. H5).

Tableau 11.19 Régression visant à expliquer l'intention de poursuivre la relation (IPR)

	Bloc 1		Bloc 2		Colinéarité	
	B	Sig.	B	Sig.	Tol.	VIF
Constante	5,412	0,000	3,424	0,003		
Intention de répondre (IR)	1,137	**0,000**	1,079	**0,000**	0,923	1,083
Implication			0,153	**0,000**	0,870	1,149
Sexe			0,505	0,064	0,977	1,023
Age			0,090	0,413	0,836	1,196
CSP			0,064	0,761	0,801	1,249
Niveau d'Etudes			-0,388	**0,025**	0,772	1,295
EXP d'Internet			-0,148	0,473	0,935	1,070
UTIL (messagerie)			-0,019	0,942	0,689	1,452
UTIL (surf web)			0,257	0,244	0,700	1,428
Nom de l'opérateur			-0,061	0,739	0,909	1,100
Désir de changer d'opérateur			0,112	0,752	0,894	1,119
R^2	0,289		0,384		-	
R^2 ajusté	0,287		0,356		-	

Les résultats confirment l'impact de l'intention de répondre sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données (p = 0,000). De plus, l'influence a lieu dans le sens attendu : plus la personne a l'intention de répondre à la sollicitation (i.e. de fournir ses données), plus elle a l'intention de poursuivre la relation avec l'entreprise l'ayant sollicitée, et inversement (i.e. un individu n'ayant pas l'intention de répondre a une plus faible probabilité de rester client de ce fournisseur). Ceci nous conduit **à valider l'hypothèse H5**, confirmant ainsi les principes issus de la théorie de l'engagement exposés au chapitre 7. L'effet positif de l'intention de répondre est toutefois soumis au contrôle de deux variables individuelles : le niveau d'implication (p = 0,000) et le niveau d'études (p = 0,025)[288]. L'impact de l'implication sur l'intention de poursuivre la relation est positif (plus un individu apprécie la téléphonie mobile, plus il a l'intention de rester client chez son opérateur actuel), ce qui est, là aussi, logique. L'effet du niveau d'études est quant à lui négatif : plus le niveau d'études s'accroît, moins l'individu a l'intention de rester client de cet opérateur, suite à ce type de sollicitation. Ce résultat peut se comprendre aisément : un individu au niveau d'études

[288] Nous ne tenons pas compte ici de l'effet du sexe dont l'influence se rapproche du niveau de significativité (p=0,064) tout en lui étant légèrement supérieur (p = 0,05).

élevé qui n'apprécie pas de se voir sollicité par son opérateur actuel de téléphonie mobile (intention de répondre négative) envisagerait facilement les démarches à effectuer pour changer d'opérateur et disposerait des capacités (cognitives, intellectuelles, ..) nécessaires pour le faire. Au contraire, une personne peu éduquée se montrerait plus résignée dans ce type de situation, imaginant mal comment elle pourrait faire pour changer (facilement) d'opérateur.

En termes de poids respectif des variables influentes, il apparaît que l'intention de répondre explique, à elle seule, près de 29% du modèle, quand les variables de contrôle n'en expliquent que 7%. L'intention de poursuivre la relation serait donc majoritairement expliquée par l'intention de répondre (suivie de l'implication et du niveau d'études), ce qui est confirmé par l'examen des coefficients standardisés (bêtas = 0,510 ; 0,253 et – 0,130 respectivement).

Pour conclure sur cette quatrième partie des résultats, nous pouvons dire que **nous validons l'hypothèse H5**, soulignant l'impact de l'intention de répondre sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données. Il apparaît ainsi que la probabilité pour un individu de rester client de son fournisseur actuel, suite à une sollicitation de ses données personnelles, serait majoritairement dépendante des dispositions favorables (ou non) que celui-ci a développé à son égard et de son intention de répondre positivement (ou non) à la requête. Un individu ayant l'intention de fournir ses données serait donc plus enclin à poursuivre sa relation avec l'entreprise qu'un individu peu prompt à se dévoiler.
Cette influence serait toutefois soumise au contrôle de 2 variables individuelles (implication et niveau d'études) dont l'impact serait cependant mineur (7% de variance expliquée).

Nous avons vérifié toutes les hypothèses correspondant à notre modèle à tester et indiquées au point 3.2 de ce chapitre. Nous souhaitons cependant poursuivre cette analyse par l'examen de deux effets médiateurs qui apparaissent à l'examen du modèle (cf. figure 11.2). Il s'agit de l'attitude et de l'intention de répondre dont l'effet de médiation a été présenté à la figure 11.3.

6.5 Le test des effets médiateurs de l'attitude et de l'intention de répondre

Bien que n'ayant formulé d'hypothèses à ce sujet, nous avons remarqué que deux variables du modèle (l'attitude et l'intention de répondre) se positionneraient en tant que médiateurs de l'effet de l'évaluation de la sollicitation (i.e. valeur perçue, VAL) sur les intentions comportementales de réponse (i.e. IR et IM) (pour la première) et de l'effet de l'attitude

(ATT) sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données (IPR) (pour la seconde). Nous avons identifié au paragraphe 5.4 les conditions à vérifier pour tester l'existence d'un tel effet. Pour rappel, il faut :

- régresser Y (IR, IM et IPR) sur X (Valeur perçue et Attitude) et montrer que les coefficients correspondant aux variables indépendantes sont significatifs (condition 1)
- régresser M (Attitude et IR) sur X (Valeur perçue et Attitude) et montrer que les coefficients correspondant aux variables indépendantes sont significatifs (condition 2)
- régresser Y (IR, IM et IPR) sur X (Valeur perçue et Attitude) et M (Attitude et IR) et montrer que les coefficients correspondant aux variables médiatrices sont significatifs (condition 3) et que ceux des variables indépendantes ne le sont pas (condition 4)

Concrètement, nous vérifierons l'effet médiateur de chaque variable concernée (Attitude et IR) séparément[289]. Vous trouverez dans le tableau ci-dessous les résultats de ces analyses.

Tableau 11.20 Test de l'effet médiateur de l'attitude (ATT) et de l'intention de répondre (IR)

Conditions à valider	Résultats	
Variables indépendantes (VI) et dépendantes (VD) concernées	VI : valeur perçue (VAL) VD 1 : Intention de répondre (IR) VD 2 : Intention de mentir (IM)	VI : Attitude (ATT) VD : Intention de poursuivre la relation (IPR)
Variables médiatrices testées	Attitude face à la communication de données personnelles (ATT)	Intention de répondre (IR)
Condition 1 : lien entre la variable indépendante et la (les) variable(s) dépendante(s)	***Validée pour IR*** (B = 0,087 ; p= 0,000) ***Validée pour IM*** (B = 0,040 ; p= 0,000)	***Validée*** (B = 0,263 ; p= 0,000)
Condition 2 : lien entre la variable indépendante et la variable médiatrice	***Validée*** (B = 0,336 ; p= 0,000) (cf. tableau 11.16)	***Validée*** (B = 0,149 ; p= 0,000) (cf. tableau 11.17)
Condition 3 : lien entre la variable médiatrice et la (les) variable(s) dépendante(s)	***Validée pour IR*** (B = 0,149 ; p= 0,000) ***Validée pour IM*** (B = - 0,046 ; p= 0,000) (cf. tableaux 11.17 et 11.18)	***Validée*** (B = 1,079 ; p= 0,000) (cf. tableau 11.19)
Condition 4 : l'effet de la variable indépendante est nul quand l'effet de la variable médiatrice est contrôlé	Non validée pour IR (B = 0,123 ; p= 0,000 pour ATT ; B = 0,034 ; p= 0,000 pour VAL) ***Validée pour IM*** (B= - 0,043; p= 0,000 pour ATT ; B= - 0,004 p= 0,569 pour VAL)	Non validée (B = 0,696 ; p= 0,000 pour IR ; B = 0,148 ; p= 0,000 pour ATT)
Conclusions	Médiateur partiel de l'IR Médiateur complet de l'IM	Médiateur partiel de l'IPR

[289] En ce qui concerne l'attitude, deux variables dépendantes sont concernées (IR et IM). Des analyses correspondant à chaque variable dépendante seront donc menées séparément. Nous faisons toutefois figurer les résultats de ces deux analyses, conjointement, dans la deuxième colonne du tableau 11.20.

Pour conclure sur ce point, on peut donc dire que l'attitude face à la communication des données (ATT) est un médiateur partiel de l'effet de l'évaluation de la sollicitation (i.e. de la valeur perçue) sur l'intention de fournir les données demandées (i.e. de répondre IR) et un médiateur complet de l'effet de l'évaluation de la sollicitation (i.e. de la valeur perçue) sur l'intention de fournir des données erronées (i.e. de mentir IM). Pour mieux appréhender l'intention de répondre, il est donc nécessaire d'inclure, au sein du modèle (cf. tableau 11.17), l'effet de la valeur perçue, en plus de celui lié à l'attitude. Au-delà des deux blocs introduits précédemment, nous y ajoutons donc un troisième[290], correspondant à la valeur perçue (VAL).

L'impact de cette variable (i.e. la valeur perçue), bien que significatif (p= 0,000), n'explique que 4% du nouveau modèle, contre 44% pour l'attitude. L'effet direct de la valeur perçue sur l'intention de répondre doit donc être pris en compte, malgré un pouvoir explicatif limité.

En ce qui concerne l'intention de mentir, nous n'avons pas à tester de nouveau modèle puisque l'attitude apparaît comme un médiateur complet de l'effet de la valeur perçue.

Le deuxième effet médiateur testé concerne l'intention de répondre. Il ressort, à la lecture des résultats, que cette variable serait un médiateur partiel de l'effet de l'attitude face à la communication des données sollicitées sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données (IPR). Pour mieux appréhender l'intention de poursuivre la relation, nous testons donc un nouveau modèle, en introduisant l'intention de répondre dans un premier bloc, l'attitude dans un second et les variables de contrôle dans un troisième. L'impact de l'attitude, bien que significatif (p= 0,000), n'explique que 5% du modèle contre 29% pour l'intention de répondre et 5% pour les variables de contrôle dont l'effet s'avère significatif (i.e. implication et niveau d'études). L'effet direct de l'attitude sur l'intention de poursuivre la relation doit donc être pris en compte même si son pouvoir explicatif est relativement faible.

Pour rappel, nous n'avions pas formulé d'hypothèse concernant l'impact de l'intention de mentir (IM) sur l'intention de poursuivre la relation (IPR) avec l'entreprise ayant sollicité les données, pour les raisons évoquées au chapitre 7 (paragraphe 4.4.2). Par souci de cohérence, nous avons décidé, dans le cadre de cette expérience, de vérifier l'existence éventuelle d'un lien de cause à effet entre ces deux variables, ce que nous soupçonnions, sans l'avoir énoncé.

[290] En réalité, la valeur perçue est introduite dans le deuxième bloc, les variables de contrôle faisant partie du troisième.

Le modèle testé à cet effet[291] (qui consistait à régresser l'intention de poursuivre la relation sur l'intention de mentir) expliquerait 35% de la variance, l'intention de mentir expliquant à elle seule environ 10%, contre 20% pour l'attitude et 5% pour les variables de contrôle dont l'effet est significatif (i.e. implication et niveau d'études). L'intention de mentir aurait ainsi un effet négatif sur l'intention de poursuivre la relation, les individus prêts à mentir se montrant peu enclins à rester client de leur fournisseur actuel.
Le test de l'effet médiateur de l'intention de mentir a également été réalisé. Il indique que cette intention jouerait un rôle de médiateur partiel entre l'attitude et l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données.

Nous terminons ce chapitre par une discussion des résultats et des limites de l'expérience 2.

7 Récapitulatif des résultats et limites inhérentes à cette expérience

7.1 Récapitulatif des résultats et commentaires

Nous présentons ici le récapitulatif des hypothèses testées dans le cadre de cette deuxième expérience puis commentons brièvement en-dessous les principaux résultats obtenus.

Tableau 11.21 Récapitulatif des hypothèses à tester et des résultats obtenus (Expérience 2)

Hypothèses testées dans l'expérience 2		Résultats
H 3 (e) (10)	Plus la préoccupation pour le RVP est élevée (vs faible), moins l'attitude face à la communication des données est favorable (vs défavorable)	**Validée (p = 0,022)**
H 1.3 (d) (11)	Plus la quantité de données demandées est élevée (vs. faible), moins (plus) la valeur perçue à répondre est élevée	**Validée (p = 0,076)**
H 1.4 (d) (12)	La valeur perçue à répondre diffère selon le contexte dans lequel les données sont sollicitées (Jeu-Concours / Newsletter / Espace Personnel)	Non validée (p = 0,479)
H 2 (d) (13)	Plus la valeur perçue à répondre est élevée (vs faible), plus l'attitude face à la communication des données sollicitées est favorable (vs défavorable)	**Validée (p = 0,000)**
H 4 (f, g) (14, 15)	Plus l'attitude face à la communication des données sollicitées est favorable (vs défavorable), plus (vs moins) l'individu a l'intention de répondre (de communiquer les données demandées) (f) et moins (vs plus) il a l'intention de mentir (g)	**Validées (p = 0,000) (p = 0,000)**
H 5 (16)	Plus (vs moins) l'individu a l'intention de communiquer les données demandées, plus (vs moins) il a l'intention de poursuivre la relation avec l'entreprise qui l'a sollicité	**Validée (p = 0,000)**

[291] Par souci de simplification, nous ne présentons pas les résultats complets de ce modèle ni du test de l'effet médiateur, qui sont néanmoins disponibles auprès de l'auteur.

Les résultats de cette deuxième expérience prouvent que la quantité de données demandées aurait une influence non négligeable sur le comportement de divulgation en ligne, à travers son influence sur l'évaluation que l'individu fait de la sollicitation, représentée ici par la valeur perçue à répondre. Le fait d'accroître le nombre de données demandées pourrait ainsi diminuer la probabilité que la personne remplisse le formulaire, en réduisant la valeur perçue à fournir de telles informations. Des variables exogènes, tel le niveau d'implication vis-à-vis de la catégorie de produit/services concernée, semblent également susceptibles d'influencer cette phase d'évaluation. Ces résultats viennent à la fois confirmer et compléter ceux obtenus lors de l'expérience 1. Nous avions déjà constaté l'impact de la quantité de données sur deux autres variables perceptuelles (i.e. la sensibilité et la pertinence perçues des données sollicitées). Nous déduisons ici que la quantité influence également la valeur perçue à répondre, complétant ainsi le processus d'évaluation de la sollicitation.
L'hypothèse concernant l'effet du contexte sur la valeur perçue à répondre n'a, par contre, pas pu être validée, faute d'écart suffisant dans les moyennes obtenues pour chaque contexte étudié (i.e. participation à un jeu-concours, inscription à une newsletter et création d'un espace personnel sur le site). Pour plusieurs raisons expliquées au paragraphe 6.1.2 de ce chapitre, nous pensons que cela ne veut pas dire que le contexte n'aurait aucun impact sur le processus de réponse face à une sollicitation de données personnelles. Nous estimons plutôt qu'il serait nécessaire de tester cet effet d'une autre manière, ce que nous aurons l'occasion de faire dans l'expérience 3, présentée au chapitre suivant (i.e. chapitre 12).

La seconde partie des résultats concerne l'influence relative de la situation de collecte (à travers la valeur perçue à répondre) et des caractéristiques individuelles (i.e. de la préoccupation pour le respect de la vie privée) sur l'attitude de l'individu face à la communication de ses données. Comme dans l'expérience 1, il apparaît que les caractéristiques situationnelles seraient prépondérantes et pourraient dépasser les simples convictions personnelles, confortant ainsi ce résultat.

La dernière partie des résultats concerne la suite du processus de réponse (i.e. de l'attitude aux intentions comportementales) qui n'avait jamais été testée jusqu'à présent. Il ressort que l'attitude est bien un médiateur de l'effet de l'évaluation de la sollicitation (évaluée ici par la valeur perçue à répondre) sur les intentions comportementales de réponse, partiel en ce qui concerne l'intention de répondre, et complet en ce qui concerne l'intention de mentir. La prédisposition de l'individu relative au formulaire à compléter (mesurée par l'attitude) est

donc bien un antécédent de l'intention de se comporter face à cette situation. Nous devrons aussi ajouter à notre modèle un lien direct entre la valeur perçue et l'intention de répondre.
L'attitude aurait également un effet direct sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données, confirmant ainsi la prédiction de Helme-Guizon et Amato (2004) qui indiquent que « même un acte aussi simple qu'exprimer une attitude peut conduire à de l'engagement ». Notre modèle doit ainsi être complété par l'ajout d'un lien direct entre l'attitude (ATT) et l'intention de poursuivre la relation (IPR).
Nos résultats confirment enfin l'influence des intentions comportementales de réponse (IR et IM) sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données (IPR), ce que nous avions supposé et formulé pour la première (i.e. l'intention de répondre) et simplement imaginé, sans pouvoir pour autant le justifier, pour la seconde (i.e. l'intention de mentir)[292]. Le modèle à tester dans le cadre de cette deuxième expérience doit finalement être complété d'un lien entre l'intention de mentir et l'intention de poursuivre la relation[293].

Nous résumons dans la figure ci-après (cf. figure 11.7) l'ensemble des hypothèses validées et invalidées, auxquelles nous ajoutons les relations « validées hors hypothèses ».

Comme dans toute démarche expérimentale, celle menée ici n'est pas exempte de limites qu'il convient de présenter maintenant. Sans remettre en cause totalement les résultats obtenus, elles amènent toutefois à prendre ces derniers avec précaution et à réfléchir aux solutions à mettre en œuvre dans l'expérimentation suivante (i.e. l'expérience 3) pour en réduire l'effet.

7.2 Les limites de l'expérience 2

Trois limites majeures ont pu être détectées dans le cadre de cette deuxième expérience. Elles concernent l'opérationnalisation des facteurs manipulés, le biais du répondant (et celui de non-réponse) et la mise en situation des répondants, qui viennent toutes réduire la validité des résultats et devront donc être pris en compte dans la dernière expérience (i.e. Expérience 3).

Par limite relative à l'opérationnalisation des facteurs, nous faisons essentiellement référence à la manière dont nous avons choisi de manipuler le contexte de sollicitation dans le cadre de cette expérience. Le choix des contextes testés et la manière de les opérationnaliser pourrait en effet expliquer, en partie au moins, l'absence d'influence de ce facteur.

[292] Nous avions justifié l'absence de cette hypothèse au paragraphe 4.4.2 du chapitre 7.
[293] En réalité, ce lien existait déjà et nous pouvons donc le garder alors qu'il était abusif dans le modèle initial.

Figure 11.7 Récapitulatif des hypothèses validées dans le cadre de l'expérience 2

En trait plein, les hypothèses validées ——

En pointillé, les hypothèses non validées .....

En double trait, les relations validées hors hypothèse ═══

Variables de contrôle

- *Niveau d'implication vis-à-vis de la catégorie de produits (IMPL)*
- *Variables sociodémographiques (âge, sexe, CSP, niveau d'éducation)*
- *Expérience (EXP)et Utilisation (UTIL) d'Internet*
- *Comportement habituel de réponse (COMP_HAB)*

Nous pensons que les contextes choisis ici n'étaient peut être pas suffisamment opposés, et surtout, que leur manipulation n'était pas suffisamment « réaliste » pour trouver un effet significatif. Nous devrons donc remédier à cela dans le cadre de l'expérience 3, ce qui nous permettra de tester de manière plus « sûre » l'impact du facteur « Contexte de sollicitation ».

La deuxième limite fait référence aux différents biais liés au choix de l'échantillon interrogé et qui ont déjà été vus au chapitre 8 (paragraphe 2.3). Ceux-ci concernent principalement le biais du répondant (i.e. la manière de se comporter face à l'expérimentateur) et le biais de non réponse. Ce dernier, bien que non négligeable est toutefois fortement réduit ici dans la mesure où plus de 60% des répondants sollicités ont accepté de participer à l'expérience. De plus, ce biais n'est pas lié au sujet de l'étude, ce qui vient encore limiter sa portée. La limite liée au biais du répondant est plus préoccupante, d'autant qu'elle est renforcée ici par les caractéristiques mêmes de l'échantillon, composé de personnes qui se sont volontairement inscrites à ce panel pour participer à des enquêtes. Nous avons donc en partie affaire à des « professionnels » des questionnaires. Il y a alors fort à parier que le simple fait d'être inscrit à un panel modifie l'attitude et le comportement du répondant, ce qui est d'ailleurs confirmé par plusieurs recherches sur le sujet.

En particulier, les motivations pour lesquelles les participants ont souhaité s'inscrire à un panel auraient une incidence profonde sur la manière de se comporter face aux enquêtes (Deutskens, Ruyter et Wetzels 2005). Même s'il existe des techniques visant à contrôler les effets de l'appartenance au panel, nous n'avons pas pu les mettre en œuvre ici, faute de matériau pour le faire[294]. Nous estimons que le choix de clients réels dans la dernière expérience, nous permettra de limiter substantiellement ces deux biais.

La dernière limite identifiée est propre à réduire la validité interne de cette expérience. En effet, même si les répondants interrogés ici l'ont été par le biais d'un questionnaire en ligne, ce type de mise en situation n'est pas tout à fait identique à un contexte réel, d'autant que les participants n'avaient pas à naviguer sur un site et qu'ils n'avaient pas à compléter, non plus, le formulaire en ligne. Là encore, mettre les individus dans une situation réelle constituerait une démarche plus adéquate pour tester de manière « efficace » l'impact de l'exposition à une sollicitation de données personnelles sur l'ensemble des variables du processus de réponse.

[294] Il faudrait en effet avoir une connaissance précise des caractéristiques du panel et de ses membres.

Conclusion du chapitre 11

Ce chapitre a présenté les résultats d'une deuxième expérimentation visant à tester, d'une part, l'effet de l'exposition à une sollicitation de données personnelles sur l'évaluation de celle-ci et, d'autre part, l'impact de cette évaluation sur la suite du processus de réponse, i.e. de l'attitude jusqu'aux intentions comportementales.

Après avoir indiqué les objectifs de l'expérimentation, nous avons présenté les facteurs manipulés, les variables contrôlées et la manière dont l'expérience s'est déroulée. Deux hypothèses spécifiques à l'étude ont ensuite été développées, conduisant à la présentation du modèle à tester. L'exposition des caractéristiques des répondants a ensuite permis d'insister sur les biais engendrés par le choix d'un échantillon de panélistes.

Les analyses préalables ont permis de souligner la validité du plan d'expérience. Nous avons également été soucieux de vérifier l'effectivité des manipulations réalisées. Après avoir établi la qualité psychométrique des instruments de mesure, nous avons présenté le plan de traitement utilisé pour mener à bien nos analyses. Ayant démontré la possibilité de tester nos hypothèses dans le respect des conditions exigées, nous avons procédé à l'étude des résultats.

Les résultats obtenus confirment l'hypothèse d'un effet de la quantité de données demandées sur l'évaluation de la sollicitation par le consommateur, ce qui avait déjà été démontré dans l'expérience 1, même si les variables de perception étaient différentes de celle étudiée ici.
Nous confirmons aussi l'existence d'antécédents situationnels (par le biais de l'évaluation de la sollicitation) et individuels (par l'intermédiaire de la préoccupation pour le RVP) à l'attitude face à la communication des données demandées, les premiers se révélant largement prépondérants, ce que nous avions également démontré dans l'expérience 1. Cette attitude s'impose d'ailleurs comme une variable médiatrice majeure, influençant à la fois les intentions comportementales de réponse (i.e. de répondre et de mentir) et l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données. Enfin, cette dernière serait également influencée par les intentions comportementales de réponse, validant ainsi nos hypothèses liées à la théorie de l'engagement. Cette dernière indique en effet qu'un individu peut maintenir la relation avec son fournisseur parce qu'il se sent engagé par ses actes antérieurs (en l'occurrence ici par le fait d'avoir accepté de fournir des données personnelles).

Partie III – EXPLIQUER

Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne

Analyse empirique des résultats

CHAPITRE 12

RÉPONSES DU CONSOMMATEUR FACE A UNE SOLLICITATION DE SES DONNÉES PERSONNELLES (ÉTUDE 5)

« The additional information provided when using the Web to collect data (e.g., automatically-generated log files, visitor tracking programs etc.) can provide a valuable insight into understanding nonresponse and response behaviors ».

Bosnjak and Tuten (2001)

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

Introduction du chapitre 12

Les deux premières expériences, menées dans le cadre de ce travail doctoral, se sont déroulées en « laboratoire ». Dans ce cas, le chercheur crée une situation « artificielle », propice à la manipulation de variables prédictives et mesure leur impact sur les réactions des unités expérimentales aux différents traitements. Quatre facteurs ont ainsi été manipulés, trois dans la première expérience et deux dans la seconde. Les critères utilisés pour juger de l'efficacité des traitements concernaient les premières étapes du processus de réponse d'un individu face à une sollicitation de ses données personnelles, à savoir de l'évaluation de la sollicitation jusqu'aux intentions comportementales de réponse.

Cette troisième expérience se distingue des deux premières par plusieurs aspects. Tout d'abord, l'expérimentation a lieu cette fois-ci sur le terrain, dans un milieu dit « naturel ». Elle a en effet fait l'objet d'un partenariat avec un site existant, qui a organisé une collecte de données auprès d'un échantillon issu de son fichier de clientèle. Les unités expérimentales sont donc constituées de « vrais » clients, et non pas d'étudiants (cas de l'expérience 1) ou de panélistes (comme dans l'expérience 2). Ces individus sont, en outre, mis dans une situation réelle et non « artificielle », comme c'est généralement le cas dans une expérimentation. Ce choix n'est pas sans incidence. Il offre l'avantage de pouvoir mesurer le comportement réel plutôt que d'étudier des comportements déclarés ou des intentions comportementales. Il permet donc de tester de façon réaliste les effets de l'exposition à une sollicitation de données personnelles sur la réponse de l'internaute. Cela donne alors la possibilité de réduire certaines sources d'erreur, notamment celles qui sont liées au statut de répondant puisque, dans ce cas, ce dernier n'a même pas conscience d'être impliqué dans une démarche expérimentale. Dans le même temps, cela réduit les possibilités de contrôler l'effet de certaines variables externes, susceptibles de biaiser les résultats. En effet, sans pour autant affranchir le chercheur des règles fondamentales de la méthode expérimentale, sortir du milieu protégé du laboratoire rend le principe du contrôle plus difficile (Bhaskar, 1975).

Après avoir identifié les objectifs de l'expérience 3 (Section 1), nous présenterons le design expérimental (Section 2), avant d'énoncer les hypothèses de recherche (Section 3). Nous évoquerons alors le recueil des données (Section 4) et les caractéristiques des répondants ayant participé à l'expérience (ainsi que les biais que ces choix soulèvent) (Section 5). Nous indiquerons ensuite le type d'analyses menées pour tester les hypothèses énoncées (Section 6). Nous présenterons enfin les résultats obtenus (Section 7) avant de les discuter (Section 8).

1 Objectifs et apports de l'expérience 3

Plusieurs objectifs sont assignés à cette troisième expérience.

Il s'agit, tout d'abord, d'évaluer l'impact d'une sollicitation de données personnelles en ligne, dans un contexte réel. Une expérience ne peut avoir lieu sans la coopération de l'ensemble des unités expérimentales. Or, le fait d'être sollicité pour participer conduit certains individus à adopter un rôle qui amène des réponses différentes de celles qu'ils donneraient dans la réalité. Mener l'expérimentation dans un contexte « naturel » permet donc de limiter ce biais. De plus, cela permet d'évaluer la phase finale du processus de réponse : le fait d'accepter, ou non, de compléter le formulaire contenant les informations personnelles demandées.

La réponse à une sollicitation de données peut prendre différentes formes. Outre la mesure du taux de participation, à travers la validation du formulaire une fois celui-ci complété, il est également intéressant, notamment managérialement parlant, d'évaluer le taux de réponse (nombre de réponses sur l'ensemble des champs proposés) ainsi que le taux d'abandon (nombre d'individus qui arrêtent avant la dernière question et/ou qui ne valident pas le formulaire). Peuvent ainsi être identifiés quatre principaux types de comportements de réponse : la réponse complète (fait d'être allé au bout du processus et d'avoir répondu à toutes les questions), la non réponse partielle (être allé au bout du questionnaire mais avoir laissé certaines questions sans réponse) ou totale (absence de participation) et enfin l'abandon (avoir arrêté avant la fin). Cette dernière catégorie peut elle-même se scinder en deux, selon que la personne a répondu ou non à toutes les questions avant de stopper (que l'on peut appeler respectivement abandon avec réponse complète et abandon avec non réponse partielle).

Les auteurs qui travaillent sur la méthodologie d'enquête étudient régulièrement l'impact du design du questionnaire sur la réponse des personnes sollicitées. Certains soulignent que des éléments du design auraient un impact limité sur la décision de participer mais un impact plus prononcé sur le taux d'abandon. Le design expérimental prévu dans cette expérience permettra de mesurer l'existence éventuelle d'un tel effet, dans le cadre d'une sollicitation de données personnelles.

Un autre élément intéressant à analyser concerne le comportement de réponse aux questions sensibles (notamment aux questions portant sur des données de contact ou des données financières). Nous chercherons donc à mesurer l'impact des caractéristiques situationnelles sur le niveau de dévoilement de l'individu.

Enfin, une autre question qui se pose régulièrement, que ce soit dans le cadre des enquêtes par sondage ou dans le cas qui nous intéresse ici (la sollicitation de données personnelles), est de savoir si les gens qui répondent diffèrent de ceux qui ne le font pas. Pour y répondre, il est nécessaire d'obtenir des informations auprès des non répondants. Un des objectifs de cette troisième expérience sera d'identifier le profil du répondant et du non-répondant et de les comparer afin de mettre au jour d'éventuelles différences.

Nous abordons, dans la section suivante, le design mis en place pour cette expérience.

2 Le design expérimental

L'expérimentation a été menée en partenariat avec la Société Millemercis[295], qui dispose de plusieurs sites en propre (Millemercis.com, Millemercismariage.com, …) et d'une base de données mutualisée de plus de 6 millions d'individus (la Base Elisa[296]). L'expérience a été lancée début septembre 2005 et s'est terminée trois mois plus tard (i.e. fin novembre 2005).

Les facteurs manipulés au cours de l'expérimentation sont tout d'abord présentés (2.1) avant d'évoquer les variables contrôlées (2.2) et la mise en œuvre du design expérimental (2.3).

2.1 Les facteurs manipulés

En accord avec l'entreprise Millemercis, 4 facteurs[297] ont été manipulés au sein de cette expérimentation, chacun limité à deux niveaux :
1) la familiarité avec l'entreprise sollicitante (déjà manipulée au cours de l'expérience 1) ;
2) la quantité de données demandées et 3) le contexte de sollicitation (tous deux manipulés conjointement au sein de l'expérience 2) ;
4) l'ordre des questions (non manipulé jusqu'ici).

Il s'agit donc d'un design 2x2x2x2 en inter-sujets : chaque participant ne sera soumis qu'à un seul des 16 traitements possibles[298]. Nous remarquons que le dernier facteur manipulé (ordre

[295] Nous remercions les dirigeants de Millemercis (Thibaut Munier et Yseulis Costes) de nous avoir fait confiance et de nous avoir donné les moyens de réussir. Nous remercions aussi les personnels qui furent associés à cette démarche (et notamment Stéphanie Gautier et Anne Guérin) pour leur forte implication dans ce projet.
[296] Cette base correspond à un programme relationnel auquel l'ensemble des membres sont invités à participer.
[297] En réalité, comme nous le justifierons ultérieurement, seuls les trois premiers seront réellement étudiés.

des questions) ne s'inscrit pas dans notre modèle à tester, volontairement limité à 4 facteurs manipulés, afin de conserver un caractère parcimonieux. Il ne sera donc pas détaillé outre mesure, même si nous en expliquerons brièvement le principe puisqu'il est inclus dans le design expérimental. Nous ne développerons pas non plus d'hypothèse à son sujet. Enfin, nous ne testerons pas son impact dans le cadre des résultats obtenus, tout en nous réservant la possibilité d'aborder la question dans le dernier chapitre de la thèse correspondant à la discussion et à la mise en perspective des résultats de la recherche.

Nous présentons brièvement la manière dont ces facteurs ont été opérationnalisés dans le tableau 12.1. Nous reviendrons plus précisément sur la manipulation de ces facteurs dans le paragraphe présentant la mise en œuvre du design expérimental (i.e. paragraphe 2.3). A noter que, contrairement aux deux expériences précédentes, qui portaient sur une sollicitation de données par le biais d'un formulaire à remplir (que nous définissons comme un ensemble de questions d'ordre personnel regroupées sur une même page), nous étudions ici la réponse à un questionnaire (plusieurs questions réparties sur plusieurs pages). Nous nous situons cependant toujours dans le contexte d'une sollicitation de données personnelles, dans la mesure où ce questionnaire est précédé d'un court formulaire d'identification permettant de relier les réponses à une personne identifiée et identifiable. Même si ce cadre diffère sensiblement de celui étudié précédemment, il se justifie par la pratique de l'entreprise partenaire qui a l'habitude de soumettre ses clients à ce type d'instrument de collecte. Il constitue en outre une extension de la sollicitation de données par le biais d'un formulaire, ajoutant ainsi de la validité externe à notre plan expérimental.

Tableau 12.1 Facteurs manipulés dans l'expérience 3

Facteurs	niveaux	Opérationnalisation	Indication des niveaux
Familiarité avec l'entreprise	2	Sélection de l'individu sur la base de l'ancienneté de sa participation au programme Elisa	- familiarité faible : individu « récent » (inscrit depuis moins de 3 mois) - familiarité élevée : individu « ancien » (inscrit depuis plus d'un an)
Quantité de données demandées	2	Nombre de données à fournir (correspondant au nombre de questions posées dans le questionnaire) (questionnaire administré en ligne, avec 1 question par écran)	- quantité faible : questionnaire « court » (16 questions posées) - quantité élevée : questionnaire « long » (42 questions posées)

[298] Selon la terminologie utilisée par Millemercis, ces traitements sont appelés des cibles. Ainsi, chaque traitement correspond à un « IdCible » numéroté de 1 à 16. Dans la suite de ce chapitre, nous utiliserons l'un ou l'autre de ces termes (i.e. traitement et cible), de manière interchangeable.

Contexte de sollicitation	2	Occasion pour laquelle on sollicite les informations personnelles et types de bénéfices proposés en échange (se retrouve dans l'intitulé de la ligne sujet et au sein même de l'email de sollicitation)	- contexte « enquête » (pour « mieux [vous] connaître ») (service davantage personnalisé) - contexte participation à un « jeu-concours » (espérance d'un gain, ici un voyage aux Antilles)
Ordre des questions	2	Place de la question sur les Revenus au sein du questionnaire	- questionnaire « sensible » (question sur les Revenus au début du questionnaire, i.e. à la question 6) - questionnaire « neutre » (question posée à la question 15, soit au milieu ou à la fin du questionnaire, selon la longueur de celui-ci [i.e. court vs long])

On peut remarquer que la manière dont ces facteurs sont manipulés ici diffère sensiblement de leur opérationnalisation, dans le cadre des deux expériences précédentes. D'abord, il a fallu tenir compte de différentes contraintes opérationnelles. Concernant la familiarité par exemple, il n'y aurait eu aucun sens à opposer des membres du programme Elisa avec des non membres, comme nous l'avions fait dans le cadre de l'expérience 1 (i.e. abonnés à l'opérateur de téléphonie mobile sollicitant les données vs. sollicitation par un opérateur inconnu).

Ensuite, cette expérimentation s'inscrit dans le cadre du partenariat avec une entreprise et nous devions donc nous adapter à sa manière habituelle de procéder. Ainsi, le questionnaire « long » visait à récolter un maximum d'informations sur les consommateurs interrogés. Plutôt que d'opposer 5 à 20 données comme nous l'avons fait dans les expériences 1 et 2, nous opposons ici 16 à 42 questions. Ganassali et Moscarola (2002) indiquent que la longueur moyenne d'un questionnaire s'établirait en-deçà de 30 questions, les enquêtes sur le web pouvant être légèrement plus longues. En opposant 16 et 42 questions, nous nous situons donc bien dans une répartition respectivement faible et forte du nombre de données demandées.

De même, concernant le contexte, nous opposons ici l'« enquête » (selon la terminologie utilisée par Millemercis) au jeu-concours, sans pouvoir reprendre les autres contextes testés dans l'expérience 2 (i.e. Newsletter et Espace Personnel). Le contexte « enquête » est en effet rapidement apparu comme le plus à même de faire l'objet d'une opération lancée dans le cadre du programme Elisa et sa justification est donc fortement liée à son caractère réaliste. De plus, bien que différent du contexte « Espace personnel » (étudié dans l'expérience 2), il s'en approche toutefois, de par les bénéfices qu'il permet d'offrir aux répondants (i.e. une meilleure personnalisation de l'offre, suite aux informations récoltées).

Au final, l'opérationnalisation proposée ici s'inscrit parfaitement dans le cadre de notre modèle, tout en correspondant aux attentes de notre partenaire. Cette situation offre autant d'avantages que d'inconvénients. Elle accroît la validité externe de notre recherche, en permettant de tester de nouvelles « combinaisons » de facteurs. A l'opposé, elle réduit les possibilités de comparer les résultats avec les expériences précédentes, dans la mesure où la manipulation de certains facteurs y est différente.

2.2 Les variables contrôlées

Nous présentons les variables à contrôler (2.2.1), puis le type de contrôle opéré (2.2.2).

2.2.1 Le type de variables à contrôler

Plusieurs raisons (autres que les facteurs manipulés dans le cadre de cette expérience) pourraient expliquer le comportement de réponse à la sollicitation. Parmi celles-ci, on trouve notamment : l'attitude qui s'est formée suite aux réponses aux premières questions, l'existence de problèmes techniques ou la survenue d'un événement extérieur (ex : un appel téléphonique). Toutefois, comme ces raisons devraient être également réparties entre les traitements, le comportement de réponse ne devrait pas être influencé outre mesure par ces différents éléments. De plus, des variables telles que le design du questionnaire, la présentation des questions, etc … identifiées au sein de la littérature comme potentiellement influentes (Knapp et Heidingsfelder 2001) seront maintenues constantes dans toutes les cellules, de manière à éviter toute influence de leur part.

D'autres variables exogènes, susceptibles d'influencer nos résultats, devront toutefois être contrôlées. Deux catégories de variables externes ont été identifiées dans le cadre de cette expérience : des variables dites « classiques », déjà contrôlées dans les expériences précédentes et des variables particulières au niveau d'analyse (le comportement de réponse).

Les variables dites « classiques » font référence aux données sociodémographiques (âge, sexe, CSP et niveau d'études), à l'expérience (EXP) et à la fréquence d'utilisation (UTIL) d'Internet. A noter que le niveau d'implication vis-à-vis de la catégorie de produits/services (ici vis-à-vis des services proposés par le programme Elisa) est omis de cette démarche, car nous soupçonnions un manque de variance à son sujet. Tous les individus de l'échantillon

étant membres du programme Elisa, il est en effet fort à parier qu'ils se sentent tous (à des degrés divers cependant) impliqués par le type de services proposés par ce programme.

Un deuxième type de variables de contrôle (particulières au niveau d'analyse) a été introduit dans le questionnaire. Il s'agit du comportement habituel de réponse (COMP_HAB), déjà contrôlé dans l'expérience 2, et de la qualité perçue de la relation avec l'entreprise sollicitante (QUALITE), plus précisément ici, la qualité perçue du programme relationnel nommé Elisa.

2.2.2 Le type de contrôle

Nous contrôlerons l'effet de ces variables « externes » par trois biais :

1) une affectation aléatoire des individus aux différents traitements, qui permet une répartition *a priori* homogène des influences externes sur les unités tests

2) un contrôle *a priori* pour les variables d'ordre sociodémographique

Nous souhaitions que les individus affectés aux différents traitements (16 au total) aient un profil sociodémographique qui soit le plus proche possible (par exemple, la même répartition homme/femme quel que soit le traitement). Dans la réalité, la distinction selon que les individus se sont récemment (familiarité faible) ou anciennement (familiarité élevée) inscrits sur Elisa, nous a conduits à réduire cette ambition. En effet, le recrutement des membres du programme Elisa évolue régulièrement, conjointement à l'évolution du profil de l'internaute moyen (i.e. davantage de très jeunes et de séniors, plus de femmes et de provinciaux). Au final, compte tenu de cette évolution de profil des membres du programme Elisa (source de recrutement dans le cadre de cette expérience), il apparaît, qu'à ancienneté égale, les individus des différentes cibles ont un profil sociodémographique quasi identique eu égard à leur sexe, leur âge, leur région d'habitation et leur CSP (cf. Annexe 12.1).

En plus, puisque les répondants seront sollicités par email et qu'il pourrait exister un écart entre la répartition initiale des individus aux différents traitements (i.e. au moment de l'envoi de l'email) et la répartition « finale » (i.e. après clic sur l'email), nous contrôlerons également l'effet de ces variables *a posteriori*, par inclusion dans les analyses (cf. point suivant).

3) un contrôle statistique pour toutes les variables : l'ensemble des variables seront incluses dans les analyses en tant que co-variables, après vérification de leur adéquation à des tests statistiques (i.e. normalité et absence de colinéarité avec d'autres variables).

Nous détaillons ci-dessous les instruments de mesure adoptés pour chaque variable exogène. Nous avons essayé, si possible, de conserver une mesure similaire à celle utilisée dans les expériences précédentes. La mesure du comportement habituel sera toutefois légèrement différente de celle utilisée dans l'expérience 2, afin d'être adaptée au modèle testé ici. Nous distinguons ainsi le comportement habituel de réponse aux sollicitations de données de celui lié aux offres proposées dans le cadre du programme Elisa. La qualité perçue n'ayant jamais été mesurée auparavant fera, en outre, l'objet d'une attention particulière. Nous reviendrons sur la validité des instruments relatifs à ces deux variables dans la suite de ce chapitre.

Tableau 12.2 Instruments de mesure utilisés pour les variables externes contrôlées

Variables	Items utilisés	Types de mesures
Variables sociodémographiques		
Age	Date de naissance (1 item) JJ/MM/AAAA	Continue ou ordinale à 7 modalités (de 1 : - de 18 ans, à 7 : + de 65 ans)
Sexe	Civilité (1 item) M/Mme/Melle	Recodée en variable binaire (1 : Homme ; 2 : Femme)
CSP	Situation professionnelle (actifs/inactifs) et profession (dont autres) (2 items)	Recodée en variable ordinale à 4 modalités (0 : inactif ; 1 : CSP - ; 2 : CSP moyen ; 3 : CSP +)[299]
Niveau d'études	Plus haut niveau d'études atteint (1 item)	Ordinale à 6 modalités (de 1 : collège, à 6 : Bac+5 et plus)
Variables expérientielles		
Utilisation d'Internet	Fréquence de consultation de la messagerie (1 item)	Ordinale à 5 modalités (de 1 : plusieurs fois par jour, à 5 : moins d'une fois par semaine)
Expérience d'Internet	Ancienneté (1 item)	Ordinale à 3 modalités (1 : - de 2 ans ; 2 : 2 à 5 ans ; 3 : + de 5 ans)
Variables comportementales		
Comportement habituel de réponse face aux sollicitations de données personnelles	- habitude de répondre (1 item) - habitude de tout donner (1 item) - habitude d'abandonner avant la fin (2 items)	Likert en 7 points (de 1 : pas du tout d'accord à 7 : tout à fait d'accord)
Comportement habituel de réponse face aux emails envoyés dans le cadre du programme Elisa	- habitude de cliquer dessus (1 item)	Ordinale à 4 modalités (de 1 : très souvent à 4 très rarement)
Variables attitudinales et/ou perceptuelles		
Qualité perçue du programme Elisa	- qualité perçue des offres et des annonceurs (3 items)	Likert en 7 points (de 1 pas du tout d'accord à 7 tout à fait d'accord)

[299] Il s'agit de la classification retenue par l'INSEE. Sont considérés comme Inactifs : les étudiants, les personnes en recherche d'emploi, au foyer, ou retraitées ; les « CSP moins » sont composées des agriculteurs, artisans et ouvriers ; les « CSP moyen » des cadres, commerçants, employés, enseignants et techniciens ; et les « CSP plus » des cadres supérieurs, des chefs d'entreprise, des professions libérales et des ingénieurs.

2.3 La mise en œuvre du design expérimental

L'expérimentation, menée en collaboration avec l'entreprise Millemercis, s'est déroulée en 2 phases. La phase 1 concerne l'expérience proprement dite, menée auprès d'environ 240 000 individus. Elle permettait à notre partenaire de réaliser une opération de requalification sur une partie de sa base de données, raison pour laquelle nous appellerons aussi cette phase : « Requalification ». La phase 2 concerne l'envoi d'un questionnaire post-test à une partie de l'échantillon initial (plus de 40 000 individus) afin d'identifier, d'une part, les raisons pouvant justifier leur comportement de réponse au cours de la phase 1, tout en mesurant, d'autre part, l'effet de variables susceptibles d'expliquer ce comportement (préoccupation pour le respect de la vie privée et variables de contrôle).
Ces deux phases et leurs particularités respectives seront tour à tour présentées.

2.3.1 La phase 1 de l'expérimentation (dite de « requalification »)

En phase 1, un email à été envoyé à un échantillon d'individus membres du programme Elisa (programme relationnel mutualisé par MilleMercis), les invitant à se connecter à un site créé spécifiquement pour l'occasion et sur lequel ils étaient amenés à répondre à un questionnaire comprenant un certain nombre de questions à caractère personnel. L'objectif était alors d'identifier leur comportement de réponse dès qu'ils arrivaient sur le site et de mesurer l'impact des facteurs manipulés sur ce comportement.

Le programme Elisa est une gigantesque base de données qui compte environ 6 millions de personnes. A partir de cette base, un échantillon d'environ 240 000 individus (pour la grande majorité français) a été prélevé, de manière aléatoire, pour participer à cette expérience.
Conformément à la manipulation du premier facteur (familiarité avec l'entreprise sollicitante), la moitié de cet échantillon, soit environ 120 000 personnes, est constituée de clients anciens qui sont inscrits depuis plus d'un an sur Elisa (précisément ils sont entrés entre le 06/12/02 et le 30/07/04). Ce sont donc des individus familiers du programme et de son mode de fonctionnement (i.e. envoi d'emails proposant des promotions et des informations dans différents domaines [mode, tourisme, téléphonie, Automobile]). L'autre moitié (également 120 000 individus) est constituée de clients récemment inscrits (depuis moins de 3 mois, soit entre le 17/06/05 et le 02/09/05), et qui y sont donc moins familiers. Ils ont notamment moins l'habitude de participer aux opérations menées dans le cadre de ce programme.

L'opération a été lancée le 08 septembre 2005 et a duré jusqu'au 28 septembre, soit près de trois semaines. Les emails envoyés ont eux-mêmes été scindés en deux, selon le contexte pouvant amener à solliciter des données personnelles (facteur 2). La moitié de l'échantillon (120 000 individus, 60 000 « anciens » et autant de nouveaux) a ainsi reçu un email invitant à participer à un jeu-concours (permettant de gagner un voyage aux Antilles) tandis que l'autre moitié (120 000 également) se voyait proposer de répondre à une enquête afin d'améliorer les services proposés par le programme Elisa (et de bénéficier par la suite d'offres davantage personnalisées). Ces emails étaient respectivement intitulés : « Explorez la jungle et partez aux Antilles ! » et « Votre opinion est essentielle » (cf. Annexe 12.2). A noter qu'ils offraient tous la possibilité de se désabonner du programme Elisa, si le destinataire le souhaitait.

Ces deux critères (familiarité et contexte) amènent donc à distinguer, dans un premier temps, 4 cibles de l'email : les anciens recevant l'email « jeu-concours », les anciens recevant l'email « enquête », les récents recevant le jeu-concours et les récents recevant l'enquête. Rappelons qu'il existait au total 16 traitement différents, les deux autres facteurs (quantité et ordre) étant manipulés au niveau du questionnaire à remplir, c'est-à-dire une fois que l'individu a cliqué sur le lien fourni et qu'il est arrivé sur le site expérimental.

Les personnes appartenant à ces 4 cibles étaient susceptibles de réagir différemment à l'envoi de cet email. Plusieurs questions se posaient ici : allaient-elles ouvrir l'email ? Si oui, allaient-elles cliquer sur le lien proposé ? Si oui, allaient-elles accepter de fournir les informations demandées, une fois arrivées sur le site expérimental ? Nous détaillons les résultats de la réaction des participants à l'envoi de cet email en Annexe 12.3 (et 12.3 bis).

Nous ne nous intéresserons en fait qu'au comportement de l'ensemble des individus ayant cliqué sur l'email et s'étant ainsi retrouvés sur le site expérimental (soit 20 805 individus sur les 240 000 ayant reçu l'email). Ces individus (qu'ils soient anciens ou récents) devaient alors s'identifier avant de répondre à un questionnaire comprenant des questions d'ordre personnel. A noter qu'afin de mesurer l'impact du contexte sur le comportement de réponse, tout en contrôlant au maximum l'effet de facteurs externes, le questionnaire proposé était identique d'un cas à l'autre (i.e. jeu-concours vs. enquête). Seul variait l'« habillage » du questionnaire (couleurs et images) qui était adapté au contexte affecté (i.e. fond noir et images de la jungle dans le cas du jeu-concours, fond blanc et orange dans le cas de l'enquête).

Le formulaire d'identification présenté sur la page d'accueil du site comprenait 10 données à fournir, toutes obligatoires et identiques d'un contexte à l'autre (cf. Annexe 12.4). Il s'agit, dans l'ordre : de l'email, de la civilité, du nom, du prénom, de la date de naissance, de l'adresse, du code postal, de la ville, du pays et du choix de la fréquence d'envoi des emails en provenance d'Elisa. Une fois ce formulaire validé, les individus tombaient alors sur le questionnaire à compléter comprenant des questions d'ordre sociodémographique, financier ou concernant les loisirs, le logement, l'utilisation d'Internet ou encore les intentions d'achat. A noter que, pour faciliter le remplissage en ligne, on ne trouvait (hormis quelques rares exceptions) qu'une question par écran, laissant en outre la possibilité de ne pas répondre à la question posée (par le biais d'une case à cocher), tout en passant à la question suivante. Ce procédé offre l'avantage d'identifier *a posteriori* les questions les plus sensibles, tout en analysant leur impact sur le comportement de réponse (i.e. poursuite ou abandon notamment). A chaque question, les individus avaient donc 3 possibilités : y répondre et passer à la question suivante, ne pas y répondre (en cochant la case prévue à cet effet) et passer à la question suivante, arrêter le questionnaire à cette question (abandonner). Dans la version jeu-concours, l'individu était amené à participer à une « chasse au trésor ». Chaque question à laquelle il acceptait de répondre lui permettait ainsi de gagner une « statuette », qui lui donnait une chance supplémentaire d'être tiré au sort et de gagner le prix (i.e. un voyage aux Antilles). Dans ce contexte, l'individu était donc incité à répondre à chaque question, ce qui n'était pas forcément le cas pour l'enquête. Vous trouverez en Annexe 12.5 des exemples de pages relatives à la version jeu-concours du questionnaire expérimental.

En réalité, 8 sites expérimentaux, semblables au niveau du design général, mais différents en ce qui concerne les deux autres facteurs étudiés, ont été créés. Ces derniers n'ont en effet été manipulés qu'au niveau du questionnaire auquel les individus ont eu à répondre. Il s'agit et du nombre de questions (Facteur Quantité) et de la place de la question portant sur les revenus (Facteur Ordre) au sein du questionnaire. Un quart des individus a ainsi été soumis à un questionnaire long (42 questions) et « sensible » (question sur les revenus en 6ème position), un autre quart à un questionnaire long et « neutre » (question sur les revenus en 15ème position), un troisième à un questionnaire court (16 questions) et « sensible », le dernier à un questionnaire court et « neutre ». Si l'on ajoute à cela le contexte dans lequel le questionnaire était posé (i.e. jeu-concours vs. enquête), on arrive aux 8 versions différentes. Enfin, si l'on tient compte de la familiarité de l'individu avec le programme Elisa (ancien vs. ancien), on aboutit alors aux 16 traitements (cibles) différents.

Ces 16 cibles ont en réalité été constituées aléatoirement avant même le lancement de l'opération, sur la base de l'échantillon initial de 240 000 individus (finalement de 239 923) soit une moyenne de 14 995 individus par traitement. Le tableau 12.3 présente la répartition des individus selon les cibles. On s'aperçoit que le nombre d'individus par traitement varie de 14 991 à 14 999[300], ce qui représente moins de 0,05% d'écart, soit une différence négligeable (compte tenu de la taille de l'échantillon) et non susceptible de remettre en cause - outre mesure - l'hypothèse d'égalité des échantillons.

Tableau 12.3 Répartition des individus selon les traitements et les facteurs manipulés

Familiarité	Contexte	Quantité (Longueur)	Ordre (Sensibilité)	Traitement	Taille initiale (emails envoyés)
Faible (Récent)	Jeu	Elevée (Long)	Sensible	1	14 997
			Neutre	2	14 998
		Faible (Court)	Sensible	3	14 997
			Neutre	4	14 998
	Enquête	Elevée (Long)	Sensible	5	14 997
			Neutre	6	14 997
		Faible (Court)	Sensible	7	14 997
			Neutre	8	14 999
Elevée (Ancien)	Jeu	Elevée (Long)	Sensible	9	14 995
			Neutre	10	14 992
		Faible (Court)	Sensible	11	14 995
			Neutre	12	14 994
	Enquête	Elevée (Long)	Sensible	13	14 992
			Neutre	14	14 992
		Faible (Court)	Sensible	15	14 992
			Neutre	16	14 991

En résumé, une fois l'échantillon divisé en deux sur la base de la familiarité avec le programme Elisa (Facteur 1) (i.e. récents vs anciens), deux emails différents étaient alors envoyés (i.e. un pour l'enquête, l'autre pour le jeu-concours), chacun à 8 des 16 cibles préalablement définies. Les emails pointaient chacun vers un des huit questionnaires possibles selon le contexte, la longueur et l'ordre des questions posées (correspondant respectivement aux facteurs 2, 3 et 4). A ce stade, rien ne permettait au destinataire d'identifier le nombre de questions auxquelles il serait amené à répondre, ni le type de questions posées. De même, hormis ces différences (design, longueur et ordre), les questions ainsi que l'ordre dans lequel elles étaient posées étaient identiques d'un contexte à l'autre (jeu ou enquête), et ce, quel que soit le niveau de familiarité avec le programme Elisa.

[300] La différence entre l'effectif prévu et l'effectif final est liée à la sélection aléatoire des individus

La figure 12.1 résume la démarche expérimentale suivie au cours de cette première phase.

Figure 12.1 Démarche expérimentale de l'expérience 3 (phase 1)

2.3.2 La phase 2 de l'expérimentation : le post-test

La phase 2 de l'expérience correspond à une phase de post-test, qui consiste à administrer un nouveau questionnaire (dit de post-test) à un sous-échantillon de la population initiale.
Suite à l'envoi du premier email, quatre types de comportements de réponse ont été identifiés, correspondant à autant de populations cibles pour cette seconde phase. Ces quatre «segments» sont présentés ci-dessous selon leur degré d'« implication » dans l'opération, et donc leur attitude plus ou moins bienveillante face à la communication de données personnelles, dans le cadre de ce dispositif expérimental. A noter que seuls les individus disposant d'une adresse email valide (ceux n'ayant pas reçu l'email envoyé en phase 1 ont été supprimés de la base) et ne s'étant pas désinscrits suite à la première phase, ont été inclus dans cette seconde analyse. Un de nos objectifs était de faire en sorte qu'un même nombre d'individus soit interrogé pour chacun de ces segments, de manière à connaître l'opinion de l'ensemble des participants.

Le premier segment (les « inscrits ») concerne les individus les plus réceptifs en phase 1. Il s'agit des personnes qui ont accepté de participer à l'« opération » proposée (en remplissant puis en validant le formulaire d'identification) et qui ont donc répondu, au moins en partie, aux questions posées à travers le questionnaire expérimental administré en phase 1. Le 25 septembre 2005, date de lancement de la phase 2, 11 678 individus (sur les 12 154 inscrits au total au cours de la phase 1[301]) s'étaient inscrits pour participer, à l'enquête pour 6 926 d'entre eux, et au jeu-concours pour les 5 228 restants. Il a donc été décidé d'envoyer le questionnaire post-test à l'ensemble des individus de ce segment, soit aux 11 678 inscrits à cette date (96% de l'ensemble des inscrits au final), puisqu'il s'agissait de la population la plus intéressante à nos yeux. Interroger ces personnes doit en effet permettre de savoir ce qui les a incitées à participer et de distinguer les raisons pratiques (i.e. familiarité avec l'entreprise, intérêt pour l'opération proposée, ...) de raisons plus personnelles (i.e. habitude de participer à des enquêtes et/ou à des jeux concours, faible préoccupation pour le respect de la vie privée, ...).

Le second segment (les « cliqueurs non inscrits » ou « cliqueurs ») est constitué d'individus ayant cliqué sur l'email envoyé en phase 1 mais ne s'étant pas identifiés (i.e. n'ayant pas rempli et/ou validé le formulaire d'identification), une fois arrivés sur le site expérimental. 7 940 personnes (sur les 8 651 identifiées au total en phase 1) correspondent à cette cible. Ce

[301] La différence avec l'effectif final de 12 154 inscrits dans le schéma 12.1 s'explique par la différence de date, la deuxième phase ayant débuté avant la fin de la phase 1, décidée au 28 septembre 2005.

chiffre étant inférieur au nombre d'individus composant le premier segment, il a été décidé de soumettre au post-test ces 7 940 individus, 39% d'entre eux ayant reçu l'email « jeu-concours » (soit 3 110 individus) en phase 1 et 61% l'email « enquête » (soit 4 830 individus). Ces individus constituent un segment dont il est particulièrement important de comprendre la motivation. Ce sont en effet des personnes qui manifestaient *a priori* une attitude favorable face à la communication d'informations puisqu'elles ont cliqué sur l'email[302]. Une fois arrivés sur le formulaire initial, ces individus ont cependant renoncé à participer. Nous savons qu'ils n'ont donc pas pu être influencés dans cette décision, ni par la longueur du questionnaire ni par l'ordre des questions posées, puisqu'ils n'y ont pas été confrontés. De plus, les facteurs manipulés dans l'email (contexte et familiarité) ne les ont pas empêchés d'arriver jusque là et ne peuvent donc pas constituer de véritables facteurs inhibiteurs. Il doit donc exister d'autres raisons à leur comportement, que le post-test doit tenter d'identifier et de quantifier.

Le troisième segment (les « ouvreurs non cliqueurs » ou « ouvreurs ») correspond aux individus ayant ouvert l'email mais n'ayant pas cliqué sur le lien proposé. Au total, plus de 70 000 individus correspondent à cette catégorie. Afin de garder un volume équivalent au premier segment (le second étant inférieur en volume mais ayant été sélectionné en totalité), il a été décidé de prélever aléatoirement un échantillon de 11 000 individus parmi eux, dont la moitié (5 500) ayant reçu l'email « jeu » et l'autre moitié (5 500) l'email « enquête » en phase 1. Finalement, pour des raisons de sélection aléatoire, 10997 individus de cette catégorie ont été soumis au post-test. Dans la mesure où ces individus ont ouvert le premier email (en phase 1), on peut supposer qu'ils avaient suffisamment confiance dans l'expéditeur pour l'ouvrir et pour lire, au moins partiellement, son contenu. De prime abord, le contexte de sollicitation (identifiable à travers la ligne sujet) ne les a pas non plus empêchés d'ouvrir l'email, fut-ce par curiosité. Cependant, après avoir lu de quoi il s'agissait, ils ont alors renoncé à poursuivre et se sont abstenus de cliquer sur le lien proposé. Les raisons ayant conduit à cette décision peuvent être multiples et sont susceptibles d'être situées en dehors du cadre étudié ici (à savoir ce qui amène un individu à fournir ou non des données personnelles). De plus, nous avons indiqué, dans le chapitre 7 notamment, que nous nous intéressions exclusivement à la réaction du consommateur une fois celui-ci soumis à un formulaire d'identification. Même si nous envoyons le questionnaire post-test aux 11 000 personnes appartenant à ce segment, nous ne nous intéresserons donc que très partiellement à leurs réponses dans le cadre de la validation de nos hypothèses de recherche.

[302] On peut toutefois penser que certaines d'entre elles ont également cliqué par « curiosité ».

Le quatrième et dernier segment (les « non ouvreurs ») concerne les individus ayant reçu l'email en phase 1 mais ne l'ayant pas ouvert. Il s'agit donc de personnes qui n'ont pas voulu connaître en détail l'objectif de l'opération proposée, et qui ne souhaitaient pas y participer. Elles se donc décidées à la simple vue de l'expéditeur et de l'intitulé de l'email. Interroger ces personnes pourrait donc permettre d'identifier les raisons de leur réticence et de séparer celles liées aux facteurs manipulés (familiarité avec l'expéditeur et contexte de la sollicitation) de raisons plus « personnelles », telles que l'habitude de ne pas répondre à ce type d'enquête ou une forte préoccupation envers le respect de la vie privée. Plus de 130 000 personnes appartiennent à cette catégorie. De la même façon qu'avec le segment 3, un échantillon de 11 000 individus a été prélevé au sein de ce groupe dont la moitié ayant reçu l'email « jeu » et l'autre moitié l'email « enquête » au cours de la phase 1. Comme précédemment également, puisque nous nous intéressons aux réactions des personnes ayant été soumises au formulaire, nous ne tiendrons pas compte de leurs réponses, pour tester notre modèle.

Pour ce post-test, un second email a été envoyé comportant lui aussi un lien conduisant sur un site permettant d'administrer en ligne le questionnaire post-test, dans sa version « jeu » ou « enquête »[303], selon la cible initiale définie en phase 1. Cet email était intitulé respectivement « Votre opinion sur les jeux sur Internet » et « Votre opinion sur les enquêtes en ligne ».
80% des emails envoyés en phase 2 (soit 33 263) mettaient en avant le partenariat avec l'Université de Paris Dauphine, notamment grâce à la présence du logo au sein de l'email (emails dits « brandés »). Les 20% restant, soit 8 352 emails ont été envoyés sans qu'aucune mention ne soit faite du partenariat. L'ensemble de ces emails (dont vous trouverez un exemple en Annexe 12.6) ont été envoyés le 05 octobre, soit plus d'un mois après l'envoi des premiers emails (correspondant à la phase 1) et presque une semaine après la fin de l'opération (le jeu se terminait en effet le 28 septembre).

La figure 12.2 en page suivante résume la démarche suivie au cours de cette seconde phase.

[303] En réalité, seul l'intitulé de la question 4 du questionnaire post-test était différent d'un contexte à l'autre. Nous souhaitions en effet poser une question précise plutôt que d'utiliser des termes « généraux » qui risquaient d'être mal compris et/ou interprétés et donc de conduire à des réponses inappropriées.

Figure 12.2 Démarche expérimentale de l'expérience 3 (phase 2)

Nous abordons maintenant la section 3 de ce chapitre qui présente le modèle à tester (3.1) ainsi que les hypothèses de recherche que nous seront amenés à tester par la suite (3.2).

3 Présentation du modèle à tester et des hypothèses de recherche

3.1 Présentation du modèle à tester

Rappelons que l'objectif majeur de cette troisième expérience est, d'une part, de mesurer le comportement de réponse des individus face à une sollicitation de leurs données personnelles, et ce, de manière réelle (plutôt qu'en déclaratif) et, d'autre part, d'évaluer l'impact direct de l'exposition à la sollicitation (à travers différents facteurs manipulés) sur ce comportement. L'impact sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données sera également testé.

Pour mémoire, les facteurs sont identifiés par un numéro correspondant à l'ordre dans lequel ils sont manipulés au sein du dispositif expérimental. Ainsi, la familiarité avec l'entreprise ayant sollicité les données est numérotée « 1.2 », la quantité de données demandées « 1.3 » et le contexte de sollicitation « 1.4 ».
Les différentes variables dépendantes (ou médiatrices) du modèle sont, pour leur part, identifiées par une lettre (par exemple, « h » pour l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données).
Enfin, la variable « Réponse » (REP) notée « i » dans le modèle général (cf. figures 7.6 et 8.2), sera décomposée ici en plusieurs variables, notées respectivement « i », « j » et « k ». Nous reviendrons sur chacune d'elles dans les paragraphes qui suivent.

Le modèle à tester dans le cadre de cette troisième et dernière expérience[304] se présente donc de la manière suivante (cf. figure 12.3 page suivante) :

[304] Comme nous l'avons indiqué et justifié précédemment, nous ne tiendrons pas compte du quatrième facteur manipulé (i.e. l'ordre des questions) dans le cadre de notre modèle à tester et de nos hypothèses de recherche. Nous nous réservons cependant le droit d'aborder cette question dans la discussion finale de ce document.

Figure 12.3 Modèle à tester dans le cadre de l'Expérience 3

Variables de contrôle

- *Variables socio-démographiques (âge, sexe, CSP, niveau d'éducation)*
- *Expérience (EXP) et Utilisation d'Internet (UTIL)*
- *Comportement habituel de réponse (COMP HAB)*
- *Qualité perçue de la relation (du programme relationnel) (QUALITE)*

On distingue donc, au sein de ce modèle, 4 variables dépendantes (à expliquer), 3 variables indépendantes (correspondant aux facteurs manipulés) et 4 types de variables de contrôle. Dans le paragraphe suivant, nous revenons plus particulièrement sur les variables dépendantes, les 2 autres catégories ayant été présentées en détail dans la section précédente.

Il existe en réalité, dans ce modèle, deux types de variables dépendantes :

1) trois variables de « réponse », correspondant à la variable dépendante majeure du modèle général à tester (notée REP, cf. figure 8.2). Il s'agit de : la Communication de Données Personnelles (ou inscription[305]) (notée i), de l'Abandon (j) et du Niveau de Dévoilement (k), correspondant à la réponse à une question sensible. Ces variables correspondent respectivement au fait d'avoir accepté (ou non) de communiquer les données demandées (en s'inscrivant à l'opération proposée au cours de la phase 1), au fait d'avoir été (ou non) jusqu'au bout du questionnaire à compléter et au fait d'avoir accepté (ou non) de fournir une information particulièrement sensible, comme le numéro de téléphone mobile[306]. Concernant la réponse à une question sensible (numéro de téléphone mobile), nous avons volontairement choisi une question posée en début de questionnaire (question 14) de manière à maximiser les chances que les participants soient encore présents à ce moment là. Une question posée après la seizième nous aurait en effet obligés à « éliminer » les individus affectés au questionnaire court (limité aux 16 premières questions) et n'aurait alors pas permis de tester l'effet de la quantité. Nous reconnaissons que ce choix a l'inconvénient potentiel de limiter cet effet, puisque tous les individus ont répondu aux mêmes questions jusque là. Rappelons cependant que les répondants pouvaient visualiser l'avancement du questionnaire grâce à une « barre de progression » et que les individus affectés au questionnaire court pouvaient considérer cette question comme l'une des dernières alors que ceux affectés au questionnaire long visualisaient être, à ce moment là, au tiers du questionnaire (14/42).

2) une variable dépendante « finale » : l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données (IPR), conséquence des variables de réponse précédentes, mais présentée comme antérieure dans le modèle conceptuel, puisqu'il s'agit d'une intention comportementale.

---

[305] Cette variable correspond au fait d'avoir (ou non) validé le formulaire d'identification donc de s'être inscrit.

[306] Le choix du numéro de téléphone mobile tient au fait qu'il est souvent considéré comme l'information la plus sensible par les individus. Les résultats de l'étude menée dans le cadre de cette expérience confirmeront d'ailleurs ce constat (taux d'abstention moyen de 71% contre 6% pour les revenus et 11% pour l'adresse email).

Les trois variables de réponse sont binaires (de type oui/non), tandis que l'intention de poursuivre la relation est mesurée sur une échelle de Likert en 7 points. Nous reviendrons sur la qualité psychométrique de ces instruments de mesure ultérieurement dans ce chapitre.

Ce modèle à tester comprend plusieurs hypothèses que nous allons à présent formuler.

3.2 Formulation des hypothèses de recherche

Les hypothèses de recherche relatives à cette troisième et dernière expérience peuvent se scinder en trois blocs selon le type de variables indépendantes et dépendantes concernées. Nous verrons tour à tour les hypothèses relatives à :

- l'influence de l'exposition à la sollicitation sur les variables de réponse (H1.2, H1.3 et H1.4)
- l'influence de la préoccupation pour le RVP sur les variables de réponse (H3 i, j et k)
- l'influence sur l'intention de poursuivre la relation (H1.2 h, H1.4 h et H6 i et j)

3.2.1 L'influence de l'exposition à la sollicitation sur la réponse

Dans le chapitre présentant le cadre conceptuel (i.e. chapitre 7), nous avions fait l'« hypothèse » de l'existence d'un lien direct entre l'exposition à la sollicitation (correspondant aux facteurs situationnels manipulés) et le comportement de réponse (identifié par la lettre i), formulée de la manière suivante :

Proposition 1 (i) : *Les facteurs situationnels influencent de manière directe la réponse du consommateur face à une sollicitation de ses données personnelles (i).*

Nous déclinons ici cette proposition en différentes hypothèses relatives à l'effet des facteurs situationnels manipulés dans cette expérience (i.e. Familiarité, Contexte et Quantité, notés respectivement 1.2 ; 1.4 et 1.3) sur les variables mesurant le comportement de réponse, et présentées dans le paragraphe précédent, à savoir l'inscription (notée i), l'abandon (noté j) et le niveau de dévoilement (acceptation de fournir son numéro de téléphone mobile) (notée k).

L'influence de la familiarité avec l'entreprise sollicitant les données

Dans le domaine interpersonnel, il a été montré que les individus sont plus enclins à répondre favorablement aux demandes émanant de gens qu'ils apprécient. Or, les relations passées figurent parmi les facteurs qui augmentent le fait d'apprécier l'autre (Frankfurt 1965).
En matière de sollicitation de données personnelles, dans un cadre commercial, Culnan et Armstrong (1999) soulignent que l'expérience passée avec la firme permet de distinguer les consommateurs qui acceptent de fournir des informations et ceux qui ne le souhaitent pas. De même, Zhang, Wang et Chen (2001) montrent que le fait de ne pas être client de l'entreprise sollicitante serait un frein dans l'acceptation de divulguer des données personnelles. On peut en effet supposer que les clients peu familiers auront une réticence à se dévoiler car les facteurs bâtisseurs de confiance sont encore peu établis (Luo 2002). Reichheld et Schefter (2000) notent ainsi que lorsque les consommateurs ont confiance dans un e-vendeur, ils sont plus enclins à lui fournir des informations. Le fait que la relation soit établie amènerait en effet le consommateur à se sentir moins vulnérable et l'encouragerait alors à fournir les renseignements demandés. Enfin, en matière de réponse aux enquêtes, Comley (2000) souligne qu'un taux de réponse élevé intervient lorsque les personnes sollicitées ont une relation forte avec l'annonceur. Tezinde et al. (2002) confirment que les relations passées avec l'entreprise influencent le taux de réponse.

L'ensemble de ces résultats conforte à penser que, dans le cadre d'une sollicitation de données personnelles, la familiarité avec l'entreprise sollicitant les données (notée 1.2) a une influence non négligeable sur les variables de réponse identifiées dans notre modèle, à savoir : la communication de données personnelles (i), l'abandon (j) et le niveau de dévoilement (réponse à une question sensible). Bien que situés à des niveaux d'analyses différents, ces comportements découlent en effet tous d'un même processus : le fait d'accepter (ou non) de fournir des informations personnelles à une entreprise.

Nous formulons donc cette triple hypothèse de la manière suivante :

Hypothèse 1.2(i,j,k) : *plus (vs moins) l'individu est familier avec l'entreprise qui le sollicite :*

- plus (vs. moins) la probabilité qu'il accepte de communiquer des données personnelles est élevée (vs. faible) (i) ;

- plus la probabilité qu'il abandonne avant la fin est faible (vs. élevée) (j) ;

- plus la probabilité qu'il réponde à une question sensible est élevée (vs. faible) (k).

L'influence du contexte de sollicitation

Outre la familiarité avec l'entreprise sollicitante, la raison de la divulgation et donc le contexte dans lequel les informations sont sollicitées ainsi que les bénéfices pouvant être obtenus en échange (par exemple : accès au site vs. achat en ligne) affectent aussi la décision du consommateur. Andrade, Kaltcheva et Weitz (2002) montrent ainsi que 7% des répondants évoquent ce facteur comme étant le principal élément de leur prise de décision.

Phelps, Nowak et Ferrell (2000) assurent que l'existence d'un intérêt personnel à répondre est un antécédent important de la volonté du consommateur de divulguer des données à une entreprise. Dans l'expérience 2, nous avons testé différentes formes d'intérêts personnels, pouvant prendre la forme d'une récompense (jeu-concours), d'une offre personnalisée (espace personnalisé sur le site) en passant par le fait d'obtenir des informations sur un sujet qui intéresse le répondant (newsletter).

D'après Hoffman, Nowak et Peralta (1999), les consommateurs seraient moins intéressés par le fait de donner des informations en échange d'avantages utilitaires (notamment financiers) que par une forme d'échange basée sur un contrat social[307] implicite, fondé sur la confiance. Dès lors, leur proposer de participer à une enquête visant à mieux les connaître et à leur faire des offres adaptées pourrait se révéler plus incitateur que la participation à un jeu-concours. De plus, une enquête portant sur les goûts et préférences génère souvent une forte implication et se traduit donc par un taux de participation élevé.

Pour Dinev et Hart (2002), l'existence d'un intérêt personnel à répondre serait une importante monnaie d'échange pour réduire les préoccupations relatives au respect de la vie privée. Dès lors, le comportement de réponse face à une question sensible pourrait différer selon le contexte dans lequel elle est posée. L'individu se montrerait ainsi plus coopératif si le contexte répond à un intérêt personnel supérieur, comme c'est probablement le cas dans le cadre de l'enquête (comparativement à celui occasionné par un jeu-concours).

Enfin, l'existence d'un intérêt personnel serait également susceptible de réduire le risque d'ennui et donc d'abandon. Frick, Baechtinger et Reips (1999) montrent ainsi que la chance de gagner un prix (en participant à une loterie) diminue le taux d'abandon tandis que Tuten, Bosnjak et Bandilla (2000) trouvent, à l'inverse, que la part de non réponse est plus élevée quand on offre de gagner un prix plutôt que quand des motivations altruistes sont offertes (exemple : la contribution à la recherche). De plus, Bickart et Schmittlein (1999) pensent que l'intérêt porté au contexte de la sollicitation pourrait réduire la sensation de fatigue ressentie

[307] Pour rappel, la notion de contrat social a été développée au chapitre 4 de ce document (dans la section 3).

pendant le remplissage et motiver le répondant à continuer. Le taux d'abandon serait donc plus faible pour les contextes liés à un intérêt personnel élevé.

L'ensemble de ces considérations nous amène donc à formuler les hypothèses suivantes :

Hypothèse 1.4 (i, j, k) : *la probabilité que l'individu accepte de communiquer des données personnelles (i), aille jusqu'au bout (vs. abandonne) (j) et réponde à une question sensible (k) diffère selon le contexte dans lequel s'opère la sollicitation et les avantages qu'il peut obtenir en échange. Elle serait plus forte pour l'enquête comparativement au jeu-concours.*

L'influence de la quantité d'informations demandées

Nous justifions à présent l'effet supposé de la quantité de données demandées sur la probabilité d'abandonner d'une part et sur la probabilité de répondre à une question sensible (consistant à fournir son numéro de téléphone mobile) d'autre part. Nous ne formulerons pas d'hypothèse relative à l'effet de ce facteur sur la probabilité de s'inscrire, car elle ne s'applique pas à l'expérience menée. Au moment de participer à l'opération proposée, les répondants n'avaient en effet aucune idée du nombre de questions qui allaient leur être posées par la suite. Cette quantité ne pouvait alors pas influencer leur décision et nous n'avions donc aucune chance (hormis celle liée au hasard) de déceler un effet de la quantité sur la probabilité de communiquer des données personnelles (c'est-à-dire de s'inscrire)[308].

Un point de vue réaliste recommande de penser qu'un questionnaire trop long entraînera un effet de démission de la part du répondant. Or, les travaux empiriques menés sur le sujet ne confirment que partiellement cette hypothèse. La majorité des études concernant l'effet de la longueur du questionnaire sur le taux de réponse donnent en effet des résultats contradictoires. Si Kanuk et Berenson (1975) ainsi que Heberlein et Baumgartner (1978) montrent que le taux de réponse est plus faible pour un questionnaire long, d'autres études ne confirment pas cette hypothèse (Roscoe, Lang et Sheth 1975) et certaines vont même jusqu'à la contredire (Berdie 1973, Champion et Sear 1969). Une explication possible à ces divergences est apportée par Friedman et Goldstein (1975) qui suggèrent que la longueur du questionnaire est perçue par le répondant comme étant multidimensionnelle avec plusieurs facteurs susceptibles d'agir ensemble pour influencer le taux de réponse tels que notamment : le nombre de questions, le nombre de pages et la taille des feuilles (dans le cas d'un questionnaire papier).

[308] Nous avons cependant vérifié cette hypothèse d'absence d'influence du facteur quantité sur la probabilité de s'inscrire, qui s'avère confirmée par les résultats (chi-deux = 0,012, p = 0,931).

S'agissant des questionnaires administrés en ligne, Manfreda et al. (2002) soulignent que le risque d'abandon est potentiellement plus fort quand le questionnaire figure sur plusieurs pages et, surtout, dans les cas extrêmes où il y a une question par page. Et ce pour plusieurs raisons, dont les principales sont : le fait que l'on mette plus de temps à répondre et le fait que l'effort demandé au répondant soit plus important (donner la réponse, cliquer sur le bouton suivant, répondre à la question qui suit, etc …) (Dillman et Bowker 2001).

Dans l'ensemble, il semble donc que l'hypothèse d'un effet négatif de la longueur soit globalement validée pour les questionnaires auto-administrés alors que, dans les méthodes assistées, la longueur produirait peu d'effet significatif. De plus, il apparaît que la coopération serait corrélée avec le niveau d'effort requis pour répondre (Lysaker 1989, Manfreda et al. 2002). Enfin, Dillman, Sinclair et Clark (1993) indiquent que, dans la majorité des cas, le taux d'omission aux questions difficiles ou embarrassantes (sensibles) est particulièrement élevé. Dès lors, si en plus d'une longueur importante apparaît une question sensible, la probabilité que le répondant accepte d'y répondre sera plus faible que lorsque le questionnaire est court. D'où l'hypothèse suivante :

Hypothèse 1.3 (j, k) : *Plus la quantité d'informations demandées est élevée (vs. faible), plus la probabilité que l'individu abandonne avant la fin est élevée (vs. faible) (j) et plus la probabilité qu'il réponde à une question sensible est faible (vs. élevée) (k).*

3.2.2 L'influence de la préoccupation pour le RVP sur la réponse

Dans le chapitre présentant le cadre conceptuel (i.e. Chapitre 7), nous avions développé une proposition (notée P3i) indiquant l'influence de la préoccupation pour le RVP sur le comportement de réponse (noté i). Rappelons que cette proposition (justifiée au chapitre 7) était formulée de la manière suivante :

Proposition 3 i : *Plus la préoccupation pour le RVP est élevée (vs. faible), moins (vs. plus) l'individu acceptera de communiquer ses données personnelles (i).*

Comme nous mesurons, dans le cas présent, trois types de comportements de réponse (l'inscription, notée i ; l'abandon, noté j et le niveau de dévoilement, noté k), nous faisons l'hypothèse que la préoccupation pour le RVP influencera tout autant ces trois variables. En effet, la littérature reconnaît que la préoccupation est susceptible d'influencer certains comportements relatifs aux pratiques de collecte de données personnelles. En particulier, Sheehan et Hoy (1999) constatent que plus la préoccupation est élevée, moins les individus

acceptent de s'enregistrer sur les sites. Dans le domaine des enquêtes par sondage, il a aussi été montré que le comportement de réponse pouvait être affecté par le rang que les problèmes de RVP pouvaient tenir dans le système de valeur des individus (Goyder et McKenzie, 1985). Cette hypothèse sera confirmée plus tard par Singer, Mathiowetz et Couper (1993) qui trouvent un lien entre la préoccupation pour le RVP et le taux de réponse, tout en spécifiant qu'il ne représente qu'une très faible part (3% environ) de la variance. Comme il n'y a pas de raison pour que la préoccupation pour le RVP n'affecte pas d'autres variables que le taux de réponse, on peut supposer que celle-ci aura également un effet sur le comportement de réponse au sein du questionnaire (taux d'abandon, réponse aux questions sensibles, etc …).
Ceci nous amène à formuler les hypothèses suivantes :

Hypothèse 3 (i, j, k) : *Plus la préoccupation pour le respect de la vie privée est élevée (vs. faible), plus la probabilité que l'individu accepte de communiquer des données personnelles est faible (vs. élevée) (i) ; plus la probabilité qu'il abandonne avant la fin est élevée (vs. faible) (j) et plus la probabilité qu'il refuse de répondre aux questions sensibles est élevée (vs. faible) (k).*

3.2.3 L'influence sur l'intention de poursuivre la relation

Dans le chapitre présentant le cadre conceptuel (i.e. Chapitre 7), nous avions développé une proposition (P6) concernant l'effet du comportement de réponse sur l'intention de poursuivre la relation avec l'entreprise sollicitante. Rappelons que cette proposition (justifiée au chapitre 7) était formulée de la manière suivante :

Proposition 6 : *Plus (vs. moins) l'individu accepte de communiquer les données demandées, plus (vs. moins) il a l'intention de poursuivre la relation avec l'entreprise qui l'a sollicité.*

Comme nous mesurons, dans le cas présent, deux types de comportements de réponse majeurs (l'inscription, notée i et l'abandon, noté j), nous déclinons cette proposition en 2 sous-hypothèses correspondant à ces deux variables dépendantes[309], soit :

Hypothèse 6 (i, j) : *L'individu ayant accepté de communiquer des données personnelles (vs. n'ayant pas accepté) (i) et n'ayant pas abandonné avant la fin (vs. ayant abandonné) (j) a une plus forte (vs. faible) probabilité de poursuivre la relation avec l'entreprise qui l'a sollicité.*

309 Nous ne pensons pas qu'il existe un lien entre le niveau de dévoilement (réponse aux questions sensibles) et l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données, c'est pourquoi nous ne formulons pas d'hypothèse à ce sujet.

Outre le comportement adopté par l'individu face à la collecte de ses données personnelles, nous pensons que la sollicitation elle-même est également susceptible d'influencer l'intention de poursuivre la relation. Nous avions, dans le chapitre concernant le cadre conceptuel, formulé une proposition relative à cet effet, de la manière suivante :

Proposition 1 (h) : *Les facteurs situationnels influencent de manière directe l'intention du consommateur de poursuivre la relation avec l'entreprise ayant sollicité les données (h).*

Nous estimons que l'expérience passée avec l'entreprise sollicitante ainsi que le contexte dans lequel la sollicitation s'est opérée (i.e. Jeu-concours vs. Enquête, dans le cadre de cette expérience) sont susceptibles d'influencer l'intention de poursuivre la relation.

Les théories de l'engagement et de l'échange social (développées au chapitre 7) offrent une justification quant à l'influence de l'expérience passée avec l'entreprise (la familiarité) sur l'intention de continuer à commercer avec celle-ci dans le futur (engagement).
Ainsi, la théorie de l'échange social explique le lien entre les échanges passés et l'engagement par le fait que de fréquents échanges réduisent l'incertitude : des acteurs qui ont de nombreux échanges entre eux trouvent le comportement de l'autre plus prédictible et estiment qu'ils partagent une vision commune de l'échange (Cook et Emerson 1978). De leur côté, Lawler et Yoon (1996) et Homans (1961) montrent que les échanges répétitifs ont d'importantes composantes affectives. Ils développent ainsi l'idée qu'il existe un processus qui lie la fréquence des échanges, les émotions positives et la perception de la relation comme un objet d'attachement et d'engagement. Il sera d'ailleurs démontré que la répétition des interactions avec le partenaire (engagement comportemental) peut conduire au désir de continuer la relation avec lui (dimension temporelle), mais probablement de façon moins forte et moins durable que celui provenant de l'engagement affectif (liens émotionnels issus de la relation) (Gutierrez, Cillan et Izquierdo 2004). Enfin, plusieurs études déduisent que l'engagement est positivement lié à la dépendance mutuelle entre les parties, à l'expérience avec le partenaire et plus particulièrement à la satisfaction vis-à-vis des expériences passées (Ganesan 1994, Morgan et Hunt 1994, Moorman Desphandé et Zaltman 1993). Ceci nous permet d'établir l'hypothèse selon laquelle, dans le cadre d'une sollicitation de données personnelles, la familiarité avec l'entreprise sollicitant les données a un impact sur l'intention de l'individu sollicité de poursuivre la relation avec cette entreprise, que nous formulons ainsi :

Hypothèse 1.2 h : *Plus (vs. moins) l'individu est familier avec l'entreprise qui le sollicite, plus la probabilité qu'il poursuivre la relation avec elle est élevée (vs. faible).*

De même, nous supposons l'existence d'un effet du contexte de sollicitation sur l'intention de poursuivre la relation avec l'entreprise. Ainsi, on peut penser qu'une personne n'appréciant pas de se voir demander de participer à un jeu-concours pourrait se poser la question de l'intérêt à continuer à entretenir une relation avec une entreprise qui ne respecte pas son intimité. Puisque nous estimons que l'intérêt porté à l'enquête est plus élevé que celui porté au jeu, nous pensons que l'incidence sur la poursuite de la relation ira dans le même sens. D'où l'hypothèse suivante :

Hypothèse 1.4 h : *L'intention de poursuivre la relation avec l'entreprise diffère selon le contexte dans lequel s'est opérée la sollicitation et les avantages que l'individu a pu obtenir en échange. Elle serait plus élevée pour l'enquête, comparativement au jeu-concours.*

En résumé, les hypothèses relatives à cette expérience sont indiquées dans le tableau suivant :

Tableau 12.4 Récapitulatif des hypothèses à tester dans l'expérience 3

H 1.2 (i, j, k) (17,18,19)	Plus (vs. moins) l'individu est familier avec l'entreprise qui le sollicite : - plus (vs. moins) la probabilité qu'il accepte de communiquer des données personnelles est élevée (vs. faible) (i) ; - plus la probabilité qu'il abandonne avant la fin est faible (vs. élevée) (j) ; - plus la probabilité qu'il réponde à une question sensible est élevée (vs. faible) (k).
H 1.4 (i, j, k) (20,21,22)	La probabilité que l'individu accepte de communiquer des données personnelles (i), aille jusqu'au bout (vs. abandonne) (j) et réponde à une question sensible (k) diffère selon le contexte dans lequel s'opère la sollicitation et les avantages qu'il peut obtenir en échange. Elle serait plus forte pour l'enquête comparativement au jeu-concours.
H 1.3 (j, k) (23,24)	Plus la quantité d'informations demandées est élevée (vs. faible), plus la probabilité que l'individu abandonne avant la fin est élevée (vs. faible) (j) et plus la probabilité qu'il réponde à une question sensible est faible (vs. élevée) (k).
H 3 (i, j, k) (25,26,27)	Plus la préoccupation pour le respect de la vie privée est élevée (vs. faible), plus la probabilité que l'individu accepte de communiquer des données personnelles est faible (vs. élevée) (i) ; plus la probabilité qu'il abandonne avant la fin est élevée (vs. faible) (j) et plus la probabilité qu'il refuse de répondre aux questions sensibles est élevée (vs faible) (k).
H 6 (i, j) (28,29)	L'individu ayant accepté de communiquer des données personnelles (vs. n'ayant pas accepté) (i) et n'ayant pas abandonné avant la fin (vs. ayant abandonné) (j) a une plus forte (vs. faible) probabilité de poursuivre la relation avec l'entreprise qui l'a sollicité.
H 1.2 (h) (30)	Plus (vs. moins) l'individu est familier avec l'entreprise qui le sollicite, plus la probabilité qu'il poursuive la relation avec elle est élevée (vs. faible).
H 1.4 (h) (31)	L'intention de poursuivre la relation avec l'entreprise diffère selon le contexte dans lequel s'est opérée la sollicitation et les avantages que l'individu a pu obtenir en échange. Elle serait plus élevée pour l'enquête, comparativement au jeu-concours.

Ayant formulé nos hypothèses de recherche, nous abordons, à présent, le recueil des données et les instruments de collecte élaborés pour ce faire.

4 Le recueil des données et les instruments de collecte

Nous présenterons brièvement l'instrument de collecte élaboré pour chacune des 2 phases de l'expérimentation, à savoir l'expérience en elle-même (ou « requalification ») (phase 1) (4.1) et le post-test (phase 2) (4.2), avant d'évoquer les données à exploiter (4.3).

4.1 L'instrument de collecte de la phase 1

Le questionnaire administré en phase 1, comprend en réalité 4 versions différentes (8 si l'on considère le design correspondant à la manipulation du facteur « Contexte » [jeu-concours vs. Enquête]). Ces quatre versions correspondent respectivement aux questionnaires Court-Sensible, Court-Neutre, Long-Sensible et Long-Neutre (cf. section précédente pour les explications relatives à ces « intitulés »), relatifs à la manipulation des facteurs « Quantité » et « Ordre ». Vous trouverez en Annexe 12.7 un exemple de questionnaire concernant cette première phase, dans sa version Long-Sensible. A noter qu'une barre de progression permettait au répondant de visualiser l'avancée du questionnaire et donc de pouvoir déterminer, de manière approximative toutefois, le nombre de questions restant. Ainsi, à la question 15, les répondants au questionnaire court pouvaient voir qu'ils étaient à la fin alors que ceux affectés au questionnaire long voyaient qu'ils étaient à peine à la moitié.

Ces questionnaires comprennent un certain nombre de questions (16 dans le cas du questionnaire court, 42 dans le cadre du questionnaire long), d'ordre personnel (d'autant que les personnes se sont identifiées au préalable) et de nature différente. On trouvait ainsi des questions d'identification (I), d'ordre sociodémographique (S), relatives à l'utilisation des Médias (M), d'ordre financier (F), relatives au logement (L), aux possessions et intentions d'achat (P), aux centres d'intérêt (C), et sur des sujets divers (D). L'ensemble de ces questions étaient réparties uniformément au sein du questionnaire. Le tableau 12.5 ci-après résume le type de données demandées (et leur nombre) en fonction de la longueur du questionnaire.
Nous rappelons également que la manipulation du facteur « Ordre » consistait à inverser l'ordre des questions 6 et 15. La version « Sensible » comprenait ainsi la question sur les revenus en sixième position et la question sur les intentions d'achat en matière de véhicules

(voitures et deux-roues) en quinzième place et inversement pour la version « Neutre ». L'impact de ce facteur sur les variables de réponse ne sera toutefois pas étudié, pour les raisons déjà évoquées dans le paragraphe 2.1.

Tableau 12.5 Types de données selon la longueur du questionnaire (Facteur Quantité)

Types de données	Questionnaire court	Questionnaire long
I : identification	1	1 + 1* = 2
S : sociodémographique	4	4 + 1* = 5
M : utilisation Médias/ Internet	4	4 + 4* = 8
F : financier	2	2 + 1* = 3
L : logement	1	1 + 2* = 3
P : possessions et intentions d'achat	1	1 + 6* = 7
C : centres d'intérêt, loisirs et préférences	1	1 + 7* = 8
D : divers	2	2 + 5* = 7
TOTAL	16	42

* spécifiques au questionnaire long (c'est-à-dire relatives aux questions 17 à 42)

4.2 L'instrument de collecte de la phase 2

Dans la mesure où les deux versions du questionnaire post-test ne diffèrent que sur l'intitulé d'une question (Q4), nous présenterons ci-dessous le cas général. La version « jeu-concours » de ce questionnaire est consultable en Annexe 12.8.

Ce questionnaire post-test, administré en ligne, comprend 9 écrans différents, pour un total de 29 questions, soit une moyenne de 3 à 4 questions par écran. La réponse aux questions était obligatoire, hormis la dernière (PTQ29[310]), de type ouverte, portant sur d'éventuelles remarques ou propositions à formuler.

Les deux premières questions (page 1) visaient à savoir si l'individu se rappelait avoir reçu un email le sollicitant pour participer à une opération organisée dans le cadre du programme Elisa (enquête ou jeu-concours) et se souvenait de la manière dont il avait réagi. Tous les individus ayant accepté de participer à ce post-test ont eu à répondre à ces deux questions.

[310] Pour différencier les questions du questionnaire expérimental (dit de requalification) et celles du post-test, nous notons, par convention, P1 à P16 (ou P42) celles relatives au premier et PTQ1 à PTQ29, celles du second.

La page suivante (page 2) comportait des questions visant à identifier les raisons expliquant la réaction face à cet email (absence d'ouverture ou absence de clic). Seuls les « non ouvreurs » et les « ouvreurs non cliqueurs » de la phase 1 ont donc eu à répondre à ces questions.
La page 3 concernait plus particulièrement les raisons expliquant le comportement des cliqueurs ne s'étant pas inscrits ou des inscrits ayant abandonné avant la fin. Seuls les « cliqueurs » et les « inscrits » de la phase 1 ont donc eu à répondre à ces questions.
Selon le segment auquel ils appartenaient (parmi les quatre présentés précédemment), les individus ont donc répondu soit aux questions de la page 2 soit à celles de la page 3, ou bien encore, ont directement été affectés à la page 4 (cas des « inscrits » n'ayant pas abandonné). A noter que les individus ne se souvenant pas de la manière dont ils avaient réagi face à l'email envoyé en phase 1 ont également été directement dirigés sur la page 4.
En page 4 se trouvaient les questions visant à mesurer le comportement de réponse habituel face à une sollicitation de données personnelles (facteurs susceptibles de déclencher un comportement de méfiance ou d'abandon).
La page 5 contenait des questions visant à évaluer les facteurs susceptibles d'expliquer la participation à un jeu-concours sur Internet et à mesurer la préoccupation pour le RVP.
La page 6 visait à mesurer comment ces personnes perçoivent le programme Elisa, la manière dont elles ont l'habitude de réagir face aux emails reçus dans le cadre de ce programme et dont elles imaginent la relation future (intention de rester ou non inscrites).
Les pages 7, 8 et 9 concernaient respectivement l'attitude et le comportement habituel de l'individu face aux sollicitations commerciales par email, l'utilisation et la consultation de la messagerie électronique ainsi que les usages et l'expérience en matière d'Internet.

4.3 Les données à exploiter

A l'issue de ces deux phases, nous disposions donc de deux jeux (ou fichiers) de données.
Le premier fichier (dit de « requalification ») comprend l'ensemble des réponses au questionnaire issu de la phase expérimentale (précédées du formulaire d'identification), pour tous les internautes s'étant retrouvés sur la page d'accueil du site et s'étant ou non enregistrés. Il s'agit en particulier des « inscrits » (personnes qui ont rempli et validé le formulaire d'identification et qui ont ensuite répondu ou non au questionnaire qui suivait[311], soit 12 154 individus) et des « cliqueurs non inscrits » (qui ont quitté le site avant même de s'identifier et qui n'ont donc répondu à aucune des questions qui suivaient, soit 8 651 individus).

[311] Plus de 3% des individus inscrits, donc préalablement identifiés, n'ont ensuite répondu à aucune question.

Ce fichier comprend donc 20 805 lignes (correspondant aux 2 types d'individus identifiés ci-dessus) et une cinquantaine de colonnes correspondant aux variables suivantes :

- les 4 facteurs manipulés : codés 1 ou 2
- des variables d'ordre sociodémographique (âge, sexe, etc …) : correspondant aux réponses données au formulaire d'identification (à condition qu'il ait été rempli)
- les réponses à chacune des questions posées dans le questionnaire (de P1 à P16 ou P42 selon la longueur) : codées 1 quand l'individu y a répondu, 0 quand il est passé à la question suivante sans y répondre et « manquant » s il a abandonné auparavant.

Il faut souligner que pour les « cliqueurs non inscrits » (soit 8 651 individus), nous disposions de peu d'informations puisqu'ils n'ont répondu ni au formulaire d'identification ni au questionnaire. Les concernant, seuls les facteurs manipulés étaient donc renseignés. Pour pallier cet inconvénient, nous avons décidé de compléter le profil sociodémographique de l'ensemble des répondants (renseigné par les réponses au formulaire d'identification) par les données déjà disponibles sur eux dans la base de données Elisa (si elles existent[312]), notamment afin de pouvoir disposer de cette information pour les « cliqueurs non inscrits ».

Le second fichier (appelé « post-test ») concerne les individus sollicités lors de la phase 2 (post-test) et leurs réponses à ce questionnaire, présenté au paragraphe 4.2. Il est composé de 7 641 colonnes, correspondant à l'ensemble des individus ayant répondu (dont 6 241 « inscrits » et « cliqueurs non inscrits » de la phase 1) et d'une quarantaine de colonnes, qui correspondent aux variables suivantes :

- le segment auquel l'individu a été affecté suite à sa réaction lors de la phase 1 : codé 1 à 4 (pour rappel, on distingue les non ouvreurs, les ouvreurs, les cliqueurs, les inscrits)
- les 4 facteurs manipulés lors de la phase 1 : codés 1 ou 2
- chacune des questions posées dans le questionnaire post-test (de PTQ1 à PQ28), correspondant, en particulier, aux variables de contrôle

Puisque nous nous intéressons surtout au comportement des inscrits et des cliqueurs de la phase 1 (car exposés concrètement à la sollicitation de données) et aux raisons possibles d'un tel comportement (mesurées dans le « post-test »), nous avons constitué un troisième fichier (dit « mixte »). Celui-ci a consisté à regrouper les deux fichiers précédents (réponses au

[312] On constate, à cet égard, que les données de la base Elisa sont beaucoup plus renseignées pour les individus appartenant au segment des « anciens » que pour les récents », ce qui est tout à fait logique.

questionnaire expérimental et au post-test) pour l'ensemble des inscrits et des cliqueurs ayant accepté de répondre au post-test (soit 6 241 individus sur les 7 641 au total[313]). Nous reviendrons dans la section suivante sur le profil des individus appartenant à chacun de ces fichiers (particulièrement aux fichiers 1 et 3) et sur les biais engendrés par cette procédure.

La figure 12.4 ci-dessous reprend le processus opéré pour obtenir le fichier « mixte » :

Figure 12.4 Processus de sélection des données

Fichier 1 : Requalification

N° Individu	Traitement *	Familiarité	Contexte	Quantité	Ordre	Age	Sexe	Région	P1	…	P16	P17	…	P42
1	16	2	2	1	1	26	2							
2	1	1	1	2	2	32	1							
3	7	1	2	1	2	21	1							
4	10	2	1	2	1	54	2							
5	2	1	1	1	1	44	2							
…														
20 805														

Fichier 2 : Post-Test

N° Individu	Segment	Traitement	Familiarité	Contexte	Quantité	Ordre	PTQ1	…	PTQ28
1	2	16	2	2	1	1			
2	3	1	1	1	2	2			
3	1	7	1	2	1	2			
4	2	10	2	1	2	1			
5	1	2	1	1	1	1			
…									
7 641									

Fichier 3 : Mixte (fichier de travail)

N° Individu	Traitement	Familiarité	Contexte	Quantité	Ordre	Age	Sexe	Région	P1	…	P16	P17	…	P42	Segment	PTQ1	…	PTQ28
1	16	2	2	1	1	26	2								2			
2	1	1	1	2	2	32	1								3			
3	7	1	2	1	2	21	1								1			
4	10	2	1	2	1	54	2								2			
5	2	1	1	1	1	44	2								1			
…																		
6 241																		

* Pour rappel, le numéro de traitement affecté (de 1 à 16 puisqu'il s'agit d'un design 2*2*2*2) est fonction du niveau correspondant de chaque facteur manipulé.
Ainsi, un individu « ancien » (codé 2), sollicité pour l'enquête (codée 2) et affecté à un questionnaire court (codé 1) et neutre (1) correspond au traitement « 16 » (cf. tableau 12.3).

Voyons à présent les caractéristiques des individus appartenant à ces échantillons de travail.

[313] La différence entre les 2 (7 641 – 6 241), soit 1 400 individus, correspond à l'ensemble des ouvreurs et des non ouvreurs de la phase 1 ayant répondu au post-test mais supprimé du fichier final (dit « mixte »), de par leur moindre intérêt dans le cadre de cette recherche.

5 Choix et caractéristiques des répondants

Dans cette expérience, l'échantillon a été sélectionné aléatoirement au sein de la base de données de l'entreprise partenaire (Base Elisa, mutualisée par Millemercis). Les participants ont tous été sollicités par email, sans avoir pour autant conscience de participer à une expérience. L'email envoyé, qui proposait de participer à une opération particulière (jeu-concours ou simple enquête), correspond en effet, en tout point, au type de courriel qu'ils reçoivent régulièrement dans le cadre de leur participation au programme Elisa. Cette solution, tout en introduisant un biais de non réponse, fait donc entièrement disparaître le biais du répondant. De plus, l'auto-sélection s'opère sur l'attractivité de la proposition et non sur les problèmes de Respect de Vie Privée, ce qui réduit donc son impact dans le cadre de notre sujet d'étude. En outre, l'échantillon sélectionné se veut représentatif de la population de base (la base de données de l'entreprise). Sans être pour autant représentatif de la population des internautes français, il y a cependant des chances pour qu'il s'en rapproche, au moins en ce qui concerne certains critères (âge, sexe, …), ce qui limite donc le biais de couverture.

Nous allons tout d'abord présenter les caractéristiques de l'échantillon correspondant à l'opération de requalification (phase 1 de l'expérience) (5.1) avant de comparer les deux échantillons utilisés pour tester nos hypothèses dans le cadre de cette étude (requalification puis mixte) et d'évoquer le problème du biais engendré par ce choix (5.2).

5.1 Les caractéristiques de l'échantillon de requalification

Pour rappel, l'email envoyé en phase 1 a été adressé à 239 923 individus et a été reçu par 227 994 d'entre eux (une fois retirées les adresses invalides). Parmi eux, 136 771 ne l'ont pas ouvert (non ouvreurs), 70 418 l'ont ouvert mais n'ont pas cliqué dessus (ouvreurs) et 20 805 ont cliqué dessus et se sont alors retrouvés sur la page d'accueil du site expérimental. 8 651 ont alors quitté le site sans s'enregistrer (cliqueurs non inscrits), les 12 154 autres ayant choisi de remplir et de valider le formulaire d'identification puis de compléter (ou non) le questionnaire qui suivait (inscrits).

A partir des renseignements issus de la base de données Elisa (spécialement le sexe, l'âge, la CSP et la région d'habitation), utiles pour identifier le profil des cliqueurs non inscrits, et de ceux fournis par les participants eux-mêmes (inscrits), soit à travers le formulaire (sexe, âge, région), soit en remplissant le questionnaire, nous pouvons connaître le profil de l'échantillon

de requalification (soit 20 805 individus) d'une part et comparer le profil des répondants (inscrits) et des non-répondants (cliqueurs non inscrits)[314] d'autre part.

Le tableau 12.6 ci-dessous présente le profil des répondants (inscrits) et des non répondants (cliqueurs non inscrits) à la phase 1 de l'expérimentation. Il indique également la significativité du test de différence de profil sociodémographique entre ces deux populations.

Tableau 12.6 Profil sociodémographique des répondants et des non répondants

Variables descriptives		Répondants (inscrits)	Non répondants (cliqueurs non inscrits)	Test de différence (Chi-deux)
Taille de l'échantillon		N = 12 154	N = 6 241	
Sexe	Hommes Femmes	37,8% 62,2%	52,0% 48,0%	Chi-deux : 355,9 p = 0,000
Age	- 18 ans 18-34 ans 35-54 ans + 55 ans	2,3% 49,3% 37,8% 10,6%	2,1% 40,7% 43,2% 14,3%	Chi-deux : 156,3 p = 0,000
CSP	Inactifs CSP – CSP moyen CSP +	32,9% 7,8% 42,9% 16,4%	21,1% 14% 52,6% 14,3%	Chi-deux : 146,9 p = 0,000
Région	Ile de France (IdF) Province	27,7% 72,3%	32,6% 67,4%	Chi-deux : 77,5 p = 0,000

Les conclusions issues de ce tableau permettent de répondre à un des points que nous cherchions à vérifier à travers cette troisième expérience, à savoir : existe-t-il une différence entre les profils du répondant et du non-répondant face à une sollicitation de données personnelles en ligne ? La réponse est donc affirmative puisque l'ensemble des tests de chi-deux sont significatifs. Ainsi, on trouve que le répondant est majoritairement une femme jeune (18 à 34 ans), inactive ou de CSP moyenne tandis que le non-répondant est plutôt un homme, d'âge moyen (35-54 ans) et de CSP moins ou moyenne.

L'annexe 12.9 présente le profil sociodémographique de l'ensemble des individus sollicités lors du Post-Test selon le type d'email envoyé d'une part (jeu vs. enquête et brandé Dauphine vs. non brandé) et le segment auquel ils ont été affectés suite à leur réaction lors de la phase 1 d'autre part (inscrit, cliqueur non inscrit, ouvreur, non ouvreur). Nous observons, là encore,

[314] Nous les qualifions respectivement de répondants et de non-répondants car, sollicités pour fournir des données personnelles, ils ont accepté (pour les premiers) ou non (pour les seconds) de les communiquer.

des différences significatives entre les individus appartenant respectivement à ces 4 segments, ce qui renforce l'idée que les répondants diffèrent des non répondants.

Voyons à présent les différences entre nos deux échantillons et les biais que cela engendre.

5.2 Comparaison des échantillons des phases 1 et 2

Concernant le post-test, rappelons que l'email a été envoyé à 41 615 individus et reçu par 41 205 d'entre eux. Parmi eux, 7 641 personnes ont accepté de répondre au questionnaire post-test, dont 4 571 inscrits et 1 655 cliqueurs de la phase 1. Nous avons alors choisi de considérer cet échantillon « réduit » de 6 241 individus (fichier mixte), sans tenir compte des 923 ouvreurs et des 492 non ouvreurs de la première phase, qui ont néanmoins accepté de répondre au post-test. Ceci permet de travailler, dans les deux cas (i.e. fichier de requalification et fichier mixte), sur les « mêmes » populations (inscrits et cliqueurs) et de confronter leur comportement lors de l'opération de requalification (phase 1) à leurs réponses lors du post-test (phase 2). A noter que ces 6 241 individus constituent un sous-échantillon de celui obtenu lors de l'opération de requalification (20 805 personnes), correspondant à l'ensemble des individus ayant cliqué sur le premier email et ayant répondu au questionnaire post-test. Nous devrons donc vérifier, dans un second temps, que le choix de travailler sur ce troisième « échantillon » n'entraîne pas de biais.

Le tableau 12.7 ci-dessous compare le profil des répondants de la phase 1 et de la phase 2.

Tableau 12.7 Description du profil des individus appartenant aux différents échantillons

Variables descriptives		Fichier 1 (Requalification)	Fichier 3 (Mixte)	Population des internautes français fin 2004 (source Médiamétrie et Ipsos)
Taille de l'échantillon (ou de la population)		N = 20 805	N = 6 241	23,9 Millions (26,8 en 2006)
Caractéristiques sociodémographiques				
Sexe	Hommes Femmes	42,9% 57,1%	39,4% 60,6%	53% 47%
Age	- 18 ans 18-34 ans 35-54 ans + 55 ans	2,2% 46,5% 39,5% 11,7%	1,6% 47,7% 36,9% 11,9%	- 14 ans : ND 15-34 ans : 52% 35-49 ans : 29% + 50 ans : 18%

CSP	Inactifs CSP – CSP moyen CSP +	31,4% 8,6% 43,9% 16,1%	34,3% 7,5% 43,1% 15,1%	32% 27% 42%
Région	Ile de France Province	29,7% 70,3%	27,4% 72,6%	26% 74%
Niveau d'études	Collège-Lycée Niveau bac Bac +1 ou 2 Bac +3 ou 4 Bac +5 et plus	24,3% 20,1% 28,0% 17,4% 10,2%	23,2% 19,6% 29,2% 18,0% 10,1%	ND
Pays	France Autre	96,7% 3,3%	97,3% 2,7%	/
Situation familiale	En couple Célibataire	63,4% 36,6%	63,4% 26,6%	ND
Revenu Net Mensuel	< 900 € 900-3 000 € > 3 000 €	7,1% 61,2% 19,6%	7,0% 62,9% 20,3%	< 1 000 € :20% 1 000-3 000 € : 39% > 3 000 € : 14%
Caractéristiques liées à Internet				
Nb moyen d'adresses emails	1 2 + de 3	44,5% 29,5% 26,0%	43,5% 28,9% 27,6%	ND
Durée moyenne d'utilisation	- de 2 ans 2 à 5 ans + de 5 ans	ND	14,5% 37,9% 47,6%	ND
Fréquence de connexion au bureau	Jamais - d'1 fois/mois 1 fois/semaine Tous les jours	56,6% 2,6% 6,4% 32,9%	57,6% 2,4% 6,2% 32,0%	Lieu de connexion : Domicile : 77% Travail : 38%
Fréquence de connexion au domicile	Jamais - d'1 fois/mois 1 fois/semaine Tous les jours	9,3% 0,9% 11,8% 76,6%	8,7% 0,7% 10,5% 78,9%	
Expérience d'achat		93,7%	ND	44,7%

Il apparaît, à la lecture de ce tableau, plusieurs résultats intéressants. Tout d'abord, d'un point de vue sociodémographique, les deux échantillons interrogés semblent proches de la population des internautes français, notamment en ce qui concerne l'âge, la CSP et la région d'habitation. Des différences apparaissent néanmoins concernant la répartition hommes/femmes, favorables aux femmes dans l'échantillon Elisa alors qu'elle est majoritairement masculine dans la population des internautes français. Dans l'ensemble cependant, ceci confirme l'idée selon laquelle le biais de couverture serait, ici, limité.

Le second résultat concerne le biais engendré par l'analyse qui sera menée sur l'échantillon mixte et qui peut être considéré comme un biais de sélection. Dans ce troisième échantillon en

effet, nous analysons les réponses d'une partie (6 241) des 20 805 individus appartenant au premier échantillon. Nous observons alors de nettes différences dans la structure de ces deux échantillons, l'« effet » de non-réponse étant, dans ce cas, renforcé. Par cela, nous entendons que les individus s'étant inscrits lors de la première phase ont également davantage répondu à la seconde. Comme il existe des différences significatives entre les répondants (inscrits) et les non répondants (cliqueurs non inscrits) (cf paragraphe 5.1.1), cet écart est donc renforcé dans l'échantillon mixte. Pour vérifier ce biais, de manière objective, nous avons effectué des comparaisons de moyenne sur variables qualitatives (tests de chi-deux) entre le profil sociodémographique de l'individu et son statut de répondant (ou non) au post-test, pour les 20 805 individus du premier échantillon. Le tableau 12.8 ci-dessous présente ces résultats.

Tableau 12.8 Profil sociodémographique des répondants et des non répondants au post-test

Variables descriptives		Répondants au post-test	Non répondants au post-test	Test de différence (Chi-deux)
Taille de l'échantillon		N = 6 241	N = 14 564*	
Statut lors de la phase 1	INSCRIT CLIQUEUR	73,5% 26,5%	52% 48%	Chi-deux : 834,5 p = 0,000
Sexe	Hommes Femmes	39,4% 60,6%	44,5% 55,5%	Chi-deux : 43,1 p = 0,000
Age	- 18 ans 18-34 ans 35-54 ans + 55 ans	1,6% 47,7% 36,9% 11,9%	2,5% 45,9% 39,8% 11,7%	Chi-deux : 19,2 p = 0,004
CSP	Inactifs CSP – CSP moy CSP +	34,3% 7,5% 43,1% 15,1%	29,6% 9,2% 44,5% 16,7%	Chi-deux : 34,3 p = 0,000
Région	IdF Province	27,4% 72,6%	30,7% 69,3%	Chi-deux : 55,9 p = 0,000

* 14 564 = 20 895 (taille de l'échantillon 1) – 6 241 (taille de l'échantillon 2)

Les résultats confirment l'existence d'un biais à travailler sur l'échantillon mixte, parce qu'il est différent, au moins sociodémographiquement parlant, du premier. Toutefois, pour pouvoir mener nos analyses avec rigueur, et spécialement contrôler l'effet possible des variables exogènes, nous ne pouvons nous contenter du premier échantillon, étant donné qu'il ne contient aucun instrument de mesure de cet ordre. Vouloir limiter ce biais de sélection pourrait alors entraîner un autre problème, lié à une moindre validité des résultats, de par l'absence de contrôle, démarche pourtant primordiale dans toute expérimentation.

Pour limiter ce biais de sélection, nous ferons donc, si possible, nos analyses sur les deux échantillons, que nous comparerons ensuite. De plus, nous procéderons à une pondération des individus de l'échantillon mixte, afin de tenir compte de cette répartition différente. Cette pondération s'effectuera sur la base du statut de répondant lors de la phase 1 (inscrit vs. cliqueur) afin de conserver la répartition initiale entre ces deux catégories dans l'échantillon mixte. Le tableau suivant explique le principe de cette pondération.

Tableau 12.9 Principe de pondération appliqué dans l'échantillon mixte

Statut de répondant	Echantillon 1 (requalification)	Répartition (%)	Echantillon 3 (mixte)	Répartition (%)	Coefficient de pondération
Inscrit	12 154	58,4%	4 587	73,5%	0,795 *
Cliqueur	8 651	41,6%	1 654	26,5%	1,570
TOTAL	20 805	100%	6 241	100%	-

* 0,795 = 58,4 / 73,5

Ayant expliqué sur quelles données nous allions travailler, nous présentons maintenant les principes à respecter pour mener à bien nos analyses et ainsi tester nos hypothèses.

6 L'analyse des données

Nous présentons dans cette section les analyses préalables, nécessaires au test des hypothèses. Nous analyserons tout d'abord la validité interne du plan d'expérience (6.1), avant d'étudier la qualité psychométrique des instruments de mesure (6.2). Nous indiquerons ensuite le plan de traitement utilisé dans le cadre de cette étude (6.3), avant d'aborder l'adéquation des données aux tests statistiques envisagés (6.4).

6.1 La validité interne du plan d'expérience

Un plan d'expérience doit répondre à plusieurs conditions pour conserver sa validité et ainsi permettre d'exécuter les analyses requises pour tester les hypothèses. Précisons toutefois que, contrairement à la majorité des expériences en marketing et à celles menées dans le cadre des expériences 1 et 2, nous ne procéderons pas ici à des analyses de variance. Les variables dépendantes étudiées sont en effet, pour la majorité, binaires et requièrent donc d'autres méthodes d'analyse. Certaines des conditions nécessaires à l'analyse de variance (identifiées notamment par Howell 1998) sont donc exclues de cette démarche. Nous retenons alors 2

conditions majeures à vérifier avant de mener à bien nos analyses. Il s'agit du contrôle des variables exogènes (6.1.1) et de l'indépendance des observations (6.1.2).

6.1.1 Le contrôle des variables exogènes

Nous avons déjà indiqué, dans le paragraphe 2.2, le nom des variables contrôlées ainsi que le type de contrôle réalisé. Outre l'affectation aléatoire des individus aux différents traitements, nous avons prévu un contrôle statistique *a posteriori* sur l'ensemble des variables externes identifiées comme potentiellement influentes. Ces variables seront donc introduites dans chacune des analyses menées, avec le statut de covariables. Nous pourrons ainsi contrôler et évaluer leur influence sur les variables dépendantes étudiées.

6.1.2 L'indépendance des observations

Il s'agit de vérifier que les différences entre traitements ne sont pas dues à une différence dans la composition de l'échantillon. Cette condition est assurée *a priori* par l'affectation aléatoire des unités expérimentales aux différents traitements. Un contrôle *a posteriori* est cependant nécessaire pour vérifier que les traitements ne se distinguent pas au niveau de leurs caractéristiques individuelles. Nous avons déjà indiqué précédemment dans ce chapitre que l'échantillon avait été construit de manière à ce que, à ancienneté égale, il n'y ait aucune différence significative entre les individus affectés aux différents traitements. Nous vérifions à présent cette indépendance des observations pour les deux autres facteurs manipulés : le contexte et la quantité.[315]. Nous testons le lien entre les facteurs et les caractéristiques individuelles principales (âge, sexe et CSP) par des tests de Chi-deux, dont nous présentons les résultats dans le tableau suivant :

Tableau 12.10 Résultats des tests de chi-deux pour tester l'indépendance des observations

Variables individuelles	Contexte : Jeu vs. Enquête	Quantité : Court vs. Long
Age	*p = 0,027*	p = 0,072
Sexe	p = 0,168	p = 0,253
CSP	p = 0,055	p = 0,697

[315] Le facteur Ordre n'étant pas intégré dans les analyses, nous nous abstenons pour l'instant de vérifier l'indépendance des observations pour ce critère.

Au seuil habituel de 5%, les observations apparaissent donc indépendantes, sauf pour l'âge qui se distingue sur les traitements Jeu et Enquête. A noter que la CSP et l'âge sont respectivement proches du seuil de significativité pour les traitements Jeu vs Enquête et Court vs. Long. L'ensemble de ces variables individuelles seront donc incluses dans les analyses en tant que co-variables, afin de tester leur influence sur les variables dépendantes étudiées.

6.2 La qualité psychométrique des instruments de mesure.

L'étude de la qualité psychométrique des instruments de mesure consiste à tester leur validité et leur fiabilité. Il faut s'assurer de la validité et de la fiabilité des échelles de mesure multi-items (6.2.1) et de la validité discriminante de tous les instruments de mesure (6.2.2), afin d'éviter, en particulier, les problèmes de multi-colinéarité.

Pour rappel, nous disposons, dans le cadre de cette troisième expérience, de 3 catégories de variables, mesurées sur des échelles mono ou multi-items :

- des variables dépendantes, dont
    - 3 variables binaires : la communication de données personnelles (CDP), l'abandon et le niveau de dévoilement (réponse à la question sur le mobile) ;
    - l'intention de poursuivre la relation avec l'entreprise sollicitante (IPR), mesurée sur une échelle de Likert à 3 items.
- des variables indépendantes, dont :
    - les 3 facteurs manipulés ;
    - la préoccupation pour le respect de la vie privée (PREOC), mesurée en 3 items
- des variables de contrôle (sauf d'ordre sociodémographique et expérientiel), dont :
    - le comportement habituel de réponse face aux sollicitations de données personnelles et aux emails envoyés dans le cadre du programme Elisa, mesuré sur des échelles mono-items, hormis l'habitude d'abandonner (2 items)[316] ;
    - la qualité perçue du programme Elisa (QUALITE), mesurée sur une échelle de Likert en 3 items

Les items adoptés pour mesurer le comportement habituel sont présentés ci-dessous. Soulignons que l'item employé pour l'habitude de répondre (CHR) est identique à celui utilisé dans l'expérience 2. Les autres items sont créés spécifiquement pour l'expérience 3.

[316] Comme la mesure de l'habitude d'abandonner ne comporte que 2 items, nous ne considérons pas celle-ci comme une mesure multi-items. Nous nous « contentons » dès lors de calculer un score à partir de ces 2 items.

Tableau 12.11 Echelle de mesure du comportement habituel

Concept	Items	Mesure
Comportement habituel de réponse face aux sollicitations de données personnelles		
habitude de répondre (CHR)	« Quand une entreprise me demande des informations personnelles, je les lui donne facilement »	Likert en 7 points (de 1 : pas du tout d'accord, à 7 : tout à fait d'accord)
habitude de tout donner (CHD)	« Lorsque je réponds à un questionnaire, je réponds à toutes les questions qui me sont posées »	
habitude d'abandonner avant la fin (2 items) (CH_abandon, score sur ces deux items)	- « Lorsque je réponds à un questionnaire, je m'arrête lorsque ce questionnaire est trop long » - « Lorsque je réponds à un questionnaire, je m'arrête lorsqu'une question est trop personnelle »	
Comportement habituel de réponse face aux emails envoyés par le programme Elisa		
habitude de cliquer (CH_clic)	« [Quand je reçois des emails du programme Elisa], je clique dessus »	Ordinale à 4 modalités (de 1 très souvent à 4 très rarement)

6.2.1 La validité et la fiabilité des échelles de mesure multi-items

Nous vérifions donc, dans un premier temps, la qualité psychométrique des échelles d'intention de poursuivre la relation (IPR), de préoccupation pour le respect de la vie privée (PREOC) et de qualité perçue du programme (QUALITE). La validité et la fiabilité de ces instruments de mesure seront évaluées par le biais d'une analyse factorielle et par le calcul du coefficient de fiabilité (Alpha de Cronbach). Rappelons que le développement et la vérification de la qualité psychométrique des deux premiers instruments de mesure ont déjà été présentés dans le chapitre 9. Nous nous contenterons donc d'exposer les résultats obtenus dans le cadre de cette étude. Seul le troisième instrument, correspondant à la qualité perçue du programme, est réellement nouveau. Il a été développé en collaboration étroite avec l'entreprise Millemercis, qui utilisait déjà ce type d'items, dans le cadre de ses enquêtes de satisfaction. Nous avons donc repris ces items, tout en les adaptant au contexte évoqué ici. Dans la mesure où cette échelle ne comporte que 3 items, nous nous contenterons de mener une analyse factorielle exploratoire[317], suivie de la vérification de sa fiabilité par le biais de l'alpha de Cronbach.

[317] La réalisation d'une analyse confirmatoire, par le biais d'un modèle de mesure sous AMOS, nécessite en effet un minimum de 4 items, pour pourvoir obtenir les indices d'ajustement.

L'analyse factorielle de l'échelle d'intention de poursuivre la relation, menée sur les 3 items conservés suite à l'analyse confirmatoire présentée au chapitre 9, donne les résultats suivants (KMO = 0, 645 ; sign du test de Bartlett = 0,000) :

Tableau 12.12 Validité et cohérence interne de l'échelle d'intention de poursuivre la relation

Items	Saturation	Alpha de Cronbach si on enlève cet item
IPR 1 : je pense rester inscrit sur Elisa pendant longtemps	0,846	0,504
IPR 2 (i) : je ne vois pas l'intérêt de participer aux opérations proposées par le programme Elisa dans le futur	0,760	0,651
IPR 3 : je recommanderai à mes amis de se rendre sur le site du programme Elisa afin de bénéficier des mêmes offres que moi	0,766	0,653
% de variance expliquée	62,66%	
Alpha de Cronbach	0,70	

La validité interne et la fiabilité de cette échelle sont donc satisfaisantes, même si l'alpha atteint tout juste les limites acceptables (Peterson 1995). A noter qu'il s'avère comparable à celui obtenu sur l'échantillon de l'Expérience 2 (alpha = 0,73 si l'on considère les 4 items initiaux et alpha = 0,70 en enlevant l'item supprimé lors de l'analyse confirmatoire).

L'analyse factorielle de l'échelle de préoccupation pour le respect de la vie privée, menée sur les 3 items conservés suite à l'analyse confirmatoire présentée au chapitre 9, donne les résultats suivants (KMO = 0, 623 ; sign du test de Bartlett = 0,000) :

Tableau 12.13 Validité et cohérence interne de l'échelle de préoccupation pour le RVP

Items	Saturation	Alpha de Cronbach si on enlève cet item
PREOC 1 : lorsque je donne des informations personnelles à une entreprise, je suis inquiet(e) à l'idée que ces entreprises puissent détenir des données que je considère comme privées	0,854	0,480
PREOC 2(i) : lorsque je donne des informations personnelles à une entreprise, l'idée que mes données puissent être utilisées sans mon accord ne m'inquiète pas particulièrement	0,689	0,729
PREOC 3 : lorsque je donne des informations personnelles à une entreprise, je suis préoccupé(e) par le fait que les informations que je donne puissent être utilisées à mon insu	0,816	0,567
% de variance expliquée	62,28%	
Alpha de Cronbach	0,69	

La validité interne et la fiabilité de cette échelle sont donc globalement satisfaisantes, même si l'alpha atteint tout juste les limites acceptables (Peterson 1995). A signaler que l'alpha est beaucoup moins bon que celui obtenu pour les Expériences 1 et 2 (respectivement de 0,87 et de 0,76 sur ces 3 items). Avoir précisé en début d'énoncé « lorsque je donne des informations personnelles à une entreprise »[318] (ce qui n'était pas le cas dans les deux expériences précédentes) pourrait expliquer en partie ce résultat, car cela alourdirait la formulation et rendrait les items moins « lisibles ».

En ce qui concerne l'échelle de qualité perçue de la relation (en l'occurrence du programme Elisa pour les individus interrogés), nous avons retenu les trois items suivants :

Tableau 12.14 Echelle de mesure de la qualité perçue du programme Elisa

Items	Mesure
QUALITE 1 : les offres envoyées par le programme Elisa correspondent à mes centres d'intérêt QUALITE 2 : les offres proposées par les partenaires du programme Elisa sont intéressantes QUALITE 3 : j'apprécie les annonceurs qui proposent les offres envoyées par le programme Elisa	Echelle de Likert en 7 points

L'analyse factorielle de l'échelle de qualité perçue du programme donne les résultats suivants (KMO = 0, 734 ; sign du test de Bartlett = 0,000) :

Tableau 12.15 Validité et cohérence interne de l'échelle de qualité perçue du programme

Items	Saturation	Alpha de Cronbach si on enlève cet item
QUALITE 1 : les offres envoyées par le programme Elisa correspondent à mes centres d'intérêt	0,876	0,881
QUALITE 2 : les offres proposées par les partenaires du programme Elisa sont intéressantes	0,921	0,810
QUALITE 3 : j'apprécie les annonceurs qui proposent les offres envoyées par le programme Elisa	0,913	0,824
% de variance expliquée	81,62%	
Alpha de Cronbach	0,88	

[318] Ce choix est lié à la présentation du questionnaire (qui est en outre administré en ligne) qui ne permettait pas d'inclure cette précision au sein de la question et rendait donc nécessaire de la formuler dans les différents items.

La validité interne et la fiabilité de cette échelle sont donc très satisfaisantes, surtout pour une échelle en construction. L'alpha est proche du seuil de 0,9, ce qui pourrait laisser craindre un risque de redondance entre les items, mais ceci est démenti par la lecture de leurs énoncés.

La validité et la fiabilité de ces échelles ayant été établies, nous créons 3 nouvelles variables (ScorIPR, ScorPREOC et ScorQUALITE) par simple addition des items correspondants.

6.2.2 La validité discriminante des instruments de mesure

La qualité psychométrique d'un instrument de mesure s'apprécie aussi par sa différenciation avec les autres instruments de mesure utilisés.
Pour les échelles multi-items, une analyse factorielle avec rotation Varimax est utilisée afin de montrer que les trois échelles représentent bien des construits différents. Les résultats confirment que ces trois échelles se distinguent les unes des autres, les items relatifs à chacune d'elles saturant bien sur la dimension correspondante (cf. Annexe 12.10).
Tester la validité discriminante des instruments de mesure mono-item (notamment pour ceux correspondant aux variables de contrôle) est primordial car nous devons éviter tout risque de multi-colinéarité trop important qui viendrait biaiser les résultats. Nous avons donc mené des analyses bivariées (tests de corrélations), pour vérifier toute absence de forte multi-colinéarité. Les analyses de corrélations (fournies en Annexe 12.11) montrent que les mesures sont toutes significativement liées aux autres. Ce résultat n'est pas étonnant dans la mesure où, bien que correspondant à des concepts différents, la plupart de ces variables (comportement de réponse, préoccupation, qualité perçue) sont conceptuellement proches. Néanmoins, ces variables sont faiblement corrélées les unes aux autres puisque l'indice de corrélation est inférieur à 0,3 à l'exception des variables ScorQualité et CH_clic (corrélation de Pearson = -0,341). Nous renonçons donc à introduire ces deux variables simultanément. Pour les autres, nous pouvons conclure à une absence de multi-colinéarité forte.

6.3 Le plan de traitement

Le plan des traitements réalisés pour tester les hypothèses relatives à cette troisième expérience est présenté. L'objectif principal de cette étude est de tester des liens de causalité entre des variables de statut différent. Le type de traitement choisi dépend donc de la nature des variables testées, qu'elles soient dépendantes ou indépendantes. Le tableau ci-dessous indique le type d'analyse préconisé en fonction de la nature des variables.

Tableau 12.16 Les différentes techniques d'analyse selon la nature des variables testées

		Variables dépendantes	
		Qualitative (binaire)	Quantitative (métrique)
Variables indépendantes	Qualitative (binaire)	Test de chi-deux	ANOVA
	Mixtes	Régression logistique	ANCOVA ou Régression multiple

En l'occurrence, nous procéderons à des régressions logistiques pour tester l'effet de la sollicitation de données personnelles (et en particulier des différents facteurs manipulés) sur les 3 variables de « Réponse », à savoir : la Communication de Données Personnelles (CDP) (i), l'Abandon (j) et le Niveau de dévoilement (réponse à une question sensible) (k).

Pour tester l'effet de certains facteurs manipulés et du comportement de réponse sur l'intention de poursuivre la relation (IPR), nous aurons recours à une régression multiple plutôt qu'à une ANCOVA qui nécessite des conditions particulières relatives à l'analyse de variance. Nous résumons, dans le tableau suivant, l'ensemble des analyses qui seront menées dans le cadre de cette troisième expérience.

Tableau 12.17 Plan de traitement des hypothèses correspondant à l'expérience 3

Variable dépendante	Variables indépendantes	Hypothèses testées	Variables de contrôle	Méthode d'analyse
Communication de données personnelles (CDP) (i)	- facteur Familiarité - facteur Contexte - préoccupation RVP (ScorPREOC)	H 1.2 i H 1.4 i H 3 i	- âge et sexe - utilisation (UTIL) et expérience d'Internet (EXP) - CHR - CH_clic	Régression logistique (1)
Abandon (j)	- facteur Familiarité - facteur Contexte - facteur Quantité - préoccupation RVP (ScorPREOC)	H 1.2 j H 1.4 j H 1.3 j H 3 j	- âge, sexe, CSP - niveau d'études - utilisation (UTIL) et expérience d'Internet (EXP) - CH_abandon	Régression logistique (2)
Niveau de Dévoilement (ND) (k)	- facteur Familiarité - facteur Contexte - facteur Quantité - préoccupation RVP (ScorPREOC)	H 1.2 k H 1.4 k H 1.3 k H 3 k	- âge, sexe, CSP - niveau d'études - utilisation (UTIL) et expérience d'Internet (EXP) - CHD	Régression logistique (3)

Intention de poursuivre la relation (ScorIPR) (h)	- facteur Familiarité - facteur Contexte - CDP (i)	H 1.2 h H 1.4 h H 6 i	- âge et sexe - utilisation (UTIL) et expérience d'Internet (EXP) - ScorQUALITE	Régression multiple (4)
ScorIPR (h)	- facteur Familiarité - facteur Contexte - Abandon (j)	H 1.2 h H 1.4 h H 6 j	- âge, sexe, CSP - niveau d'études - utilisation (UTIL) et expérience d'Internet (EXP) - CHD	Régression multiple (5)

Nous reviendrons, dans le paragraphe suivant, sur les conditions à respecter pour mener ces analyses et sur les statistiques à contrôler pour vérifier si les hypothèses sont ou non validées.

Bien que n'ayant pas formulé d'hypothèse à ce sujet, nous constatons que les variables CDP (i) et Abandon (j) sont positionnées en tant que médiatrices entre les facteurs manipulés et l'intention de poursuivre la relation (IPR). De manière classique, nous testerons ce rôle médiateur, en regardant si les conditions données par Baron et Kenny (1986) et exposées dans le chapitre précédent (au paragraphe 5.4) sont bien vérifiées.

Nous symbolisons ces conditions, dans le cadre de l'expérience 3, dans la figure suivante :

Figure 12.5 Les relations de médiation dans le cadre de l'expérience 3

Dans la mesure où les variables indépendantes et médiatrices sont binaires et la variable dépendante est continue, nous aurions pu tester la relation (2) par un test de chi-deux, et, les relations (1) et (3) par des ANOVA. Toutefois, puisque nous y incluons les variables de contrôle, nous les testerons respectivement par des régressions logistiques et multiples.

6.4 L'adéquation des données à des tests statistiques

Pour mener à bien nos analyses, nous devons vérifier que les conditions d'utilisation des méthodes auxquelles nous souhaitons recourir sont respectées (6.4.1). Nous indiquons ensuite quels sont les « indices » sur lesquels nous pouvons nous appuyer pour vérifier si les hypothèses sont ou non validées (6.4.2).

6.4.1 Les conditions d'utilisation de la régression

Ces conditions varient selon que l'on a recours à une régression logistique ou multiple. Des conditions communes peuvent également être identifiées.

Les principales conditions à respecter pour la régression logistique sont les suivantes :

1. pas de variable explicative oubliée qui soit corrélée avec celles du modèle
2. pas d'erreur de mesure des variables indépendantes
3. les observations sont indépendantes
4. la variable dépendante n'agit pas sur les variables indépendantes

Les trois premières conditions ont déjà été vérifiées dans les paragraphes précédents : l'ensemble des variables exogènes susceptibles d'influencer les variables dépendantes ont été incluses dans le modèle, la qualité psychométrique des variables indépendantes a été vérifiée et l'indépendance des observations a été confirmée (hormis pour l'âge dans la condition Jeu vs. Enquête). La quatrième condition est également vérifiée dans la mesure où il est peu probable que les variables de réponse puissent influencer les facteurs manipulés ici et relatifs à l'exposition à une sollicitation de données personnelles.

Les principales conditions à respecter pour la régression multiple ont déjà été exposées au chapitre 10 (paragraphe 5.4). Les conditions 1 (nullité de l'espérance des erreurs) et 3 (indépendance des observations) ont déjà été vérifiées précédemment pour cette expérience. Les conditions 2 et 4 seront, elles, vérifiées lors de l'analyse des résultats (cf. Section 7).

Deux conditions, non énumérées jusqu'ici et valables pour les deux méthodes d'analyse, concernent respectivement l'absence de multi-colinéarité (vérifiée au paragraphe 6.2) et la normalité des distributions. Cette dernière constitue un postulat fondamental selon lequel la

distribution des observations, pour chacune des variables métriques et assimilées métriques (cas des variables ordinales), doit suivre une loi normale. En cas d'écart par rapport à la normalité, les tests statistiques comme le t de Student deviennent en effet invalides. Nous avons déjà indiqué précédemment dans ce document que nous pouvions vérifier cette condition par l'examen de deux indicateurs : le coefficient d'asymétrie (*skewness*) et le coefficient d'aplatissement *(kurtosis*) qui doivent être respectivement inférieurs à 1 et 1,5 en valeur absolue. Les données seraient considérées comme modérément non normales pour des coefficients respectivement inférieurs à 2 et 7 et substantiellement anormales au-delà.

Nous avons tout d'abord vérifié la normalité de la seule variable à expliquer métrique (en l'occurrence l'intention de poursuivre la relation, IPR) pour chacun des facteurs manipulés. Les résultats fournis en Annexe 12.12 confirment le respect de cette hypothèse, car les coefficients d'asymétrie et d'aplatissement sont conformes aux règles énoncées ci-dessus.
Nous vérifions ensuite la normalité des variables indépendantes et des variables de contrôle (variables exogènes), tous traitements confondus. Les résultats (cf. Annexe 12.12) confirment une fois encore le respect de cette hypothèse, hormis pour les variables « habitude de tout donner » (CHD) et « fréquence de consultation de la messagerie » (UTIL), dont les coefficients d'asymétrie et d'aplatissement sont respectivement de (-1,478 ; 1,198) et (1,299 ; 1,353). La violation de l'hypothèse de normalité concerne donc uniquement la symétrie (assez faiblement toutefois), la concentration étant pour sa part respectée. Ces deux variables ont donc une distribution très modérément non normale, sans que cela ne puisse toutefois remettre en cause leur introduction au sein du modèle.

6.4.2 La validation des hypothèses

Nous présentons rapidement ci-dessous la manière de vérifier les hypothèses selon la méthode d'analyse utilisée : régression logistique, d'une part et multiple, d'autre part.

Validation des hypothèses dans le cadre de la régression logistique

La régression logistique est utilisée quand la variable dépendante est binaire et les variables indépendantes quelconques (qualitatives ou quantitatives). Sur le plan théorique, cette méthode convient parfaitement dans le cas présent puisque nous souhaitons obtenir des prévisions de y que l'on puisse interpréter comme des probabilités d'occurrence d'un événement (ici : de donner ou non les informations personnelles demandées, d'abandonner ou

non avant la fin, de fournir ou non son numéro de téléphone portable). Contrairement à la régression multiple où les paramètres sont estimés à partir des moindres carrés (OLS), la méthode d'estimation utilisée ici est celle du maximum de vraisemblance qui offre l'avantage d'aboutir à des estimations meilleures (notamment quand les distributions ne sont pas normales) et surtout non biaisées (quand les distributions sont normales).

La qualité d'ajustement global du modèle peut se faire par plusieurs méthodes :

1. faire un histogramme des prévisions : le graphique doit alors laisser apparaître une séparation claire entre les deux groupes d'individus (qui répondent respectivement de façon positive et négative à l'évènement).
2. comparer les prévisions du modèle avec les données : il s'agit ici, en analysant la matrice de confusion, de vérifier que le modèle aboutit à un pourcentage d'individus « bien classés » conséquent. S'il n'existe pas vraiment de règle de décision formellement établie dans ce domaine, nous pouvons néanmoins estimer qu'un modèle pour lequel ce critère serait inférieur à 50% serait un mauvais modèle. Au contraire, nous estimons qu'un modèle supérieur à 70% est « correct », voire satisfaisant.
3. tester dans quelle mesure les paramètres estimés sont probables : on utilise alors comme indicateur le « -2LL » (moins deux fois le log de la vraisemblance). Un bon modèle est celui qui a une vraisemblance élevée ce qui se traduit par une valeur du -2LL proche de zéro. On utilise aussi le test de Chi-deux qui compare le modèle spécifié avec le modèle de départ (ou avec le modèle précédent en cas de modèles emboîtés), ce qui revient à vérifier que tous les coefficients du modèle spécifiés sont égaux à zéro.
4. regarder les indicateurs de variance expliquée (R^2 de Cox & Snell) et de la qualité du classement (R^2 de Nagelkerke)

Pour statuer sur les relations testées dans le cadre de cette troisième expérience et valider ainsi (ou non) les hypothèses formulées, nous aurons recours aux trois dernières méthodes, la première étant, selon nous, plus délicate à interpréter. En outre, il convient de tester la significativité des coefficients obtenus (pour chaque variable indépendante introduite) par le test de Wald, qui suit une loi de distribution du chi-deux.

Validation des hypothèses dans le cadre de la régression multiple

La régression multiple s'utilise quand la variable dépendante est métrique et les variables indépendantes quelconques (qualitatives ou quantitatives). Le modèle doit alors comporter

toutes les variables indépendantes (y compris des variables de contrôle) qui influencent y et qui sont faiblement corrélées entre elles. La qualité d'ajustement global du modèle est vérifiée par l'étude du coefficient de détermination R^2 qui représente la proportion de variance expliquée par les variables du modèle. La significativité du R^2 se teste avec un test F de Fisher. Dans le cas de modèles emboîtés, il convient de comparer les R^2 ajustés et de vérifier qu'ils ont augmenté, suite à l'introduction des dernières variables. Enfin, la significativité des paramètres se teste par un test t de Student.

Ayant établi les conditions à vérifier pour tester nos hypothèses, nous procédons, à présent à l'analyse des résultats.

7 Présentation des résultats

Nous allons maintenant indiquer les résultats obtenus dans le cadre de cette expérience et valider ainsi (ou non) les hypothèses formulées. Pour ne pas alourdir la présentation, seules les variables exogènes significatives, parmi celles introduites dans les différents modèles (cf. Tableau 12.17), seront présentées. Nous avons, la plupart du temps, procédé par étapes, en introduisant d'abord les variables principales que l'on voulait tester (variables indépendantes) avant d'introduire les variables de contrôle, dans des blocs successifs.

Le plan de traitement présenté précédemment nous a permis d'identifier 5 analyses différentes à mener, sous la forme de régressions logistiques ou multiples.
Nous présenterons successivement les résultats permettant de statuer sur :

- l'effet de l'exposition à la sollicitation (familiarité et contexte) et de la préoccupation RVP (PREOC) sur la probabilité de communiquer des données personnelles (CDP, i)
- l'effet de l'exposition à la sollicitation (familiarité, contexte, quantité) et de la préoccupation RVP (PREOC) sur la probabilité d'abandonner avant la fin (Abandon,j)
- l'effet de l'exposition à la sollicitation (familiarité, contexte, quantité) et de la préoccupation RVP (PREOC) sur la probabilité de communiquer son numéro de téléphone mobile (ND, k)
- l'effet de l'exposition à la sollicitation (familiarité et contexte), de la communication (ou non) de données personnelles (CDP, i) et de l'abandon (ou non) avant la fin du questionnaire (Abandon, j) sur la poursuite de la relation avec l'entreprise sollicitante (IPR, h). Cet effet sera testé par deux régressions multiples successives.

Nous commençons (point 7.1) par un préambule indiquant les principaux résultats obtenus lors de l'opération de requalification (phase 1). Ils nous semblent en effet utiles pour présenter le cadre d'analyse et donner un premier aperçu des résultats ultérieurs.
Nous terminerons cette présentation (point 7.6) par l'examen des effets médiateurs présentés au point 6.3, bien qu'ils n'aient pas fait l'objet d'hypothèses spécifiques.

7.1 Préambule concernant la participation à l'opération de requalification

Nous présentons ci-dessous un premier bilan des résultats concernant la participation des individus à l'opération de requalification (phase 1 de l'expérimentation). Rappelons que nous nous intéressons uniquement ici aux « inscrits » et aux « cliqueurs non inscrits », c'est-à-dire à l'ensemble des individus ayant cliqué sur le premier email envoyé, dit de requalification.
Sur ces 20 805 individus qui ont cliqué sur l'email et qui se sont donc retrouvés sur la page d'accueil du mini site expérimental contenant le questionnaire, 12 154 se sont inscrits pour participer, soit un taux moyen d'inscription de 58,4%. Ce taux d'inscription est cependant différent d'un traitement à l'autre. Ainsi, il est plus faible pour les individus inscrits depuis longtemps sur Elisa (anciens) et pour l'enquête (respectivement 55% et 56,5%). On constate aussi, à la lecture des résultats, une interaction possible entre ces deux types de facteurs (familiarité et contexte). Ainsi, pour l'enquête, le taux d'inscription des anciens est plus élevé que celui des clients récemment inscrits sur Elisa (respectivement 7,3% et 5,1%).

En ce qui concerne le comportement de réponse au questionnaire, on observe, là encore, des différences selon les traitements. Ainsi, en moyenne, le nombre de questions ayant obtenues réponse est sensiblement le même que le questionnaire soit sensible ou non (Facteur Ordre) (22,22 contre 22,31) et que l'individu soit anciennement ou récemment inscrit (Facteur Familiarité) sur le programme Elisa (22,44 contre 22,15). Il est par contre plus faible pour le jeu comparativement à l'enquête (respectivement 20,92 contre 23,28), cette différence paraissant significative. De plus, en moyenne, les participants ont répondu à 30,5 questions pour le questionnaire long (sur les 42 questions proposées) contre 13,7 questions pour le questionnaire court (sur les 16 proposées). Pour le questionnaire court, les gens ont donc quasiment répondu à toutes les questions alors que pour le questionnaire long, ils ont répondu à un peu moins de 4 questions sur 5. Reste à savoir si c'est le fait d'un choix de non réponse (possibilité de ne pas répondre à une question, tout en poursuivant le questionnaire) ou d'abandon (arrêt du questionnaire avant la fin), ce que nous étudierons à travers nos analyses ultérieures.

7.2 Résultats concernant la probabilité de communiquer des données

Puisque nous nous intéressons ici à la communication de données personnelles, mesurée par le fait d'avoir accepté (ou non) de remplir le formulaire d'identification et de s'inscrire ainsi à l'opération proposée (jeu ou enquête), nous mènerons l'analyse sur les deux types de fichiers disponibles (cf. point 4.3). Il s'agit du fichier de requalification (composé de 20 805 individus dont 12 154 inscrits) et du fichier mixte (composé de 6 421 individus dont 4 587 inscrits), ce dernier faisant toutefois l'objet d'une pondération (présentée au point 5.1.2).

L'analyse sera ainsi menée en 3 temps :

- régression logistique sur les facteurs manipulés (sans intégration de la préoccupation pour le RVP ni des variables de contrôle), menée sur le fichier de requalification
- même analyse, menée sur le fichier mixte
- régression logistique sur l'ensemble des variables indépendantes et de contrôle, menée sur le fichier mixte

Nous présenterons ci-après les résultats de ces trois types d'analyses, les deux premières étant toutefois comparées au sein d'un même tableau (cf. tableau 12.18).

Le modèle de régression logistique visant à expliquer la probabilité de communiquer les données personnelles demandées (CDP, notée i) par les seuls facteurs manipulés (Familiarité et Contexte uniquement) est, dans les deux cas (fichier de requalification et fichier mixte), significatif (chi-deux = 115,06 ; p = 0,000 et chi-deux = 46,11 ; p = 0,000, respectivement). L'exposition à la sollicitation a donc un impact significatif sur la probabilité de communiquer les données demandées dans les deux fichiers, ce qui renforce la solidité de ce résultat.
En particulier, le contexte dans lequel s'opère la sollicitation (jeu vs. enquête) et la familiarité avec l'entreprise sollicitant les informations (anciens vs. récents) influencent la probabilité d'y répondre favorablement (p = 0,024 et p = 0,000 respectivement, dans le fichier mixte).

Le tableau 12.18 qui suit récapitule ces résultats, selon le fichier utilisé.

Tableau 12.18 Impact des facteurs manipulés sur la communication de données personnelles

	Fichier requalification		Fichier mixte (pondéré)	
Taille	N = 20 805		N = 6 241	
VD : CDP (oui vs. non)	B	Sig.	B	Sig.
FAMILIARITE	- 0,236	**0,000**	- 0,321	**0,000**
CONTEXTE	- 0,185	**0,000**	- 0,131	**0,024**
Constante	0,971	0,000	1,027	0,000
-2LL	28 134,16		8 431,76	
R^2 de Cox & Snell	0,006		0,007	
R^2 de Nagelkerke	0,007		0,010	
% de « bien classés »	56,3%		58,4%	

Comme le confirme le tableau suivant (réalisé sur le fichier mixte), lorsque le contexte de sollicitation concerne le jeu, la probabilité de communiquer des données personnelles est plus élevée que pour l'enquête. De même, lorsque l'individu sollicité s'est récemment inscrit sur le programme Elisa (familiarité faible), la probabilité qu'il communique ses données est plus élevée que pour celui qui est inscrit depuis longtemps.

Tableau 12.19 Tableau croisé Communication des données / Familiarité et Contexte

	Récent	Ancien	Jeu	Enquête	Total
Communication	2 767 (76,2%)	1 820 (69,7%)	1 337 (75,8%)	3 250 (72,6%)	4 587
Non communication	863 (23,8%)	791 (30,3%)	428 (24,2%)	1 226 (27,4%)	1 654
Total	3 630	2 611	1 765	4 476	6 241

Voyons, à présent, si cet effet est confirmé lorsque l'on introduit l'autre variable indépendante (préoccupation RVP) ainsi que les variables de contrôle. Cette analyse ne peut être menée que sur le fichier mixte qui contient, à la différence de l'autre, l'ensemble de ces mesures.

Le modèle de régression logistique visant à expliquer la probabilité de communiquer des données personnelles par les facteurs manipulés (Familiarité et Contexte), la préoccupation pour le RVP (PREOC), tout en contrôlant l'effet des variables exogènes est, là aussi, significatif (chi-deux = 279,63 ; p = 0,000). De plus, l'ajustement global de ce modèle est meilleur que le précédent (avec les seuls facteurs manipulés), ce qui dénote l'apport substantiel de ces variables, introduites par blocs successifs. Pour rappel, seules les variables exogènes significatives sont présentées.

Tableau 12.20 Résultat concernant la probabilité de communiquer des données (CDP)

	Modèle A		Modèle B		Modèle C	
VD : CDP (oui vs. non)	B	Sig.	B	Sig.	B	Sig.
FAMILIARITE	- 0,506	**0,000**	- 0,526	**0,000**	- 0,334	**0,000**
CONTEXTE	- 0,205	**0,021**	- 0,234	**0,009**	- 0,217	**0,019**
PREOC			- 0,057	**0,000**	- 0,040	**0,000**
Habitude de répondre (CHR)					0,097	**0,000**
Habitude de cliquer (CH_vclic)					- 0,448	**0,000**
Sexe (H/F)					0,236	**0,004**
Age					- 0,012	**0,000**
Expérience d'Internet (EXP)					- 0,226	**0,000**
Constante	2,223	0,000	3,164	0,000	1,027	0,000
-2LL	4 074,31		4 027,47		3 846,34	
R^2 de Cox & Snell	0,014		0,026		0,073	
R^2 de Nagelkerke	0,021		0,039		0,108	
% de « bien classés »	75,4%		75,4%		75,8%	

Pour conclure cette première analyse, les résultats confirment l'impact significatif de la familiarité avec l'entreprise sollicitant les données, du contexte de sollicitation et de la préoccupation pour le respect de la vie privée sur la probabilité de communiquer des données personnelles, en contrôlant l'habitude de répondre (de fournir des données), l'habitude de cliquer sur les emails reçus dans le cadre du programme relationnel, le sexe, l'âge et l'expérience Internet du répondant. Plus précisément, il apparaît qu'être peu familier de l'entreprise et être sollicité dans le cadre d'un jeu renforcent la probabilité de communiquer les données personnelles demandées. A l'opposé, la préoccupation pour le respect de la vie privée (PREOC) vient diminuer cette probabilité. Enfin, l'habitude de fournir des informations personnelles aux entreprises (CHR) et le fait d'être une femme accroissent la probabilité de fournir des données personnelles, tandis que l'habitude de cliquer sur les emails envoyés par l'entreprise, l'âge et l'expérience d'Internet la réduisent.

Au final, il apparaît ainsi que ***les hypothèses H1.2i et H1.4i ne sont pas validées*** : il existe bien un impact de la familiarité et du contexte sur la probabilité de communiquer des données personnelles mais dans le sens inverse de celui imaginé.

Par contre, ***l'hypothèse H3i*** préconisant un impact de la préoccupation RVP sur la probabilité de fournir des données personnelles ***est bien validée*** : plus la préoccupation est élevée, moins l'individu accepte de communiquer des données personnelles.

7.3 Résultats concernant la probabilité d'abandonner

Comme dans le cas précédent, nous mènerons les analyses concernant l'effet des facteurs manipulés et de la préoccupation pour le RVP sur la probabilité d'abandonner avant la fin en 3 temps, à travers les deux fichiers de données (requalification puis mixte). Nous présentons dans un premier temps et conjointement, les résultats concernant l'impact des seuls facteurs manipulés (Familiarité, Contexte et Longueur) sur la probabilité d'abandon.

Le modèle de régression logistique visant à expliquer la probabilité d'abandonner avant la fin par les seuls facteurs manipulés (Familiarité, Contexte et Quantité) est, dans les deux cas (fichier de requalification et fichier mixte), significatif (chi-deux = 623,93 ; p = 0,000 et chi-deux = 71,64 ; p = 0,000, respectivement). L'exposition à la sollicitation a donc un impact significatif sur la probabilité d'abandonner dans les deux fichiers.
On observe toutefois des différences entre les deux fichiers en ce qui concerne l'impact respectif de chaque facteur manipulé. Si, dans le premier fichier (dit de requalification), les trois facteurs influencent significativement la probabilité d'abandonner (p=0,046 ; p=0,000 et p=0,000 respectivement), dans le second, seuls les deux derniers ont un impact significatif. En particulier, le contexte dans lequel s'opère la sollicitation (jeu vs. enquête) et la quantité d'informations demandées (faible vs. élevée) influencent la probabilité d'aller (ou non) jusqu'au bout (p =0,013 et p =0,000 respectivement, dans le fichier mixte). Au contraire, la familiarité (faible vs. élevée) ne semble pas avoir d'impact significatif sur cette probabilité, si l'on considère les résultats du second fichier, tandis qu'elle en a un si l'on prend en compte ceux du premier fichier. Il est difficile d'expliquer les raisons d'un tel écart, sauf à considérer le biais engendré par la sélection opérée pour obtenir le fichier mixte, biais déjà évoqué au paragraphe 5.1.2 et censé être redressé par une pondération adéquate.

Le tableau 12.21 qui suit récapitule ces résultats, selon le fichier utilisé.

Tableau 12.21 Impact des facteurs manipulés sur la probabilité d'abandonner

	Fichier requalification		Fichier mixte (pondéré)	
Taille	N = 12 154		N = 4 587	
VD : Abandon (oui vs non)	B	Sig.	B	Sig.
FAMILIARITE	- 0,094	**0,046**	- 0,002	0,988
CONTEXTE	- 0,532	**0,000**	- 0,268	**0,013**
QUANTITE	1,023	**0,000**	0,845	**0,000**
Constante	- 1,993	0,000	- 2,830	0,000
-2LL	11 889,60		2 675,11	
R^2 de Cox & Snell	0,050		0,019	
R^2 de Nagelkerke	0,078		0,037	
% de « bien classés »	73,1%		87,5%	

Comme l'indique le tableau suivant (relatif au fichier mixte), lorsque le contexte de sollicitation concerne le jeu, la probabilité d'abandonner avant la fin est plus faible que pour l'enquête. A l'opposé, lorsque la quantité de données demandées est élevée, la probabilité d'abandonner avant la fin est plus élevée que lorsque cette quantité est faible.

Tableau 12.22 Tableau croisé Abandon / Contexte et Quantité

	Jeu	Enquête	Court	Long	Total
Non abandon	1 144 (85,6%)	2 870 (88,3%)	2 080 (92,0%)	1 934 (83,2%)	4 014
Abandon	193 (14,4%)	380 (11,7%)	182 (8,0%)	391 (16,8%)	573
Total	1 337	3 250	2 262	2 325	4 587

Voyons à présent si ces effets prévus dans le fichier mixte sont confirmés lorsqu'on introduit l'autre variable indépendante (préoccupation RVP) ainsi que les variables de contrôle.

Le modèle de régression logistique visant à expliquer la probabilité d'abandonner par les facteurs manipulés (Familiarité, Contexte et Quantité) et la préoccupation RVP (PREOC), tout en contrôlant l'effet des variables exogènes est, là aussi, significatif (chi-deux = 67,61 ; p = 0,000). De plus, l'ajustement global de ce modèle est meilleur que le précédent (avec les seuls facteurs manipulés), ce qui dénote l'apport substantiel de ces variables, introduites par blocs successifs. Pour rappel, seules les variables exogènes significatives sont présentées.

Tableau 12.23 Résultat concernant la probabilité d'abandonner (Abandon)

	Modèle A		Modèle B		Modèle C	
VD : Abandon (oui vs. non)	B	Sig.	B	Sig.	B	Sig.
FAMILIARITE	0,012	0,946	0,021	0,906	- 0,003	0,987
CONTEXTE	0,259	0,212	0,284	0,171	0,416	**0,049**
QUANTITE	1,055	**0,000**	1,059	**0,000**	1,077	**0,000**
PREOC			0,046	**0,016**	0,027	0,182
Habitude d'abandonner (CH_abandon)					0,111	**0,000**
Niveau d'études					- 0,168	**0,012**
Constante	- 4,962	0,000	- 5,718	0,000	- 5,875	0,000
-2LL	1 037,82		1 031,77		1 005,51	
R^2 de Cox & Snell	0,014		0,017		0,027	
R^2 de Nagelkerke	0,040		0,047		0,077	
% de « bien classés »	94,3%		94,3%		92,2%	

Pour conclure cette seconde analyse, les résultats confirment l'impact significatif du contexte de sollicitation et de la quantité de données demandées sur la probabilité de fournir toutes les informations (vs. d'abandonner avant la fin), en contrôlant l'habitude d'abandonner et le niveau d'études du répondant. Plus précisément, il apparaît que le fait d'être sollicité dans le cadre d'une enquête augmente la probabilité d'abandonner. A l'opposé, le fait d'être sollicité pour un questionnaire court vient diminuer cette probabilité. Enfin, l'habitude d'abandonner lors du remplissage d'un questionnaire (CH_abandon) augmente la probabilité d'abandon tandis qu'avoir un niveau d'études élevé vient la diminuer. A noter que la familiarité avec l'entreprise sollicitant les données n'aurait aucun effet sur la probabilité d'abandonner avant la fin. Enfin, en ce qui concerne la préoccupation pour le RVP, significative dans le modèle B, elle devient non significative quand on contrôle l'habitude d'abandonner et le niveau d'études.

Au final, il apparaît ainsi que ***les hypothèses H1.2j et H3j ne sont pas validées*** : il n'y aurait pas d'impact, ni de la familiarité avec l'entreprise sollicitante, ni de la préoccupation RVP sur la probabilité d'abandonner. De même, ***l'hypothèse H1.4j n'est pas non plus validée*** : il y aurait bien un impact du contexte de sollicitation sur la probabilité d'abandonner, mais dans le sens inverse de celui imaginé. Par contre, ***l'hypothèse H1.3j*** préconisant un impact de la quantité d'informations demandées sur la probabilité d'abandonner ***est bien validée*** : plus la quantité est élevée, plus la probabilité d'abandonner avant la fin augmente.

Pour renforcer (et compléter) ce résultat concernant l'impact de la quantité d'informations demandées sur la probabilité d'abandonner, nous avons représenté graphiquement la courbe d'attrition eu égard à la progression des individus dans le questionnaire (cf. Annexe 12.13 a). Ainsi, si l'on s'arrête à la question 16 (partie commune aux questionnaires court et long), on s'aperçoit que l'attrition est légèrement plus faible pour le questionnaire court. Ceci confirme nos résultats : les individus affectés au questionnaire court ont dans l'ensemble moins abandonné. Toutefois, on constate aussi que l'écart entre les deux courbes est relativement faible : globalement, la quantité d'informations demandées ne modifierait donc que marginalement le taux de réponse. De plus, si l'on prolonge la courbe jusqu'à la question 42, ce qui n'est possible que pour le questionnaire long (cf. Annexe 12.13 b), on s'aperçoit que le nombre d'individus encore présents décroît de manière presque régulière. On perd donc un nombre quasi identique d'individus à chaque question mais, au final, le nombre d'informations récoltées est beaucoup plus important. L'impact managérial de ce résultat est considérable. Il est en effet possible de dire qu'une entreprise a intérêt à proposer un questionnaire long, quitte à perdre beaucoup de personnes au départ (comparativement à un questionnaire court) car elle aura alors la possibilité de récolter plus d'informations au final (notamment auprès de ceux qui accepteront de répondre à de nombreuses questions). La deuxième conclusion que ce résultat amène concerne l'ordre des questions posées. Puisque les personnes affectées à un questionnaire long ont une plus forte probabilité d'abandonner avant la fin, il serait préférable de mettre les questions importantes (celles pour lesquelles l'entreprise souhaite le maximum de réponses possibles) au début, à un moment où le taux de réponse est encore élevé. En les mettant à la fin, l'entreprise courrait en effet le risque qu'un grand nombre d'individus aient déjà abandonné. Ce critère doit cependant être rapproché du caractère sensible (ou non) des questions posées et justifie tout l'intérêt de l'analyse suivante qui concerne l'impact de la sollicitation sur la probabilité de répondre à une question sensible.

7.4 Résultats concernant la probabilité de fournir une donnée sensible

Pour ce test, nous procéderons de la même manière que précédemment, à savoir en 3 étapes successives. Nous comparerons, dans un premier temps, les résultats des analyses menées sur les seuls facteurs manipulés, et ce, sur les deux fichiers disponibles.

Le modèle de régression logistique visant à expliquer la probabilité de fournir une donnée sensible par les seuls facteurs manipulés (Familiarité, Contexte et Quantité) est, dans les deux

cas (fichier de requalification et fichier mixte), significatif (chi-deux = 341,07 ; p = 0,000 et chi-deux = 134,38 ; p = 0,000, respectivement). L'exposition à la sollicitation a donc un impact significatif sur la probabilité de fournir une donnée sensible dans les deux fichiers. En particulier, le contexte dans lequel s'opère la sollicitation (jeu vs. enquête) et la familiarité avec l'entreprise sollicitant les informations (faible vs. élevée) influencent la probabilité d'y répondre favorablement (p = 0,000 et p = 0,048 respectivement, dans le fichier mixte).

Tableau 12.24 Impact des facteurs manipulés sur la communication de données sensibles

	Fichier requalification		Fichier mixte (pondéré)	
Taille	N = 20 805		N = 6 241	
VD : Réponse à une question sensible (oui vs. non)	B	Sig.	B	Sig.
FAMILIARITE	- 0,115	**0,011**	- 0,160	**0,048**
CONTEXTE	- 0,787	**0,000**	- 0,920	**0,000**
QUANTITE	- 0,054	0,215	0,044	0,572
Constante	0,600	0,000	0,788	0,000
-2LL	12 280,60		3 877,85	
R^2 de Cox & Snell	0,032		0,039	
R^2 de Nagelkerke	0,046		0,056	
% de « bien classés »	67,9%		68,2%	

Comme le confirme le tableau suivant (pour le fichier mixte), lorsque le contexte de sollicitation concerne le jeu, la probabilité de fournir une donnée sensible est plus élevée que pour l'enquête. De même, lorsque l'individu sollicité s'est récemment inscrit au programme Elisa (familiarité faible), la probabilité qu'il réponde est plus élevée que pour celui qui est inscrit depuis longtemps.

Tableau 12.25 Tableau croisé Communication de données sensibles / Familiarité et Contexte

	Récent	Ancien	Jeu	Enquête	Total
Non réponse	1 398 (69,3%)	989 (74,2%)	544 (56,9%)	1 843 (77,0%)	2 387
Réponse	618 (30,7%)	343 (25,8%)	412 (43,1%)	549 (23,0%)	961
Total	2 016	1 331	956	2 392	3 348

Voyons, à présent, si cet effet est confirmé lorsqu'on introduit l'autre variable indépendante (préoccupation RVP) et les variables de contrôle, tout en travaillant sur le fichier mixte.

Le modèle de régression logistique visant à expliquer la probabilité de fournir une donnée sensible par les facteurs manipulés (Familiarité, Contexte et Quantité) et la préoccupation RVP (PREOC), tout en contrôlant l'effet des variables exogènes est, là encore, significatif (chi-deux = 217,44 ; p = 0,000). De plus, l'ajustement global de ce modèle est meilleur que le précédent (avec les seuls facteurs manipulés), ce qui dénote l'apport substantiel de ces variables. Pour rappel, seules les variables exogènes significatives sont présentées.

Tableau 12.26 Résultat concernant la probabilité de communiquer des données sensibles (ND)

	Modèle A		Modèle B		Modèle C	
VD : Réponse à une question sensible (oui vs. non)	B	Sig.	B	Sig.	B	Sig.
FAMILIARITE	- 0,125	0,189	- 0,132	0,165	- 0,044	0,662
CONTEXTE	- 0,997	**0,000**	- 1,017	**0,000**	- 1,127	**0,000**
QUANTITE	- 0,024	0,799	- 0,025	0,789	- 0,023	0,805
PREOC			- 0,030	**0,002**	- 0,034	**0,001**
Habitude de donner (CHD)					0,099	**0,002**
Age					- 0,030	**0,000**
CSP					- 0,129	**0,004**
Niveau d'études					- 0,122	**0,001**
Constante	0,971	0,000	1,449	0,000	2,742	0,000
-2LL	2 769,08		2 759,40		2 662,79	
R^2 de Cox & Snell	0,045		0,049		0,087	
R^2 de Nagelkerke	0,065		0,070		0,124	
% de « bien classés »	68,0%		68,2%		71,3%	

Pour conclure cette troisième analyse, les résultats confirment l'impact significatif du contexte de sollicitation et de la préoccupation pour le respect de la vie privée sur la probabilité de fournir une donnée sensible, en contrôlant l'habitude de répondre à toutes les questions posées, l'âge, la CSP et le niveau d'études du répondant. Plus précisément, il apparaît que le fait d'être sollicité dans le cadre d'un jeu renforce la probabilité de fournir l'information sensible demandée. A l'opposé, la préoccupation pour le respect de la vie privée (PREOC) vient diminuer cette probabilité. Enfin, l'habitude de fournir l'ensemble des renseignements demandés quand on complète un questionnaire (CHD) accroît la probabilité de répondre à la question sensible, tandis qu'être âgé, avoir un niveau d'études élevé et faire partie d'une catégorie socioprofessionnelle supérieure la réduisent.

Au final, il apparaît ainsi que ***les hypothèses H1.2k et H1.3k ne sont pas validées*** : il n'y aurait pas d'impact de la familiarité et de la quantité d'informations demandées sur la probabilité de répondre à une question sensible. De même, ***l'hypothèse H1.4k n'est pas non plus validée*** : il y aurait bien un impact du contexte de sollicitation sur la probabilité de fournir une donnée personnelle mais dans le sens contraire à celui imaginé.
Par contre, ***l'hypothèse H3k*** préconisant un impact de la préoccupation pour le RVP sur la probabilité de répondre à une question sensible ***est bien validée*** : plus la préoccupation est élevée, moins l'individu accepte de communiquer cette information.

7.5 Résultats concernant la poursuite de la relation avec l'entreprise

Nous testons à présent l'impact respectif de l'exposition à la sollicitation (et en particulier des deux premiers facteurs manipulés : Familiarité et Contexte) et de la réponse à celle-ci (communication de données, abandon avant la fin) sur l'intention de l'individu de poursuivre la relation avec l'entreprise l'ayant sollicité. La raison d'être de cette hypothèse est de mesurer l'incidence de la sollicitation sur la relation que l'individu entretient avec la firme.

Nous procéderons ici, en deux temps, par le biais de régressions multiples. Nous ne pouvons en effet tester à la fois l'impact de la réponse à la sollicitation (communication ou non des données demandées) et le fait d'avoir poursuivi (ou non) jusqu'au bout (vs. abandon). Nous testerons donc, dans un premier temps, l'impact de la réponse puis dans un second temps, celui de l'abandon. Nous donnons successivement les résultats pour ces deux analyses.

Tableau 12.27 Impact de la réponse sur la probabilité de poursuivre la relation

	Modèle A		Modèle B		Modèle C	
VD : Int. de poursuivre la relation (IPR)	B	Sig.	B	Sig.	B	Sig.
Communication des données (CDP)	1,674	**0,000**	1,666	**0,000**	1,077	**0,000**
FAMILIARITE			- 0,246	**0,044**	0,081	0,444
CONTEXTE			0,460	**0,001**	0,468	**0,000**
Qualité perçue du programme (QUALITE)					0,570	**0,000**
Sexe					0,223	**0,037**
Utilisation d'Internet (UTIL)					- 0,255	**0,000**
Constante	13,044	0,000	12,609	0,000	5,444	0,000
R^2 ajusté	0,037		0,041		0,325	

Le pouvoir explicatif du modèle de régression linéaire visant à expliquer l'intention de poursuivre la relation (IPR) par l'exposition à la sollicitation (Contexte et Familiarité) et par la réponse à celle-ci (communication ou non des données personnelles demandées) (modèle B) est assez faible (R^2 ajusté = 0,041 ; F = 0,000). Néanmoins, l'ensemble de ces variables est significatif et explique donc, au moins en partie, cette intention de poursuivre la relation. En particulier, les individus ayant été soumis à l'enquête, étant inscrits depuis longtemps sur Elisa et ayant accepté de fournir des données personnelles ont une plus forte probabilité de poursuivre la relation. Lorsqu'on intègre les différentes variables de contrôle et, en particulier, la qualité perçue du programme, le pouvoir explicatif du modèle augmente considérablement pour atteindre un niveau tout à fait acceptable (R^2 ajusté = 0,325 ; F = 0,000). Ce modèle expliquerait donc plus de 30% de l'intention de poursuivre la relation. Par contre, en ce qui concerne les variables explicatives, on constate que la familiarité, significative en l'absence de variables de contrôle, ne l'est plus une fois celles-ci introduites (t = 0,765 ; p = 0,444), notamment sous l'effet de la qualité perçue du programme (t = 38,568 ; p = 0,000). Le contexte de sollicitation reste quant à lui significatif (t = 4,121 ; p =0,000), tout comme le fait d'avoir répondu favorablement à la sollicitation reçue précédemment (t = 9,001 ; p = 0,000). Deux variables individuelles s'avèrent également significatives, même si elles n'améliorent pas le pouvoir explicatif du modèle : il s'agit du sexe d'une part (t = 2,089 ; p = 0,037) et de l'utilisation d'Internet (mesurée par la fréquence de consultation de la messagerie) d'autre part (t =-4,380 ; p=0,000). Plus particulièrement, les hommes et les personnes consultant leur boîte email peu souvent ont une plus faible probabilité de poursuivre la relation avec l'entreprise les ayant sollicités pour fournir des données à caractère personnel.

Tableau 12.28 Impact de l'abandon sur la probabilité de poursuivre la relation

	Modèle A		Modèle B		Modèle C	
VD : Int. de poursuivre la relation (IPR)	B	Sig.	B	Sig.	B	Sig.
Abandon	- 0,359	0,265	- 0,377	0,240	- 0,330	0,220
FAMILIARITE			- 0,220	0,147	0,101	0,444
CONTEXTE			0,423	**0,011**	0,463	**0,001**
Qualité perçue du programme (QUALITE)					0,550	**0,000**
Sexe					0,277	**0,039**
Utilisation d'Internet (UTIL)					- 0,299	**0,000**
Constante	14,767	0,000	14,352	0,000	7,577	0,000
R^2 ajusté	0,000		0,003		0,303	

Le pouvoir explicatif du modèle de régression linéaire visant à expliquer l'intention de poursuivre la relation (IPR) par l'exposition à la sollicitation (Contexte et Familiarité) et par le fait d'avoir ou non répondu à l'ensemble des questions posées (modèle B) est très faible (R^2 ajusté = 0,003 ; F = 0,025). Parmi ces variables, seul le contexte de sollicitation s'avère significatif et explique donc, au moins en partie, l'intention de poursuivre la relation. En particulier, les individus ayant été soumis à l'enquête ont une plus forte probabilité de poursuivre la relation que ceux ayant été soumis au jeu-concours. Ceci confirme l'idée selon laquelle solliciter des clients sous une forme à laquelle ils adhèrent peu (ex : le jeu) pourrait les décourager à poursuivre la relation avec cette entreprise. Ni être familier avec l'entreprise, ni avoir abandonné avant la fin n'influenceraient par contre la poursuite de la relation.

Lorsqu'on intègre les différentes variables de contrôle et, en particulier, la qualité perçue du programme, le pouvoir explicatif du modèle augmente là encore, pour atteindre un niveau équivalent au précédent (R^2 ajusté = 0,303 ; F = 0,000). Ce modèle expliquerait donc environ 30% de l'intention de poursuivre la relation. En ce qui concerne les variables explicatives, on constate que le contexte de sollicitation reste significatif (t=3,333 ; p =0,000), tout comme la qualité perçue du programme (t =30,703 ; p = 0,000). Les deux variables individuelles identifiées dans le modèle précédent (sexe et utilisation d'Internet), s'avèrent là encore significatives, et leur influence va dans le même sens que précédemment.

Au final, il apparaît que les ***hypothèses H1.2h et H6j ne sont pas validées*** : il n'y aurait donc pas d'impact de la familiarité ni du fait d'avoir abandonné avant la fin sur la probabilité de poursuivre la relation avec l'entreprise ayant sollicité les données. A l'opposé, ***l'hypothèse H1.4h est bien validée*** : il y aurait effectivement un impact du contexte de sollicitation sur la probabilité de poursuivre la relation, dans le sens imaginé : les individus soumis à l'enquête auraient plus de chance de poursuivre la relation avec l'entreprise les ayant sollicités.

De même, ***l'hypothèse H6i*** préconisant un impact de la réponse à la sollicitation sur la probabilité de poursuivre la relation ***est validée*** : les individus ayant accepté de fournir leurs données personnelles ont plus de chances de continuer la relation avec l'entreprise les ayant sollicités. Comme suggéré dans le chapitre 7, nous avons ici la confirmation du caractère « engagé » de la réponse à une sollicitation de données personnelles : fournir de telles informations est une forme d'engagement qui montre la volonté du consommateur de continuer à maintenir une relation durable avec cette entreprise.

Testons, à présent, le rôle médiateur des variables de réponse.

7.6 Test du rôle médiateur de la réponse à la sollicitation

Pour rappel, sans que nous ayons formulé d'hypothèses spécifiques à ce sujet, nous avons remarqué que les variables de réponse (Communication de données personnelles et Abandon) semblent se positionner en tant que médiateurs de l'effet de l'exposition à la sollicitation sur l'intention de poursuivre la relation. Nous avons identifié au paragraphe 6.3 les analyses à réaliser et les conditions à vérifier pour tester l'existence d'un tel effet. Concrètement, nous vérifierons l'effet médiateur de chacune des variables de réponse (CDP et Abandon) séparément. Vous trouverez dans le tableau ci-dessous les résultats de ces analyses.

Tableau 12.29 Test de l'effet médiateur des variables de réponse (CDP et Abandon)

Conditions à valider	Résultats	
Variables indépendantes (VI) et dépendantes (VD) concernées	VI : Familiarité avec l'entreprise et Contexte de sollicitation VD : Intention de poursuivre la relation (IPR)	
Variables médiatrices testées	Communication de données personnelles (CDP)	Abandon
Condition 1 : lien entre la variable indépendante et la variable dépendante	Non validée pour la familiarité (B = - 0,138 ; p = 0,267) ***Validée pour le contexte*** (B = 0,443 ; p = 0,001)	
Condition 2 : lien entre la variable indépendante et la variable médiatrice	***Validée pour la familiarité*** (B = - 0,334 ; p = 0,000) ***Validée pour le contexte*** (B = - 0,217 ; p = 0,019) (cf. tableau 12.20)	Non validée pour la familiarité (B = - 0,002 ; p = 0,988) ***Validée pour le contexte*** (B = - 0,268 ; p = 0,013) (cf. tableau 12.21)
Condition 3 : lien entre la variable médiatrice et la variable dépendante	***Validée*** (B = 1,674 ; p = 0,000) (cf. tableau 12.27)	Non validée (B = - 0,359 ; p = 0,265) (cf. tableau 12.28)
Condition 4 : l'effet de la variable indépendante est nul quand la variable médiatrice est contrôlée	***Validée pour la familiarité*** (B = 0,081 ; p = 0,444) Non validée pour le contexte (B = 0,468 ; p = 0,000) (cf. tableau 12.27)	***Validée pour la familiarité*** (B = 0,101 ; p = 0,444) Non validée pour le contexte (B = 0,463 ; p = 0,001) (cf. tableau 12.28)
Conclusions	Médiateur partiel	Pas médiateur

Pour conclure sur ce point, on peut donc dire que la communication de données personnelles (CDP) est un médiateur partiel de l'effet de l'exposition à la sollicitation sur l'intention de poursuivre la relation (IPR). L'abandon ne jouerait, quant à lui, aucun rôle médiateur, d'autant qu'il n'aurait pas d'effet direct sur l'intention de poursuivre la relation.

Nous en avons fini avec le test des hypothèses de recherches relatives à l'expérience 3. Nous proposons de récapituler les conclusions auxquelles nous avons abouties. Nous reviendrons sur les explications possibles des résultats obtenus dans la discussion finale de ce document.

8 Récapitulatif des résultats

Nous avons étudié dans ce chapitre l'effet de l'exposition à une sollicitation de données personnelles et de la préoccupation pour le respect de la vie privée sur le comportement de réponse de l'individu (communication des données demandées, abandon avant la fin, réponse aux questions sensibles) et sur son intention comportementale de poursuivre la relation avec l'entreprise l'ayant sollicité.

Au regard des résultats, la familiarité avec l'entreprise sollicitant les informations n'aurait qu'un effet très partiel sur le comportement et les intentions comportementales de l'individu :

- la familiarité aurait bien un effet sur la probabilité de communiquer les données, mais dans le sens inverse de celui attendu : les individus récemment inscrits au programme auraient davantage répondu que les plus anciens (***H1.2 i non validée***) ;
- par contre, la familiarité n'aurait aucun effet, ni sur la probabilité d'abandonner avant la fin (pour ceux ayant commencé à répondre) ni sur la probabilité de répondre à une question sensible (***H1.2 j et k non validées***) ;
- la familiarité avec l'entreprise sollicitante n'aurait également aucun effet sur l'intention de poursuivre la relation, une fois contrôlée la qualité perçue de la relation (ici, du programme relationnel) (***H1.2 h non validée***).

Sans nous avancer trop dans la discussion qui interviendra en fin de document, nous pouvons néanmoins avancer quelques explications possibles à ces résultats. Tout d'abord, le fait que la familiarité influence la probabilité de fournir des données, dans le sens contraire à celui attendu (i.e. les plus anciens sont plus réticents) peut finalement s'expliquer facilement. En effet, les individus anciennement inscrits sur Elisa ont déjà eu, à plusieurs reprises, l'occasion de fournir des données personnelles. Il n'est donc pas sûr qu'ils aient vu l'intérêt de le faire une nouvelle fois. A l'opposé, les individus qui viennent de s'inscrire sont forcément très motivés et très impliqués par cette démarche. Il est donc probable qu'ils sont très demandeurs d'opérations leur proposant des bénéfices substantiels en échange, qu'ils soient monétaires (cas du jeu-concours) ou non (cas de l'enquête).

La familiarité aurait donc une influence (même dans le sens contraire à celui attendu) sur la probabilité de communiquer des données, c'est-à-dire sur la décision initiale de l'individu. Elle n'en aurait par contre aucune ni sur la probabilité d'abandonner ni sur celle de fournir une donnée sensible, qui concernent des décisions ultérieures, c'est-à-dire une fois le « processus » enclenché. Dans ce cas, des éléments plus « matériels », tels que le design du questionnaire, ont probablement un impact plus important que les liens créés avec l'entreprise qui sollicite les données. Enfin, le fait que la familiarité n'ait pas d'impact sur la poursuite de la relation serait lié à l'influence prioritaire de la qualité perçue du programme, qui supplante alors l'effet de l'ancienneté de la relation.

Le contexte dans lequel les informations sont sollicitées et le type d'avantages que l'individu peut retirer en échange aurait un impact majeur sur la réponse de l'individu à la sollicitation et sur son intention de poursuivre la relation avec l'entreprise l'ayant sollicité, puisque :

- le contexte influencerait de manière significative la probabilité de communiquer les informations demandées, d'aller jusqu'au bout du questionnaire et de répondre à une question sensible. Toutefois, cette influence irait, à chaque fois, dans le sens inverse de celui imaginé : les individus sollicités par le biais du jeu-concours aurait systématiquement un comportement plus favorable que ceux sollicités pour une enquête (***H1.4 i, j et k non validées***) ;
- de la même manière, le contexte influencerait de manière significative l'intention de poursuivre la relation avec l'entreprise sollicitée, mais cette fois-ci dans le sens supposé : les individus sollicités pour l'enquête auraient une plus forte intention de poursuivre la relation (***H1.4 h validée***).

Le fait que le contexte influence de manière significative les comportements de réponse mais dans le sens inverse à celui attendu (i.e. le jeu-concours est plus favorable) nous semble pouvoir être lié à la manipulation de ce facteur qui, dans le cas du jeu-concours incitait davantage à répondre. Nous rappelons en effet que les participants au jeu maximisaient leurs chances de gagner en répondant à toutes les questions, ce qui n'était pas le cas pour l'enquête. Toutefois, l'opération « jeu-concours » aurait un effet plus éphémère que celle consistant à mieux connaître ses clients (cas de l'enquête), ce qui pourrait être lié à la nature même du bénéfice obtenu (incertain dans le premier cas et certain dans le second).

La quantité de données demandées aurait un impact partiel sur la réponse à la sollicitation :

- la quantité influencerait de manière significative la probabilité d'abandon : plus le nombre d'informations à fournir est important, plus la probabilité que l'individu ne poursuive pas le questionnaire est élevée (***H1.3 j validée***) ;
- la quantité n'aurait par contre aucun impact sur la probabilité de fournir une donnée sensible (***H1.3 k non validée***).

Si le premier résultat confirme ce que la littérature prévoyait, le second indiquerait qu'une donnée sensible le reste quelque soit le nombre de questions posées.

Par ailleurs, la préoccupation pour le respect de la vie privée aurait un impact non négligeable sur le comportement de réponse face à une sollicitation de données personnelles :

- la préoccupation pour le RVP influencerait de manière significative à la fois la probabilité d'accepter de communiquer les informations demandées et la probabilité de fournir une donnée sensible (***H3 i et k validées***) ;
- la préoccupation pour le RVP n'aurait par contre aucun impact sur la probabilité d'abandonner avant la fin, une fois le questionnaire commencé (***H3 j non validée***).

Les individus les plus préoccupés par le respect de leur vie privée préféreraient donc, soit ne pas participer, soit s'abstenir de répondre en cas de question délicate (sensible). Le fait que cela n'ait pas eu d'incidence sur la décision d'abandonner s'expliquerait, selon nous, par la possibilité de passer à chaque fois qu'une question délicate se présentait, ce qui représente une solution facile de « repli ».

En ce qui concerne l'impact de la réponse à la sollicitation sur l'intention de poursuivre la relation avec l'entreprise sollicitante, l'effet est partiellement validé :

- l'acceptation de communiquer les données a un impact sur la poursuite de la relation (***H6i validée***). De plus, elle serait un médiateur (même partiel) de l'effet de l'exposition à la sollicitation (Familiarité et Contexte) sur la poursuite de la relation ;
- le fait d'avoir abandonné ou non avant la fin n'aurait, par contre, aucun impact sur l'intention de poursuivre la relation (***H6 j non validée***).

Le fait d'avoir accepté de participer serait donc bien considéré comme une forme d'engagement ayant une incidence sur la suite de la relation. Le fait que l'abandon puisse s'expliquer par d'autres raisons que celles étudiées ici (comme le manque de temps par exemple) justifierait selon nous, au moins en partie, son absence d'influence.

Le tableau et la figure qui suivent récapitulent l'ensemble des relations validées.

Tableau 12.30 Récapitulatif des hypothèses à tester et des résultats obtenus (Expérience 3)

	Hypothèses testées dans l'expérience 3	Résultats
H 1.2 (i, j, k) (17,18,19)	Plus (vs. moins) l'individu est familier avec l'entreprise qui le sollicite : - plus (vs. moins) la probabilité qu'il accepte de communiquer des données personnelles est élevée (vs. faible) (i) ; - plus la probabilité qu'il abandonne avant la fin est faible (vs. élevée) (j) ; - plus la probabilité qu'il réponde à une question sensible est élevée (vs. faible) (k).	Non validée (p = 0,000)* Non validée (p = 0,987) Non validée (p = 0,662)
H 1.4 (i, j, k) (20,21,22)	La probabilité que l'individu accepte de communiquer des données personnelles (i), aille jusqu'au bout (vs. abandonne) (j) et réponde à une question sensible (k) diffère selon le contexte dans lequel s'opère la sollicitation et les avantages qu'il peut obtenir en échange. Elle serait plus forte pour l'enquête comparativement au jeu-concours.	Non validée (p = 0,019)* Non validée (p = 0,049)* Non validée (p = 0,000)*
H 1.3 (j, k) (23,24)	Plus la quantité d'informations demandées est élevée (vs. faible), plus la probabilité que l'individu abandonne avant la fin est élevée (vs. faible) (j) et plus la probabilité qu'il réponde à une question sensible est faible (vs. élevée) (k).	**Validée** **(p = 0,000)** Non validée (p = 0,805)
H 3 (i, j, k) (25,26,27)	Plus la préoccupation pour le respect de la vie privée est élevée (vs. faible), plus la probabilité que l'individu accepte de communiquer des données personnelles est faible (vs. élevée) (i) ; plus la probabilité qu'il abandonne avant la fin est élevée (vs. faible) (j) et plus la probabilité qu'il refuse de répondre aux questions sensibles est élevée (vs. faible) (k).	**Validée** **(p = 0,000)** Non validée (p = 0,182) **Validée** **(p = 0,001)**
H 6 (i, j) (28,29)	L'individu ayant accepté de communiquer des données personnelles (vs. n'ayant pas accepté) (i) et n'ayant pas abandonné avant la fin (vs. ayant abandonné) (j) a une plus forte (vs. faible) probabilité de poursuivre la relation avec l'entreprise qui l'a sollicité.	**Validée** **(p = 0,000)** Non validée (p = 0,220)
H 1.2 (h) (30)	Plus (vs. moins) l'individu est familier avec l'entreprise qui le sollicite, plus la probabilité qu'il poursuive la relation avec elle est élevée (vs. faible).	Non validée (p = 0,444)
H 1.4 (h) (31)	L'intention de poursuivre la relation avec l'entreprise diffère selon le contexte dans lequel s'est opérée la sollicitation et les avantages que l'individu a pu obtenir en échange. Elle serait plus élevée pour l'enquête, comparativement au jeu-concours.	**Validée** **(p = 0,001)**

(*) Ces hypothèses ne sont pas validées dans le sens où l'effet est contraire à celui prévu. L'impact du facteur testé est cependant démontré.

Figure 12.6 Récapitulatif des hypothèses validées dans le cadre de l'expérience 3

N.B. : Pour des raisons de lisibilité, les hypothèses non validées ne sont pas représentées sur cette figure.

Conclusion du chapitre 12

Ce chapitre a présenté les résultats d'une expérimentation visant à tester l'effet (direct) de l'exposition à une sollicitation de données personnelles sur la réponse du consommateur et sur son intention de poursuivre la relation avec l'entreprise l'ayant sollicité.

Après avoir indiqué les objectifs de l'expérimentation, nous avons présenté les facteurs manipulés, les variables contrôlées et la manière dont l'expérience s'est déroulée. Des hypothèses spécifiques à l'étude ont ensuite été développées, conduisant à la formulation du modèle à tester. Du fait de la lourdeur du dispositif et de la taille de l'échantillon, le recueil des données a fait l'objet d'un développement approfondi, afin de présenter la manière dont elles seraient utilisées. La présentation des caractéristiques des répondants a ensuite permis d'insister sur les biais limités dans le cadre de cette expérience ainsi que sur ceux engendrés et sur la manière dont on pouvait y répondre. En particulier, une pondération du fichier a été nécessaire, afin de redresser le biais de sélection occasionné par l'utilisation du post-test.

Les analyses préalables ont permis de souligner la validité du plan d'expérience, notamment du fait du contrôle des variables exogènes et de l'indépendance (bien que parfois relative) des observations. Après avoir établi la qualité psychométrique des instruments de mesure, nous avons présenté le plan de traitement pressenti pour mener à bien nos analyses. Nous avons ensuite vérifié l'adéquation des données aux tests statistiques envisagés, dans le cadre d'une régression logistique ou multiple. Ayant démontré la possibilité de tester les hypothèses dans le respect des conditions exigées, nous avons procédé à l'analyse des résultats.

Les résultats obtenus confirment l'hypothèse générale d'un effet (direct) de l'exposition à la sollicitation sur le comportement de réponse du consommateur. Cet effet est cependant lié au type de facteur étudié. En particulier, le contexte dans lequel les informations sont sollicitées, aurait un effet robuste, tandis que la familiarité et la quantité auraient un effet plus mitigé, lié au type de comportement étudié. La préoccupation pour le respect de la vie privée apparaît comme un antécédent de la réponse du consommateur, sans influencer toutefois l'intention d'abandonner. Enfin, la réponse à la sollicitation aurait aussi un effet majeur sur l'intention de poursuivre la relation avec l'entreprise sollicitante, tout comme le contexte dans lequel ces informations sont demandées.

CONCLUSION DE LA PARTIE III

La question empirique à laquelle nous souhaitions répondre dans cette troisième et dernière partie était la suivante :

Les données obtenues permettent-elles de valider les hypothèses proposées ?

Les analyses réalisées (i.e. analyses de variance, régressions linéaires et logistiques), dans chacune des expériences menées, ont permis de valider, en grande partie, nos hypothèses de recherche relatives à l'impact de la sollicitation de données sur les différentes variables du processus de réponse (i.e. attitude face à la communication des données, intentions comportementales de réponse et réponse à la sollicitation).

En particulier, l'influence de la politique de respect de la vie privée choisie par l'entreprise et celle de la quantité de données demandées sont, dans l'ensemble, établies. A l'opposé, l'impact de la familiarité avec l'entreprise sollicitant les données et du contexte de sollicitation est plus délicat à interpréter, tant les résultats obtenus sont surprenants. Ils feront ainsi l'objet d'une discussion approfondie dans la dernière partie de ce document. L'effet des variables externes de la recherche, introduites comme co-variables dans les analyses, s'avère relativement limité. Seul le niveau d'implication vis-à-vis de la catégorie de produits/services concernée semble susceptible d'influencer, de façon non négligeable, certaines variables du modèle. Enfin, le processus de réponse en lui-même est globalement validé.

La prochaine question qui se pose alors est donc la suivante :

Quelles sont les conclusions auxquels ces résultats nous conduisent ?

Dans la discussion qui suit, des conclusions théoriques et appliquées seront ainsi formulées, tandis que les limites et voies de recherche futures seront identifiées.

DISCUSSION
ET
MISE EN PERSPECTIVE DES RÉSULTATS

« Internet risque de tuer la vie privée. »

Ktitareff (Les Echos du 04/09/2006)

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

Introduction de la discussion

Les chapitres précédents (notamment les chapitres 10, 11 et 12 de la partie III) ont permis de présenter les résultats du test des hypothèses de la recherche. Ce « chapitre final »[319] propose de discuter ces résultats et de les mettre en perspective afin d'en déduire les principaux enseignements, les limites et voies de recherche.

La première section sera consacrée à une discussion des principales conclusions auxquelles cette recherche a permis d'aboutir. Nous ferons ainsi un récapitulatif de l'ensemble des résultats obtenus dans les différentes études menées à l'occasion de ce travail doctoral, et notamment des trois expériences issues du plan expérimental. En particulier, nous interprèterons ces résultats au regard de la littérature sur le respect de la vie privée et le dévoilement de soi, d'une part, et sous l'angle des travaux relatifs au comportement du consommateur face à un questionnaire d'enquête, d'autre part.

Nous discuterons, dans la deuxième section de ce chapitre, des apports de la recherche. Ces apports sont, tout d'abord, théoriques puisque ce travail contribue à améliorer la compréhension du processus de décision du consommateur lorsqu'il se trouve confronté à une sollicitation de ses données personnelles. L'apport empirique réside dans la réalisation d'une étude auprès de consommateurs français, quand toute la littérature sur le sujet est anglo-saxonne. Les apports managériaux sont directement issus d'une meilleure connaissance des effets des facteurs situationnels sur le comportement du consommateur. Enfin, nous présenterons les contributions méthodologiques de ce travail qui sont principalement liées à la mise en œuvre d'un plan expérimental original.

Dans la troisième section de ce chapitre, nous identifierons les limites liées à la validité interne et à la validité externe de la recherche afin de présenter les voies de recherches qui permettraient d'améliorer ce travail et de l'étendre.

[319] Pour des raisons, non de superstition, mais d'équilibre (conserver un nombre de chapitres pair), nous donnons à ce « chapitre final » (qui aurait donc pu être le numéro 13) un statut de « conclusion », plutôt que de « chapitre ».

1 Mise en perspective des résultats de la recherche

Pour rappel, deux études exploratoires et trois expériences successives, soit cinq études au total, ont été menées dans le cadre de ce travail doctoral. Dans cette dernière partie, nous focalisons toutefois l'essentiel de la discussion sur les trois études issues du plan expérimental, dans la mesure où ce sont celles qui ont donné lieu au test des hypothèses de recherche. Nous fournissons à cet égard, dans le tableau ci-après, un récapitulatif de l'ensemble des résultats trouvés au cours de ces trois expériences (cf. tableau D.1 parties 1 et 2).
Cette recherche ayant aussi une visée cumulative, dans le sens où elle permet de tester (confirmer ou infirmer) les conclusions issues des travaux antérieurs, nous comparons également nos résultats avec ceux trouvés dans la littérature.

Les principaux résultats de la recherche permettent, tout d'abord, de discuter de l'existence d'un processus de décision relatif à la réponse du consommateur à une sollicitation de ses données personnelles (1.1). Ensuite, ils rendent compte des effets des variables situationnelles (i.e. de la situation de collecte) sur ce processus de réponse et donc sur le comportement du consommateur (1.2). Enfin, plusieurs enseignements intéressants relatifs à l'effet des variables individuelles ressortent également des analyses (1.3).

1.1 Le processus de décision face à une sollicitation de données personnelles

Nous avons postulé théoriquement, en nous appuyant notamment sur la théorie de l'action raisonnée (cf. chapitre 7), qu'une sollicitation de données personnelles, parce qu'elle fait l'objet d'une évaluation (formation de croyances), conduit les consommateurs à former une attitude (favorable ou défavorable) relative à la communication de ce type de renseignements. Cette tendance entraîne alors des intentions comportementales de réponse, censées se traduire, en cas de sollicitation effective, par un comportement allant dans ce sens.

Nos résultats permettent de discuter de l'existence d'un tel processus et d'en expliquer le mode de fonctionnement. Nous reviendrons surtout sur les phases centrales[320] de ce processus à savoir l'impact de l'évaluation sur l'attitude et l'effet de celle-ci sur les intentions comportementales.

[320] « Centrales » signifiant à la fois importantes et situées au cœur du processus.

Tableau D.1 (partie 1) Récapitulatif des résultats de la recherche concernant la validation du processus de réponse

Hyp.	Variable indépendante	Variable dépendante	Exp. n°	Méthodologie (type d'analyse)	Résultats		Commentaires et Explications
Phase 2 du processus : impact de l'évaluation de la sollicitation sur l'attitude face à la communication des données							
2 a (7)	Confidentialité perçue (CONFID)	Attitude (ATT, e)	1	REGRESSION (tableau 10.17)	**V**	**B = 0,192 p = 0,054**	**L'attitude face à la communication des données est plus favorable quand la confidentialité perçue des données est élevée (vs. faible)**
2 b (8)	Sensibilité perçue (SENS)		1		**V**	**B = - 0,308 p = 0,001**	**L'attitude face à la communication des données est plus favorable quand la sensibilité perçue des données est faible (vs. élevée)**
2 c (9)	Pertinence perçue (PERT)		1		**V**	**B = 0,346 p = 0,002**	**L'attitude face à la communication des données est plus favorable quand la pertinence perçue des données est élevée (vs. faible)**
2 d (13)	Valeur perçue (VAL)	Attitude (ATT, e)	2	REGRESSION (tableau 11.16)	**V**	**B = 0,336 p = 0,000**	**L'attitude face à la communication des données est plus favorable quand la valeur perçue à répondre est élevée (vs. faible)**
HH *		Intention de répondre (IR, f)	2	REGRESSION (tableau 11.20)	**V**	**p = 0,000**	**L'intention de répondre est supérieure quand la valeur perçue à répondre est élevée (vs. faible)**
Phase 3 du processus : impact de la préoccupation pour le respect de la vie privée sur l'attitude face à la communication des données							
3 e (10)	Préoccupation RVP (PREOC)	Attitude (ATT, e)	1	REGRESSION (tableau 10.17)	**V**	**B = - 0,233 p = 0,000**	**L'attitude face à la communication des données est plus favorable quand la préoccupation pour le RVP est faible (vs. élevée)**
			2	REGRESSION (tableau 11.16)	**V**	**B = - 0,138 p = 0,014**	
3 i (25)		Réponse (CDP, i)	3	REGR. LOG. (tableau 12.20)	**V**	**B = - 0,040 p = 0,000**	**La probabilité de répondre est supérieure quand la préoccupation pour le RVP est faible (vs. faible)**
3 j (26)		Réponse (Abandon, j)	3	REGR. LOG. (tableau 12.23)	NV	B = 0,027 p = 0,182	La préoccupation pour le RVP n'a pas d'effet significatif sur la probabilité d'abandonner avant la fin
3 k (27)		Réponse (ND, k)	3	REGR. LOG. (tableau 12.26)	**V**	**B = - 0,034 p = 0,001**	**La probabilité de répondre à une question sensible est supérieure quand la préoccupation pour le RVP est faible (vs. élevée)**
Phase 4 du processus : impact de l'attitude face à la communication des données sur les intentions comportementales							
4 f (14)	Attitude (ATT)	Intention de répondre (IR, f)	2	REGRESSION (tableau 11.17)	**V**	**B = 0,149 p = 0,000**	**L'intention de répondre est supérieure quand l'attitude face à la communication des données est favorable (vs. défavorable)**
4 g (15)		Intention de mentir (IM, g)	2	REGRESSION (tableau 11.18)	**V**	**B = - 0,046 p = 0,000**	**L'intention de mentir est inférieure quand l'attitude face à la communication des données est favorable (vs. défavorable)**
HH *		Intention de poursuivre la relation (IPR, h)	2	REGRESSION (tableau 11.20)	**V**	**p = 0,000**	**L'intention de poursuivre la relation est supérieure quand l'attitude face à la communication des données est favorable (vs. défavorable)**
Phase 5 du processus : impact de l'intention de répondre sur l'intention de poursuivre la relation							
5 (16)	Intention de répondre (IR)	Intention de poursuivre la relation (IPR, h)	2	REGRESSION (tableau 11.19)	**V**	**B = 1,079 p = 0,000**	**L'intention de poursuivre la relation est supérieure quand l'intention de répondre est élevée (vs. faible)**
Phase 6 du processus : impact de la réponse sur l'intention de poursuivre la relation							
6 i (28)	Réponse (CDP, i)	Intention de poursuivre la relation (IPR, h)	3	REGRESSION (tableau 12.27)	**V**	**B = 1,077 p = 0,000**	**L'intention de poursuivre la relation est supérieure quand l'individu a accepté (vs. n'a pas accepté) de répondre à la sollicitation**
6 j (29)	Réponse (Abandon, j)		3	REGRESSION (tableau 12.28)	NV	B = - 0,330 p = 0,220	Le fait d'avoir abandonné (ou non) avant la fin n'a pas d'effet significatif sur l'intention de poursuivre la relation

(*) HH signifie hors hypothèse

Tableau D.1 (partie 2) Récapitulatif des résultats de la recherche concernant l'impact des facteurs situationnels (phase 1 du processus de réponse)

Hyp.	Variable indépendante	Variable dépendante	Exp. n°	Méthodologie (type d'analyse)	Résultats		Commentaires et Explications
1.1 a (1)	Mentions RVP (facteur 1)	Confidentialité perçue (CONFID, a)	1	MANCOVA (tableau 10.16)	**V**	**p = 0,011**	**La confidentialité perçue des données sollicitées est supérieure quand les mentions de RVP sont nombreuses (vs. limitées)**
1.1 b (2)		Sensibilité perçue (SENS, b)	1		**V**	**p = 0,064**	**La sensibilité perçue des données sollicitées est supérieure quand les mentions de RVP sont limitées (vs. nombreuses)**
1.2 a (3)	Familiarité (facteur 2)	Confidentialité perçue (CONFID, a)	1	MANCOVA (tableau 10.16)	NV	p = 0,179	La familiarité avec l'entreprise n'a pas d'effet significatif sur la confidentialité perçue des données sollicitées
1.2 c (4)		Pertinence perçue (PERT, c)	1		NV	p = 0,651	La familiarité avec l'entreprise n'a pas d'effet significatif sur la pertinence perçue des données sollicitées
1.2 i (17)		Réponse (CDP, i)	3	REGR. LOG. (tableau 12.20)	NV*	B = - 0,334 p = 0,000	L'effet est contraire à celui attendu : les individus les moins familiers à l'entreprise ont une plus forte probabilité de répondre
1.2 j (18)		Réponse (Abandon, j)	3	REGR. LOG. (tableau 12.23)	NV	B = - 0,003 p = 0,987	La familiarité avec l'entreprise n'a pas d'effet significatif sur la probabilité d'abandonner le questionnaire avant la fin
1.2 k (19)		Réponse (ND, k)	3	REGR. LOG. (tableau 12.26)	NV	B = - 0,044 p = 0,662	La familiarité avec l'entreprise n'a pas d'effet significatif sur la probabilité de répondre à une question sensible
1.2 h (30)		Intention de poursuivre la relation (IPR, h)	3	REGRESSION (tableau 12.27)	NV	B = 0,081 p = 0,444	La familiarité n'a pas d'effet significatif sur l'intention de poursuivre la relation quand on contrôle la qualité perçue de la relation
1.3 b (5)	Quantité (facteur 3)	Sensibilité perçue (SENS, b)	1	MANCOVA (tableau 10.16)	**V**	**p = 0,000**	**La sensibilité perçue des données est supérieure quand leur quantité demandée est élevée (vs. faible)**
1.3 c (6)		Pertinence perçue (PERT, c)	1		**V**	**p = 0,002**	**La pertinence perçue des données est supérieure quand leur quantité demandée est faible (vs. élevée)**
1.3 d (11)		Valeur perçue (VAL, d)	2	ANCOVA (tableau 11.15)	**V**	**p = 0,076**	**La valeur perçue à répondre est supérieure quand la quantité de données demandées est faible (vs. élevée)**
1.3 j (23)		Réponse (Abandon, j)	3	REGR. LOG. (tableau 12.23)	**V**	**B = 1,077 p = 0,000**	**La probabilité que l'individu abandonne avant la fin est supérieure quand la quantité de données demandées est élevée (vs. faible)**
1.3 k (24)		Réponse (ND, k)	3	REGR. LOG. (tableau 12.26)	NV	B = - 0,023 p = 0,805	La quantité de données demandées n'a pas d'effet significatif sur la probabilité de répondre à une question sensible
1.4 d (12)	Contexte (facteur 4)	Valeur perçue (VAL, d)	2	ANCOVA (tableau 11.15)	NV	p = 0,479	Le contexte dans lequel les informations sont sollicitées n'a pas d'effet significatif sur la valeur perçue à répondre
1.4 i (20)		Réponse (CDP, i)	3	REGR. LOG. (tableau 12.20)	NV*	B = - 0,217 p = 0,019	L'effet est contraire à celui attendu : les individus sollicités pour un jeu-concours (vs. une enquête) ont une plus forte probabilité de répondre
1.4 j (21)		Réponse (Abandon, j)	3	REGR. LOG. (tableau 12.23)	NV*	B = 0,416 p = 0,049	L'effet est contraire à celui attendu : les individus sollicités pour un jeu-concours (vs. une enquête) ont une plus forte probabilité d'aller au bout
1.4 k (22)		Réponse (ND, k)	3	REGR. LOG. (tableau 12.26)	NV*	B = - 1,127 p = 0,000	L'effet est contraire à celui attendu : les individus sollicités pour un jeu-concours ont une plus forte probabilité de répondre à une question sensible
1.4 h (31)		Intention de poursuivre la relation (IPR, h)	3	REGRESSION (tableau 12.27)	**V**	**B = 0,468 p = 0,000**	**L'intention de poursuivre la relation avec l'entreprise est plus élevée quand la sollicitation concerne une enquête (vs. un jeu-concours)**

(*) dont l'effet est contraire à celui supposé

La phase précédente (effet de l'exposition à la sollicitation sur la formation des croyances) sera évoquée à l'occasion de la discussion relative à l'effet des variables situationnelles (1.2). Enfin, la phase ultime (impact des intentions comportementales sur la réponse) sera laissée de côté, puisqu'elle n'a pas été testée dans notre travail (pas d'hypothèse formulées à ce sujet)[321].

La formation de croyances positives conduit-elle à une attitude favorable ?

Nos résultats suggèrent qu'une évaluation positive de la sollicitation entraîne une attitude favorable à la communication des données, **validant ainsi la phase 2 du processus de réponse**. L'étude qualitative exploratoire (présentée au chapitre 5), renforcée par les conclusions issues de la littérature (étudiée en partie I), ont permis de dégager quatre éléments (croyances) sur lesquels l'individu forme son évaluation de la sollicitation. Nommés respectivement confidentialité perçue, sensibilité perçue, pertinence perçue et valeur perçue (ou gain net à répondre), ils correspondent aux lettres « a » à « d » au sein de notre modèle conceptuel[322]. Nous avions alors fait la supposition que chacun de ces éléments contribuait, à son niveau, à renforcer ou à diminuer (dans le cas de la sensibilité perçue) l'établissement d'une attitude favorable, ce qui était formulé à travers les hypothèses H2a (7), H2b (8), H2c (9) et H2d (13).

Les résultats obtenus dans les expériences 1 et 2 conduisent à accepter ces quatre hypothèses, contribuant ainsi à valider l'étape correspondante du processus de réponse (effet de l'évaluation de la sollicitation sur l'attitude face à la communication des données demandées).

Plus précisément, il apparaît que la confidentialité perçue des données sollicitées a un effet positif sur l'attitude face à la communication des données, même si le résultat est proche du niveau de significativité ($B = 0,192$; $p = 0,054$). Nous confirmons ainsi les conclusions des recherches menées dans le domaine du respect de la vie privée (notamment par Bies 1993, Stone et Stone 1990 et Culnan et Armstrong 1999), qui indiquent que le consommateur est plus enclin à se dévoiler quand il perçoit une possibilité de contrôler l'utilisation qui sera faite de ses données (forte confidentialité perçue).

[321] Cf. chapitre 7 pour les explications relatives à ce choix.
[322] Nous renvoyons les lecteurs aux chapitres 5 et 7 pour le rappel des définitions et l'explication de la place de ces variables au sein du modèle.

Concernant la sensibilité, là encore, le résultat va dans le sens de ce qui était attendu : des données jugées très sensibles conduisent à une détérioration de l'attitude de l'individu face à la communication de ce type d'informations (B = - 0,308 ; p = 0,001). Nous confirmons ainsi, à la fois, les résultats des recherches menées dans le domaine de la vie privée (Cranor, Reagle et Ackerman 1999, Salerno 2001) et de la réponse aux enquêtes (Singer 1984). Le fait que des questions soient jugées sensibles (i.e. qu'elles ne concernent pas celui qui les demande) serait donc significativement lié à une prédisposition défavorable au fait de les fournir.

L'hypothèse relative à l'effet de la pertinence est également validée (B = 0,346 ; p = 0,002). Ainsi, le fait de percevoir que l'entreprise a une raison légitime de recueillir ce type de données personnelles amène l'individu à former une attitude favorable à leur communication. Nous apportons alors une preuve supplémentaire de l'importance du caractère non intrusif d'une sollicitation de données, qui renforce les résultats établis dans ce sens par la littérature. Nous confirmons notamment le résultat obtenu par Hine et Eve (1998) selon lequel toute demande d'informations non jugées nécessaires à la transaction en cours est vécue comme intrusive et s'accompagne alors d'hostilité.

Nous validons enfin l'effet de la valeur perçue (ou du gain net à répondre) sur l'attitude face à la communication des informations demandées (B = 0,336 ; p = 0,000). Ainsi, l'individu peut développer une attitude favorable à la condition qu'il perçoive plus de bénéfices que de coûts à fournir les données. Nos résultats sont donc conformes à la littérature sur le dévoilement de soi et la réponse à une sollicitation de données qui prévoit l'existence d'un « *privacy calculus* » mesurant les bénéfices du dévoilement contre les conséquences négatives d'un tel choix, notamment une utilisation non autorisée des données divulguées (Laufer et Wolfe 1977).

Pour finir, nous voulons souligner qu'apparaît une certaine hiérarchie dans l'impact respectif de chacune de ces perceptions. En effet, même si nous ne pouvons pas vraiment comparer ces résultats (les 3 premiers étant issus de l'expérience 1 quand le dernier correspond à l'expérience 2), l'examen des coefficients standardisés montre que la confidentialité aurait un impact moins important sur l'attitude que la sensibilité et la pertinence perçues des données sollicitées. La valeur perçue à répondre aurait, quant à elle, une influence quasi-similaire à ces deux dernières. On peut donc supposer que l'attitude est moins influencée par la confidentialité de la requête que par les trois autres perceptions.

Dans l'ensemble, nous pouvons donc en déduire que l'effet de l'évaluation de la sollicitation sur l'attitude face à la communication des données est relativement robuste. Nous confirmons ainsi les résultats des recherches issues des théories de l'action raisonnée (Ajzen et Fishbein 1980) et de la motivation (ou de l'espérance attendue) (Vroom 1964, Farag et Krishnan 2003).

Nous poursuivons notre analyse concernant la « validité » du processus de réponse par l'étude de l'effet de l'attitude sur les intentions comportementales de réponse.

Une attitude favorable à la communication de données conduit-elle à des intentions comportementales allant dans ce sens ?

Les résultats de l'expérience 2 confirment tout d'abord l'impact de l'attitude face à la communication des données demandées sur l'intention de répondre (i.e. de les fournir) (B = 0,149 ; p = 0,000). Nous montrons ainsi que, plus l'individu développe une attitude favorable à la divulgation de ses données, plus il y a de chances pour qu'il accepte de les fournir.
Nos résultats valident aussi l'hypothèse d'un effet de l'attitude sur l'intention de mentir (i.e. de fournir des données erronées) (B = - 0,046 ; p = 0,000). Ainsi, moins l'individu est prêt à communiquer ses données, plus il y a de chances pour qu'il les falsifie, en cas de sollicitation.
L'impact de l'attitude de l'individu sur les intentions comportementales de répondre et de mentir conforte les résultats de nombreuses recherches menées en comportement du consommateur, notamment sur les problématiques d'adoption d'une innovation ou de nouvelles technologies[323]. Bien qu'il ait été très peu étudié dans le contexte d'une sollicitation d'informations personnelles, hormis en ce qui concerne l'adoption d'un comportement mensonger (Lwin et Williams 2003), ce lien entre attitude et intention comportementale, déjà largement étudié et prouvé dans la littérature[324], est donc confirmé ici.

De plus, bien que n'ayant pas fait d'hypothèse à ce sujet, nous validons l'existence d'un lien direct entre l'attitude de l'individu face à la communication de ses données et son intention de poursuivre la relation avec l'entreprise qui l'a sollicité (p = 0,000 ; cf. tableau 11.20).
Nous confirmons ainsi les propos de Helme-Guizon et Amato (2004) qui supposaient que « même un acte aussi simple qu'exprimer une attitude peut conduire à de l'engagement ».

[323] Voir notamment les travaux de Davis (1993) à ce sujet.
[324] Il a notamment été démontré dans le cadre de la théorie de l'action raisonnée (Ajzen et Fishbein 1980), de la théorie du comportement planifié (Ajzen 1991) et du modèle TAM (*Technology Acceptance Model*) (Davis 1989).

Nous apportons donc la preuve de l'influence de l'attitude face à la communication des données demandées sur l'ensemble des intentions comportementales testées ici (i.e. intention de répondre, de mentir et de poursuivre la relation), **validant alors la phase 4 du processus de réponse**[325]. Nous confirmons ainsi les résultats des travaux issus des théories de l'action raisonnée (Ajzen et Fishbein 1980) et de l'engagement (Thibaut et Kelley 1959).

Nous terminons cette analyse du processus de réponse en discutant les résultats concernant l'impact de l'intention de répondre et de la réponse sur l'intention de poursuivre la relation.

L'intention de fournir des données et le fait de le faire conduisent-ils à poursuivre la relation avec l'entreprise à l'origine de la sollicitation ?

Dans l'expérience 2, nous validons l'effet de l'intention de répondre (i.e. de fournir les données demandées) sur l'intention de poursuivre la relation (B = 1,079 ; p = 0,000). Nous montrons en effet que l'intention de l'individu de continuer sa relation avec l'entreprise est supérieure quand il est prêt à communiquer des données, **validant dès lors la phase 5 du processus de réponse**. Ce résultat confirme ainsi les principes de la théorie de l'engagement (Thibaut et Kelley 1959).

De la même manière, nous montrons dans l'expérience 3, l'influence de la réponse (i.e. fait d'avoir accepté de remplir le formulaire/questionnaire) sur l'intention de poursuivre la relation (B = 1,077 ; p = 0,000). Ainsi, une personne qui ne veut pas remplir le formulaire aurait de plus grandes chances de rompre la relation que celle qui accepterait de le faire. Nous prouvons alors l'impact possible de la sollicitation sur l'engagement de l'individu dans sa relation avec l'entreprise à l'origine de la requête, confirmant les résultats obtenus par Gundlach, Achrol et Mentzer (1995) et par Frisou (2000) et **validant ainsi la phase 6 du processus de réponse**.
Nous n'avons toutefois pas été en mesure de démontrer l'impact de l'abandon sur l'intention de poursuivre la relation (p = 0,220), ce qui constitue une bonne nouvelle pour les entreprises. Cela signifie en effet que si l'individu prend la décision de remplir le formulaire, le fait qu'il n'aille pas jusqu'au bout n'a pas d'incidence sur l'avenir de la relation.

[325] La phase 3, correspondant à l'impact de la préoccupation RVP, sera évoquée ultérieurement dans cette section.

Pour conclure ce premier point, les résultats obtenus nous permettent de valider l'ensemble du processus de réponse à une sollicitation de données personnelles testé dans cette recherche. Nous montrons ainsi que la formation de croyances positives (i.e. confidentialité et pertinence perçues des données sollicitées élevées, faible sensibilité perçue et gain net à répondre élevé) produit une attitude favorable face à la communication des données demandées. Cette attitude conduit elle-même à des intentions de répondre (i.e. de fournir les données sollicitées), de mentir (i.e. de fournir des données erronées) et de poursuivre la relation avec l'entreprise à l'origine de la requête, conformes à ce qui était supposé. Enfin, l'intention de l'individu de fournir les données et le fait de le faire aboutissent à une plus forte probabilité d'engagement.
L'ensemble de ces résultats amène donc à confirmer les conclusions de la littérature, notamment celles issues de la théorie de l'action raisonnée (Ajzen et Fishbein 1980), de la théorie de la motivation (Vroom 1964) et de la théorie de l'engagement (Thibaut et Kelley 1959).

Voyons à présent l'influence des facteurs situationnels manipulés sur ce processus de réponse.

1.2 L'impact des facteurs situationnels sur le processus de réponse

Nous nous intéressons à présent à la première phase du processus de réponse, à savoir l'impact de l'exposition à la sollicitation (i.e. des facteurs situationnels) sur l'évaluation de celle-ci. Nous discuterons aussi les résultats concernant l'effet des facteurs sur le reste du processus de réponse.

1.2.1 L'impact des facteurs sur l'évaluation de la sollicitation

Certaines des variables situationnelles testées (i.e. mentions de RVP et quantité de données) ont bien un impact sur l'évaluation de la sollicitation (mesurée à travers quatre perceptions majeures, i.e. confidentialité, sensibilité, pertinence et valeur perçues) puisque les analyses de variance qui ont pu être menées au cours des deux premières expériences laissent apparaître des effets significatifs au seuil de 10%[326]. A l'opposé, l'effet d'autres variables, et notamment de la familiarité avec l'entreprise sollicitant les données et du contexte dans lequel les informations sont requises est plus difficile à démontrer, tant les résultats obtenus sont délicats à interpréter.
Nous revenons ci-après sur chaque facteur situationnel manipulé dans le cadre de notre travail.

[326] Le choix de ce seuil en lieu et place du seuil habituel de 5% a été justifié dans les chapitres correspondants.

Impact des mentions de RVP sur l'évaluation de la sollicitation

Nos résultats montrent l'impact des mentions de RVP à la fois sur la confidentialité (p = 0,011) et sur la sensibilité (p = 0,064) perçues des données sollicitées. Nous trouvons ainsi que plus ces mentions sont nombreuses (informations sur la collecte et l'utilisation des données, choix par opt-in et présence d'un logo de sécurité), plus cela contribue à rassurer le consommateur sollicité quant à l'utilisation ultérieure des données (confidentialité élevée) et à la vulnérabilité que leur divulgation peut occasionner (faible sensibilité). Nous confirmons ainsi certaines conclusions issues de la littérature sur la vie privée, qui démontrent l'impact majeur des pratiques éthiques et de l'information donnée au consommateur sur sa perception de vulnérabilité (Stone et Stone 1990) et son sentiment de contrôle (Culnan et Armstrong 1999). De plus, bien que n'ayant pas fait d'hypothèse à ce sujet, nous trouvons que les mentions de RVP ont aussi un effet significatif (p = 0,057) sur la pertinence perçue des données sollicitées. La présence de ces mentions contribuerait ainsi à renforcer la légitimité de l'entreprise à solliciter de telles informations.

Ce facteur n'ayant pas été manipulé dans l'expérience 2, nous n'avons pu tester (et *a fortiori* démontrer) son influence sur la valeur perçue à répondre, même si nous pouvons supposer que tel aurait été le cas. En effet, la présence de mentions de RVP semble contribuer à rassurer le consommateur quant à l'avenir de ses données personnelles, ce qui devrait réduire sa perception des risques à répondre et donc accroitre le gain net perçu. Des études complémentaires sont néanmoins nécessaires pour confirmer une telle hypothèse.

Au final, nous sommes donc en mesure d'établir que la politique de respect de la vie privée adoptée par la firme (à travers les mentions apposées sur le formulaire) a une influence majeure sur l'évaluation que l'individu porte à la sollicitation, et notamment sur les perceptions de confidentialité, de sensibilité et de pertinence des données sollicitées. Pour renforcer et confirmer ce résultat, de nouvelles recherches pourraient être menées, notamment en manipulant d'une autre manière la politique de respect de la vie privée (par exemple, par le biais d'une charte).

Impact de la familiarité sur l'évaluation de la sollicitation

Aucune des hypothèses concernant l'impact de la familiarité sur l'évaluation que l'individu fait de la sollicitation n'a pu être validée. En particulier, nous n'avons pas été en mesure de montrer qu'elle pouvait influencer la confidentialité et la pertinence perçue des données sollicitées. Ce résultat vient donc contredire nombre de conclusions issues de la littérature sur la vie privée (Culnan et Armstrong 1999), le marketing relationnel (Milne et Boza 1999) et le marketing direct (Rogers 1996) qui affirment le rôle de l'expérience passée sur la formation des cognitions.

Ceci amène à se poser la question des raisons ayant conduit à ce résultat. Plusieurs justifications peuvent être avancées sans que nous puissions pour autant affirmer qu'elles sont les bonnes.

La première explication, qui est d'ailleurs confirmée par nos analyses, concerne l'existence possible d'effets d'interaction entre la familiarité et d'autres facteurs situationnels tels que ceux manipulés ici. Nous trouvons en effet que, si la familiarité n'a aucun effet direct ni sur la confidentialité (p = 0,179) ni sur la pertinence perçue (p = 0,651) des données sollicitées, l'effet d'interaction avec la quantité de données demandées est significatif dans les deux cas (p = 0,085 et 0,079 respectivement). Autrement dit, l'individu ne tient pas compte uniquement de l'ancienneté de sa relation avec l'entreprise pour former son évaluation mais plutôt de l'effet conjoint de ce facteur (i.e. la familiarité) avec le nombre d'informations à fournir (i.e. la quantité). Plus précisément, nous avons pu établir qu'à un niveau de familiarité élevée, un formulaire court engendrait une plus forte confidentialité et pertinence perçues des données.

D'autres explications à l'absence d'effet direct de la familiarité peuvent être trouvées dans la manière dont nous avons choisi de définir ce terme et d'opérationnaliser ce facteur. Nous reviendrons sur ces deux possibilités dans la section relative aux limites de cette recherche (cf. paragraphe 3.2.1). Enfin, nous pensons également à la confusion qui s'est peut-être opérée dans l'esprit des répondants entre familiarité avec l'entreprise et familiarité avec le site.

Des études complémentaires sont donc indispensables afin de pouvoir déterminer avec plus de certitude laquelle de ces explications est la plus plausible et ainsi pouvoir dire si la familiarité n'a réellement aucun effet (direct) sur l'évaluation de la sollicitation.

Impact de la quantité de données demandées sur l'évaluation de la sollicitation

Nos résultats confirment le rôle de la quantité d'informations requises sur l'évaluation de la sollicitation et plus précisément sur la sensibilité (p = 0,000) et la pertinence (p = 0,002) perçues des données demandées ainsi que sur la valeur perçue à répondre (p = 0,076). Nous montrons aussi (comme nous nous y attendions) que ce facteur n'a aucun impact sur la confidentialité des données (p = 0,978), si ce n'est en interaction avec le facteur « Mentions de RVP » (p = 0,080). Plus précisément, il apparaît qu'un formulaire court engendre une plus faible sensibilité perçue, une plus forte pertinence perçue et un plus grand gain net à répondre (à bénéfice égal) que lorsque le formulaire comporte une quantité importante d'informations à fournir. Ce résultat, somme toute relativement logique, vient confirmer nombre d'études réalisées sur le sujet en matière de réponse à des enquêtes, notamment celle menée par Childers et Skinner (1996).

Au final, nous pouvons donc dire que plus la quantité d'informations figurant sur le formulaire à remplir est réduite, plus l'individu formera des cognitions (perceptions) positives à son sujet. Des

recherches complémentaires pourraient toutefois être menées afin de déterminer quel est le nombre optimal de données à demander et s'il existe un seuil « de tolérance » à ne pas dépasser.

Impact du contexte de sollicitation sur l'évaluation de celle-ci

Le contexte n'ayant pas été manipulé dans l'expérience 1, nous n'avons pas été en mesure d'évaluer son impact ni sur la confidentialité, ni sur la sensibilité ni sur la pertinence perçue des données sollicitées. Les résultats relatifs à l'expérience 2 montrent par ailleurs que le contexte[327] dans lequel sont sollicitées les données n'a aucun effet sur la valeur perçue à répondre (p=0,479). L'impact de ce facteur (tel que nous l'entendons ici) n'ayant jamais été testé auparavant, nous ne pouvons réellement nous appuyer sur la littérature pour en tirer un quelconque enseignement. Néanmoins, des recherches réalisées dans le domaine de la vie privée (Culnan et Bies 2003) ou de la réponse aux enquêtes (Childers et Skinner 1996) pouvaient laisser supposer l'influence du type d'échange proposé sur l'attractivité perçue du résultat et donc sur le gain perçu à répondre.
Les raisons d'un tel résultat sont donc difficiles à établir. Une première explication possible, fournie par nos analyses, tient au choix de la manipulation de ce facteur, et notamment à la distinction de trois niveaux (i.e. jeu-concours, newsletter, espace personnel) qui ne seraient peut-être pas suffisamment différents les uns des autres. Ainsi, il apparaît une certaine « proximité » entre la newsletter et l'espace personnel, qui se différencient tous deux du contexte « jeu-concours », bien que de façon non significative. Une opposition plus forte et plus visible entre les contextes testés aurait peut-être pu permettre d'aboutir à un résultat différent. De même, d'autres contextes, plus différenciés (i.e. demande de catalogues et inscription à un club de fidélité), auraient également pu conduire à des conclusions toutes autres.
Il est enfin possible, si ce n'est probable, que le contexte n'ait pas d'effet direct sur la valeur perçue à répondre, mais un effet indirect, en interaction avec d'autres facteurs manipulés. Si cette hypothèse n'a pas pu être vérifiée dans l'expérience 2 (i.e. pas d'effet significatif en interaction avec la quantité de données demandées : p = 0,633), des études complémentaires seraient nécessaires afin de le vérifier sur d'autres facteurs, notamment la familiarité.

Au final, nous n'avons pas été en mesure de montrer l'impact du contexte dans lequel les informations sont requises sur l'évaluation de la sollicitation, limitée dans ce cas à la valeur perçue à répondre, l'impact sur les autres perceptions n'ayant pas pu être testé. Des études

[327] En l'occurrence, trois « niveaux » de ce facteur étaient testés : le jeu concours, la newsletter et la création d'un espace personnel sur le site.

complémentaires devront donc vérifier si ce résultat se confirme ultérieurement ou s'il est lié aux choix opérés ici, notamment en termes d'opérationnalisation de ce facteur.

Les variables perceptuelles mesurées dans cette recherche (i.e. confidentialité, sensibilité, pertinence et valeur) ne seraient pas forcément des médiateurs de l'effet de l'exposition à la sollicitation (facteurs situationnels manipulés) sur le comportement de réponse, en atteste les résultats obtenus pour l'expérience 3. Ceci laisse alors penser que le processus d'évaluation ne serait pas exclusivement cognitif, signe d'un jugement parfois heuristique de la situation et de l'existence de liens directs entre les facteurs situationnels et le reste du processus de réponse. Nous abordons à présent les résultats concernant cette possibilité de liens directs.

1.2.2 L'impact des facteurs sur le reste du processus de réponse

Nous distinguons ici l'impact direct des facteurs situationnels sur le comportement de réponse et sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données.

Impact des facteurs situationnels sur le comportement de réponse

Dans l'expérience 3 qui est celle où nous avons mesuré la réponse à la sollicitation, trois types de comportements étaient distingués : le fait d'accepter (ou non) de fournir les données demandées (noté i dans le modèle à tester), d'aller (ou non) jusqu'au bout du questionnaire (ou au contraire d'abandonner avant la fin, j) et de répondre à une question sensible (niveau de dévoilement, k). L'impact des trois facteurs manipulés dans cette expérimentation était ainsi testé : celui de la familiarité avec l'entreprise, de la quantité de données demandées et du contexte de sollicitation. Nous discutons à présent brièvement les résultats obtenus pour chacun de ces facteurs.

Concernant la familiarité, les résultats montrent son influence sur la probabilité de répondre ($p = 0,000$), même si l'effet est contraire à celui supposé. Nous trouvons ainsi que les individus les moins familiers sont les plus enclins à répondre, ce qui est vient infirmer de nombreux résultats établis dans la littérature sur la vie privée (Culnan et Armstrong 1999, Zhang et al. 2001) et la réponse aux enquêtes (Comley 2000, Tezinde et al. 2002). Des explications ont été avancées dans le chapitre correspondant (i.e. 12), la plus plausible tenant au fait que les nouveaux clients ont été moins sollicités par le passé et qu'ils sont donc moins importunés par ce type de requête que les clients plus anciens. De nouvelles études doivent être menées pour savoir si cette justification est confirmée ou si le résultat est lié à d'autres raisons.

Nos résultats montrent aussi l'absence d'influence de la familiarité sur la probabilité d'abandonner avant la fin (p = 0,987) et de répondre à une question sensible (p = 0,662). Autrement dit, la familiarité n'aurait pas d'effet, une fois prise la décision de remplir le formulaire de données. Nous confirmons là un des constats majeurs des recherches sur la réponse aux enquêtes qui montrent que certains facteurs ont un impact décisif sur la décision initiale de participer tandis que d'autres exercent davantage leur influence sur le comportement qui s'en suit, mesuré notamment à travers le taux et la qualité des réponses (Vehovar et Manfreda 2000).

Les résultats concernant l'influence de la quantité de données demandées sont contrastés. D'un côté, nous montrons son impact sur la probabilité d'aller au bout du questionnaire (p = 0,000), ce qui va dans le sens des conclusions de Manfreda et al. (2002). De l'autre, nous ne trouvons pas d'influence sur la probabilité de répondre à une question sensible (p = 0,805). Ce dernier résultat nous semble cependant dû, au moins en partie, au côté facultatif des réponses dans le cadre de notre expérience, le caractère sensible de la question étant dès lors atténué.

Concernant le contexte de sollicitation, les résultats sont pour le moins déconcertants. Nous validons en effet les trois hypothèses relatives à l'impact du contexte sur la probabilité de répondre (p=0,019), d'aller au bout (p=0,049) et de répondre à une question sensible (p=0,000). Toutefois, l'influence est systématiquement inverse à ce qui était prévu. Alors que la littérature laissait supposer que le comportement serait plus favorable dans le cas de l'enquête, nous trouvons au contraire que c'est le jeu-concours qui favorise la divulgation. Ce constat pourrait néanmoins s'expliquer par la mécanique mise en place par l'entreprise avec laquelle nous avons collaboré dans le cadre de cette expérience. Le jeu proposé favorisait en effet l'implication du répondant puisqu'à chaque réponse, il multipliait ses chances de gagner. Il était donc encouragé à répondre à l'ensemble des questions (y compris les plus sensibles d'entre elles) et à aller jusqu'au bout. De plus, ce résultat vient confirmer les conclusions de Frick et al. (1999) qui montrent que la chance de gagner un prix réduit le taux d'abandon.
Dans l'ensemble, nous validons ainsi l'effet direct du contexte sur le comportement de réponse même si celui-ci s'opère dans le sens inverse à ce qui était prévu.

Impact des facteurs situationnels sur l'intention de poursuivre la relation

Nous avions fait l'hypothèse que certains des facteurs situationnels manipulés dans le cadre de l'expérience 3 auraient un effet direct sur l'intention de poursuivre la relation avec l'entreprise. Nos résultats confirment ces suppositions, au moins en ce qui concerne le contexte (p = 0,000).

De plus, l'effet va dans le sens attendu (l'enquête favorise davantage l'engagement que le jeu). L'impact de la familiarité est également validé mais il est « annulé », une fois contrôlée la qualité perçue de la relation (p = 0,444). Cette dernière supplante alors l'effet de la familiarité qui, à qualité perçue de la relation identique, n'aurait plus d'influence. Ce résultat est conforme à la littérature en marketing relationnel, qui affirme qu'au-delà de l'ancienneté, d'autres éléments doivent être pris en compte pour évaluer la relation, dont la satisfaction vis-à-vis des expériences passées (Ganesan 1994), que l'on peut considérer comme un proxi de la qualité perçue.
Dans l'ensemble, on peut donc confirmer la possibilité d'un effet direct des facteurs situationnels sur la probabilité d'engagement, cet impact pouvant cependant être réduit voire annulé lorsque certaines variables individuelles (telles que la qualité perçue de la relation) sont contrôlées.

Pour conclure ce deuxième point, les résultats de la recherche permettent de mettre en avant l'influence de variables situationnelles qui n'ont, à notre connaissance, jamais été étudiées dans le cadre de la réponse à une sollicitation de données personnelles (i.e. familiarité et contexte). Nous mettons ainsi à jour l'influence directe de la familiarité sur la probabilité de remplir le formulaire, même si l'effet est inverse à ce qui était attendu, les clients récents se montrant plus coopératifs. Nous validons aussi l'existence d'un effet direct du contexte de sollicitation sur la probabilité de répondre, d'aller au bout et de répondre à une question sensible, l'influence étant là encore contraire à ce qui était attendue puisque c'est le jeu-concours qui conduit au comportement le plus favorable. Enfin, nous montrons également l'influence du contexte sur l'intention de poursuivre la relation, l'enquête amenant cette fois-ci un engagement plus profond.
Les autres relations mises en avant par cette recherche viennent confirmer en grande partie les travaux précédents. Ainsi en est-il des mentions de RVP qui, conformément à ce que la littérature laissait supposer, influencent de manière significative l'évaluation de la sollicitation à travers la confidentialité, la sensibilité et la pertinence perçue des données sollicitées. Enfin, la quantité influence également à la fois l'évaluation de la sollicitation (sensibilité, pertinence et valeur perçues) et la réaction du consommateur, à travers le comportement d'abandon.

Nous abordons la dernière phase de discussion des résultats à travers l'analyse de l'impact des variables individuelles sur le processus de réponse à une sollicitation de données personnelles.

1.3 L'impact des variables individuelles sur le processus de réponse

Nous distinguons ci-après l'influence de la préoccupation pour le RVP (phase 3 du processus de réponse) de celle liée à l'ensemble des variables exogènes contrôlées dans cette recherche.

1.3.1 L'impact de la préoccupation pour le respect de la vie privée

Hormis l'influence de la préoccupation pour le RVP sur la probabilité d'abandon qui n'a pas pu être démontrée[328], les autres hypothèses concernant l'impact de cette variable sur le processus de réponse ont toutes été acceptées, **validant ainsi la phase 3 de notre processus de réponse**. Nous confirmons en effet le rôle de la préoccupation dans la formation de l'attitude face à la communication des données demandées dans nos deux premières expériences (p=0,000 et 0,014 respectivement). De même, nous validons son impact sur la probabilité de remplir le formulaire (p =0,000) et de répondre à une question sensible (p =0,001).

Nous montrons ainsi que, plus un individu est préoccupé par le respect de sa vie privée, plus il y a de chances pour qu'il développe une attitude défavorable à la communication de ses données, qu'il refuse de remplir le formulaire et qu'il s'abstienne de fournir une donnée sensible au cas où il aurait accepté de répondre. Ce constat est en tout point conforme à ce que la littérature sur le dévoilement de soi et le respect de la vie privée prévoyait[329].

Le deuxième résultat important à propos de la préoccupation pour le RVP concerne son influence relative par rapport à celle des facteurs situationnels. A ce sujet, il apparaît que la situation de sollicitation supplanterait, parfois très largement, certaines convictions personnelles pourtant solidement établies. Ainsi, un individu fortement préoccupé pourrait néanmoins accepter de répondre si les circonstances l'amènent à penser qu'il y a intérêt à le faire. Ce constat, qui n'avait pas donné lieu à la formulation d'hypothèses, était pourtant prévisible si l'on s'en tient aux conclusions de certains auteurs tels Acquisti (2004) qui insistent sur le caractère souvent irrationnel de la décision de fournir ou non des données personnelles. Nous confirmons donc ici la prépondérance de la situation de sollicitation sur les caractéristiques individuelles.

Nous avons terminé la discussion concernant les relations ayant donné lieu à la formulation d'hypothèses. Nous résumons à travers la figure suivante (cf. Figure D.1), l'ensemble de celles qui ont pu être validées dans cette recherche, omettant volontairement celles qui ont été rejetées. Nous constatons que l'ensemble du processus de réponse est validé et que seul l'impact de certains facteurs situationnels sur l'évaluation de la sollicitation reste sujet à caution.

[328] Après réflexion, ce résultat peut sembler logique, l'impact de la préoccupation se faisant au moment de la décision initiale de participer ou non (et au moment de répondre à une question si celle-ci s'avère sensible) mais pas une fois le processus enclenché. La décision d'abandonner serait dans ce cas davantage lié à un effet de fatigue et/ou de saturation qu'à une forte préoccupation pour ce qui peut advenir des données.

[329] Cf. chapitre 7, paragraphe 4.2.2 pour les références d'auteurs ayant établi des conclusions dans ce domaine.

Figure D.1 Ensemble des hypothèses validées dans le cadre de cette recherche

En trait plein, les hypothèses validées

En pointillé, les hypothèses partiellement validées

En double trait, les relations validées hors hypothèse

En pointillé double-trait, relation non testée dans la thèse

Variables de contrôle

- *Niveau d'implication vis-à-vis de la catégorie de produits (IMPL)*
- *Variables sociodémographiques (âge, sexe, CSP, niveau d'éducation)*
- *Expérience (EXP) et Utilisation (UTIL) d'Internet*
- *Comportement habituel de réponse (COMP_HAB)*
- *Qualité perçue de la relation (QUALITE)*

N.B. : Pour des raisons de lisibilité, les hypothèses non validées ne sont pas représentées sur la figure

(*) Abréviations pour Communication de Données Personnelles (CDP) et Niveau de Dévoilement (ND)

Parallèlement à l'ensemble des constats précédents qui ont apporté une réponse à la question de la validation (ou non) des hypothèses de recherche proposées, d'autres résultats ressortent des expérimentations, qui permettent un enrichissement des conclusions obtenues jusqu'ici. Ceux-ci concernent l'impact des variables exogènes, appelées aussi variables de contrôle.

1.3.2 L'impact des variables exogènes

Nous avions pu faire ressortir, de la littérature et de nos études exploratoires, un ensemble de variables ayant un impact présumé sur le processus de réponse, que nous avons été amenées à contrôler dans le cadre de notre dispositif expérimental. Ces variables, de nature exclusivement individuelle, ont dès lors été introduites en tant que co-variables dans toutes nos analyses.

Les analyses de covariance qui ont pu être menées dans les expériences 1 (MANCOVA) et 2 (ANCOVA) et l'introduction de ces variables dans les analyses de régression (notamment au cours de l'expérience 3) montrent que certaines d'entre-elles ont un impact significatif sur les relations testées et qu'elles ne doivent donc pas être négligées.
Nous reprenons ci-après les variables exogènes pour lesquelles l'influence s'avère significative. Nous analyserons dans l'ordre le niveau d'implication, les variables sociodémographiques et liées à Internet ainsi que le comportement habituel. Nous ne reviendrons pas sur la qualité perçue dont on a vu qu'elle avait un impact majeur sur la probabilité de poursuivre la relation.

L'impact de l'implication dans la catégorie de produits/services

Nos analyses montrent que le niveau d'implication dans la catégorie de produits/services aurait un impact global significatif sur l'évaluation que l'individu porte à la sollicitation, et ce, pour chacune des variables perceptuelles concernées. Ainsi, un niveau élevé d'implication augmenterait la confidentialité ($p = 0,089$), la pertinence ($p = 0,010$) et la valeur perçue à répondre ($p = 0,003$) et diminuerait la sensibilité perçue des données demandées ($p = 0,003$).
Ces résultats sont conformes aux conclusions issues de la littérature. En effet, il a été montré que lorsque le consommateur est impliqué, il a davantage tendance à rechercher de l'information (Amine 1993) et donc à être attentif à tout signe susceptible de le renseigner sur la situation et sur le stimulus qui lui est proposé. Ce faisant, il se montre donc plus confiant, pour autant que l'entreprise puisse le rassurer sur ses intentions aussi bien actuelles (sensibilité et pertinence) que futures (confidentialité et valeur).

L'impact des variables sociodémographiques

Quatre variables sociodémographiques ont été introduites comme co-variables dans les analyses[330]. Elles semblent dans l'ensemble avoir peu d'influence, hormis sur le comportement de réponse. Ainsi, sexe et âge influenceraient tous deux la probabilité de remplir le formulaire de données, les femmes et les jeunes étant respectivement plus enclins à accepter. Le niveau d'études est la seule variable sociodémographique à influencer la probabilité d'abandon, les plus éduqués ayant une plus forte probabilité d'aller jusqu'au bout. Enfin, la probabilité de répondre à une question sensible serait influencée aussi bien par l'âge, le niveau d'études et la CSP, à chaque fois négativement, les gens âgés, de niveau d'études et de CSP supérieure ayant une plus forte réticence à fournir une information qu'ils jugent sensible.

Le résultat concernant le sexe semble confirmer la littérature sur le dévoilement de soi qui prédit que les femmes parlent plus facilement d'elles-mêmes, sont moins préoccupées par la protection de leur vie privée et acceptent plus volontiers de fournir des données personnelles (Cosby 1973).
L'analyse sur l'âge va aussi dans le sens de la majorité des travaux sur le sujet (dont celui de Milne et Gordon 1993) qui concluent à une plus grande préoccupation chez les personnes âgées. D'après les auteurs, cet effet serait d'ailleurs renforcé sur Internet, les jeunes étant plus à l'aise avec les nouvelles technologies, ce qui se trouve confirmé ici.
S'agissant du niveau d'études, nos résultats confirment les conclusions du sondage Louis Harris (1995) qui montre que les gens éduqués sont plus enclins à refuser. D'autres travaux dans ce domaine montraient toutefois l'effet inverse voire l'absence d'effet significatif (cf. tableau 4.2).
Enfin, pour la CSP, nos résultats vont plutôt dans le sens inverse de ce que la littérature prédit, même s'ils confirment ceux trouvés par Robins (1963) en matière de réponse à un sondage.

L'absence d'uniformité dans les conclusions de la littérature concernant l'impact des variables sociodémographiques sur le comportement de réponse semble donc se retrouver ici, nos résultats étant assez hétéroclites, même s'ils confirment en grande partie ceux déjà établis.

[330] Pour rappel, dans l'expérience 1, seul le sexe était contrôlé, les autres variables sociodémographiques étant homogènes du fait de la particularité de l'échantillon interrogé (i.e. étudiants).

L'impact des variables Internet

Les variables Internet (ancienneté, fréquence de surf et/ou de consultation de la messagerie) semblent influencer à la fois certaines variables perceptuelles et des éléments du comportement de réponse. Ainsi, fréquence de surf et de consultation de la messagerie auraient respectivement un impact positif sur la sensibilité et la pertinence perçue des données sollicitées. L'expérience d'Internet (ancienneté) aurait quant à elle un effet négatif sur la probabilité de remplir le formulaire, les internautes plus anciens étant plus réticents à répondre. Ce dernier résultat va dans le sens de celui trouvé par Nyshadham (2000) qui indique que plus l'expertise en ligne est élevée, plus la préoccupation est forte. Il s'avère cependant contraire aux conclusions de Sheehan (2002) qui montrent que, par grâce au contrôle perçu, les internautes aguerris disposeraient d'un moyen de réduire leur anxiété et, par conséquent, leurs préoccupations de vie privée.

Ces résultats vont donc dans le sens de certains de ceux établis dans la littérature, tout en étant en contradiction avec d'autres. Ce constat n'est cependant pas surprenant en soi, la majorité des recherches sur le sujet ayant tendance à aboutir à ce type de conclusions contradictoires, notamment du fait de l'absence d'unité dans la manière d'opérationnaliser les variables étudiées.

L'impact du comportement de réponse habituel

Les conclusions concernant l'impact du comportement habituel de réponse sont, en tout point, conformes à ce que l'on pouvait attendre et à ce que la littérature prédit dans ce domaine. Ainsi, les personnes ayant l'habitude de donner des informations et d'aller au bout du formulaire sont celles qui, dans nos expériences, se sont montrées les plus favorables. Nous confirmons ainsi les résultats que nous avions obtenus dans l'étude exploratoire descriptive (cf. chapitre 6, tableau 6.7) et le constat établi par Ehrenberg (1974), selon lequel le comportement passé serait un plus grand prédicateur du comportement futur que ne le sont les attitudes.

Pour conclure ce troisième point, nous avons démontré l'impact des variables individuelles sur certains éléments du processus de réponse à une sollicitation de données. Cette influence est toutefois loin d'être systématique (seules quelques variables dépendantes sont concernées) et est surtout moins importante que celle liée aux facteurs situationnels, hormis ce qui concerne le comportement habituel.

En guise de synthèse, nous proposons, dans le tableau suivant, de mettre en perspective les résultats de la recherche par rapport aux conclusions mises en avant par la littérature.

Tableau D.2 Mise en perspective des résultats de la recherche

Facteurs étudiés	Résultats des travaux empiriques[331]	Résultats dans notre recherche
Impact des facteurs situationnels [332]		
Mentions RVP	- l'absence d'information et de choix engendre une sensation de perte de contrôle (Culnan 1995) - ne pas connaître l'utilisation ultérieure des données accroît la vulnérabilité du consommateur (Stone et Stone 1990)	**Résultats confirmés :** Les mentions de RVP influencent à la fois la confidentialité, la sensibilité et la pertinence perçue des données sollicitées et donc l'évaluation que l'individu porte à la sollicitation
Familiarité	- pas de recherche montrant l'impact de la familiarité sur la réponse à une sollicitation de données personnelles - nombreux travaux en marketing relationnel et marketing direct établissant l'influence de l'expérience passée avec l'entreprise sur la formation des cognitions (dont Milne et Boza 1999, Rogers 1996)	Résultats non confirmés : Pas d'effet (direct) observé ni sur la confidentialité ni sur la sensibilité ni sur la pertinence perçues des données sollicitées Mais, un effet significatif sur la confidentialité et la pertinence perçue en interaction avec la quantité de données demandées
Quantité de données demandées	- l'augmentation du nombre de données demandées accroît la perception de vulnérabilité (Stone et Stone 1990) - un questionnaire court réduit le montant d'effort nécessaire pour le remplir et donc décroît le coût de la participation (Childers et Skinner 1996)	**Résultats confirmés :** La quantité de données demandées influence positivement la sensibilité et la pertinence perçues des données demandées et négativement la valeur perçue à répondre
Contexte de sollicitation	- aucune recherche à notre connaissance en termes de réponse à une sollicitation de données personnelles - des recherches prouvant l'impact du type d'échange proposé sur l'attractivité perçue du résultat (Childers et Skinner 1996)	Résultats non confirmés : Aucun effet du contexte sur la valeur perçue à répondre
Impact des facteurs individuels[333]		
Préoccupation pour le RVP	La majorité des recherches confirment l'impact de la préoccupation pour le RVP sur le comportement de réponse (notamment celle de Sheehan et Hoy 1999)	**Résultats confirmés :** La préoccupation influence le comportement de réponse mais aussi l'attitude face à la communication des données
Variables socio-démographiques	Les résultats étant souvent en contradiction les uns avec les autres, il est difficile d'affirmer de manière certaine l'impact d'une ou de plusieurs variables socio-démographiques (cf. chapitre 4, tableau 4.2)	**Résultats confirmés :** Les résultats trouvés vont généralement dans le sens de la majorité des travaux sur le sujet
Variables Internet	Là encore, les résultats ne sont pas uniformes d'une étude à l'autre, des auteurs démontrant l'impact de certaines variables quand d'autres trouvent le contraire	Résultats ambigus : Les résultats confirment systématiquement certains de ceux établis dans la littérature tout en étant contraires à d'autres.

[331] Nous n'indiquons que quelques résultats issus de la littérature, choisis pour leur proximité avec ceux établis ici.
[332] Nous nous limitons ici à l'impact des facteurs situationnels sur les quatre variables perceptuelles testées.
[333] Nous nous limitons ici à ces trois variables individuelles qui sont les plus étudiées dans la littérature.

Synthèse de la première section

Les résultats permettent de statuer sur l'existence de différentes phases inhérentes au processus adopté par le consommateur en réponse à une sollicitation de données à caractère personnel.
En l'occurrence, nous prouvons que :
- l'évaluation de la sollicitation (en termes de confidentialité, sensibilité, pertinence et valeur perçue à répondre) conduit à la formation d'une attitude (favorable ou non) face à la communication des données demandées (validation de la phase 2[334] du processus de réponse) ;
- cette attitude engendre à son tour des intentions comportementales, en termes de réponse à la sollicitation (intention de remplir le formulaire), de mensonge (fournir des données erronées) et de poursuite de la relation avec l'entreprise à l'origine de la requête (validation de la phase 4[335]) ;
- l'intention de répondre et la réponse elle-même influencent pour finir l'intention du consommateur de poursuivre la relation avec l'entreprise l'ayant sollicité (validation des phases 5 et 6).
Ces constats sont cohérents avec les travaux en comportement du consommateur, notamment ceux relatifs au modèle TRA, TPB et TAM mais aussi concernant la théorie de l'engagement.

Les analyses menées permettent aussi d'apporter des éléments de réponse concernant l'influence des facteurs situationnels sur le processus de décision et le comportement du consommateur.
En particulier, nous observons que les mentions de RVP et la quantité de données demandées ont un impact majeur sur l'évaluation de la sollicitation et sur le reste du processus de réponse, ce qui vient confirmer nombre de travaux déjà réalisés à ce sujet.
De plus, nous identifions l'effet possible de deux facteurs jamais étudiés jusqu'ici dans les travaux sur le dévoilement de soi : la familiarité avec l'entreprise à l'origine de la requête et le contexte de sollicitation. Les résultats obtenus peinent toutefois à démontrer un tel effet, suggérant alors que des recherches complémentaires soient menées dans ce domaine.

De manière complémentaire, nous observons l'influence de diverses variables individuelles (particulièrement celle de la préoccupation pour le RVP) sur les différents niveaux du processus de réponse. Ces résultats sont, dans l'ensemble, cohérents avec ceux trouvés dans la littérature. L'influence des facteurs individuels est toutefois très inférieure à celle des facteurs situationnels.

[334] La phase 1 correspond à l'impact des variables situationnelles manipulées sur l'évaluation de la sollicitation et est donc abordée ci-dessous.
[335] La phase 3 correspond à l'impact de la préoccupation pour le RVP et est donc abordée ci-dessous.

2 Les apports de la recherche

Les apports de cette recherche sont d'ordre théorique, empirique, managérial et méthodologique. L'étude du processus de décision adopté par le consommateur lorsqu'il se trouve confronté à une sollicitation de ses données personnelles et des facteurs, notamment situationnels, qui influencent ce processus contribuent ainsi à accroître les connaissances en matière de dévoilement de soi et d'incidence des préoccupations de vie privée sur le comportement du consommateur (2.1). L'analyse de ce phénomène dans le contexte français constitue, à cet égard, un apport empirique certain, face à la quasi-absence de recherches hexagonales dans ce domaine (2.2). Le test du modèle dans le cadre d'un formulaire à remplir sur Internet permet, en outre, de dégager plusieurs implications à visée managériale (2.3). Enfin, l'apport méthodologique de la recherche repose essentiellement sur le recours à un plan expérimental original comprenant une étude menée auprès de 20 000 clients d'une société partenaire (2.4).

2.1 Les apports théoriques de la recherche

Nous ne reviendrons pas ici sur les résultats spécifiques à cette recherche (cf. section 1), qui constituent autant d'apports théoriques puisque certains n'avaient jamais été démontrés auparavant, quand d'autres permettent de mettre en perspective les conclusions issues de recherches antérieures. Nous préférons insister sur les apports académiques majeurs qui sont liés à l'amélioration de la compréhension du phénomène étudié (2.1.1), à la nouveauté de la démarche (2.1.2), ou encore à la mise en avant de résultats fondamentaux liés à l'étude de facteurs situationnels peu pris en compte jusqu'ici (2.1.3).

2.1.1 Une compréhension élargie de la décision du consommateur

Ce travail a permis l'élaboration d'un modèle conceptuel visant à expliquer le comportement de réponse du consommateur sollicité pour fournir des informations à caractère personnel. Celui-ci constitue un apport théorique certain dans la mesure où, à notre connaissance, aucun modèle de ce type n'avait été proposé auparavant dans ce champ de recherche. Il a, en outre, l'avantage d'inclure des facteurs à la fois situationnels (correspondant à l'exposition à la sollicitation) et personnels (notamment à travers la préoccupation pour le respect de la vie privée), et donc de pouvoir tester l'impact respectif de chacun d'eux.

Les résultats de cette recherche, confrontés aux travaux issus de la littérature, nous permettent donc d'enrichir la compréhension du processus de décision adopté par le consommateur, que nous synthétisons à travers le « modèle explicatif » présenté ci-dessous (tableau D.3).

Tableau D.3 Modélisation explicative du processus de réponse à une sollicitation de données

Phases	Définition	Principales caractéristiques
Exposition / Détection	Confrontation au stimulus (le formulaire de données)	L'individu se retrouve exposé à un formulaire qu'il doit choisir de compléter ou non
Perception / Evaluation	Interprétation de la sollicitation et évaluation cognitive de celle-ci	L'ensemble des éléments liés à la sollicitation (facteurs situationnels) conduit l'individu à évaluer la situation en termes de confidentialité, pertinence, valeur et sensibilité des données sollicitées
Réaction affective	Evaluation globale de la communication des données sur un continuum favorable-défavorable	L'évaluation cognitive aboutit à un jugement global, essentiellement de nature affective, de l'acte qui consiste à communiquer les données demandées
Réponse[336]	Décision de fournir (ou non) les données demandées	L'attitude développée précédemment conduit l'individu à prendre la décision de fournir (ou non) les données demandées

2.1.2 Une approche novatrice de l'étude du comportement de réponse

Le caractère novateur de cette recherche réside, tout d'abord, dans le sujet étudié (le dévoilement de soi et le respect de la vie privée), peu courant en France, bien qu'abondamment exploité outre-Atlantique. Il est également lié au choix d'une vision processuelle du comportement de réponse (cf. paragraphe précédent), dont on a vu qu'elle était largement ignorée dans la littérature. Enfin, nous avons aussi privilégié une approche « réaliste » du phénomène, en cherchant à mettre les personnes interrogées dans une situation qui se rapproche le plus possible des conditions réelles d'une sollicitation de données personnelles sur Internet. Notre troisième expérience, menée auprès de véritables clients, constitue à cet égard une originalité, puisque peu de recherches s'appuient sur de tels tests grandeur nature.

[336] Nous incluons dans cette phase aussi bien les intentions comportementales de réponse que la réponse elle-même.

2.1.3 La mise en avant de l'effet de la situation de sollicitation

L'analyse de la situation de sollicitation, à travers l'impact des facteurs situationnels, constitue en soi un apport théorique substantiel, puisque c'est plutôt l'étude des facteurs individuels qui est privilégiée dans la littérature. De plus, nous avons choisi des facteurs jusqu'ici largement ignorés dans les recherches existantes et à forte portée managériale. Enfin, nos résultats permettent de mettre en avant la prépondérance de la situation par rapport aux convictions personnelles (mesurée à travers la préoccupation pour le respect de la vie privée) dans la décision de l'individu de fournir ou non les données demandées.

L'étude de variables « originales » et à forte portée managériale

Les facteurs que nous avons choisi d'étudier sont à la fois originaux au sein de la littérature sur la réponse à une sollicitation de données personnelles et opérationnels, dans le sens où ils peuvent être facilement « manipulés » par les managers, y compris dans une optique de test.
Hormis les mentions de RVP dont l'effet a été démontré auparavant (mais pour lesquelles la manipulation expérimentale proposée ici est nouvelle dans un contexte français), les trois autres facteurs étudiés contribuent à approfondir les connaissances déjà établies quant à l'incidence d'une sollicitation de données personnelles sur la réponse du consommateur.
Ainsi, malgré les difficultés à prouver l'influence de la familiarité sur les différentes variables du processus de décision, les résultats « prometteurs » auxquels nous avons néanmoins pu aboutir[337] démontrent la nécessité d'élargir le périmètre de la réflexion et de ne pas se contenter d'étudier des variables « classiques » telles que la réputation de l'entreprise. Il s'agit en outre d'un facteur sur lequel l'entreprise peut facilement agir, en adaptant le formulaire à l'ancienneté de la relation et à ce qu'elle connaît déjà du client.
De même, l'impact avéré de la quantité de données demandées constitue une avancée, tant d'un point de vue théorique que pratique, puisqu'il confirme les résultats de recherches passées (notamment celles réalisées dans le cadre de la réponse à une enquête) et démontre la nécessité de ne pas alourdir le processus de collecte par un formulaire trop long.

[337] Les résultats de l'expérience 3 prouvent ainsi une influence de la familiarité sur certaines variables de réponse (notamment sur la probabilité de communiquer les données demandées), même si elle s'avère contraire à l'hypothèse formulée (ce sont les nouveaux clients et non les plus anciens qui se sont montrés les plus réactifs).

Enfin, l'étude du contexte de sollicitation constitue l'originalité majeure de notre travail, même si les résultats ne se révèlent pas forcément à la hauteur de nos attentes[338]. En effet, ce facteur n'avait jamais été étudié en tant que tel dans la littérature, si ce n'est indirectement, à travers le type de bénéfices proposés en échange des données, ou à travers l'objectif de visite affiché par le consommateur (i.e. recherche d'information, surf, achat …). Le fait de s'intéresser à une variable spécifique (i.e. le contexte) permet donc de comparer nos résultats avec ceux existants, et de les prolonger. De plus, le contexte est un élément plus opérationnel que l'objectif de visite, qui n'est pas « maîtrisable » par l'entreprise à l'origine de la sollicitation. L'idée d'adapter le formulaire aux circonstances dans lesquelles les données sont sollicitées (i.e. jeu-concours, newsletter …), même si elle n'est pas formellement démontrée ici, constitue à l'évidence un résultat intéressant pour des managers en quête d'éléments clés en termes de personnalisation.

La mise en avant d'un effet supérieur de la situation de sollicitation

Un des résultats majeurs de notre recherche réside dans la démonstration d'un effet supérieur de la situation de sollicitation par rapport aux variables individuelles. Ce résultat, intéressant en tant que tel puisqu'il indique la possibilité pour les entreprises d'agir sur le processus de réponse à travers la « manipulation » de certains facteurs situationnels, constitue un apport théorique certain. Il n'avait en effet jamais été démontré jusqu'ici, alors même que nombre de travaux s'attachent à déterminer les critères les plus à même d'influencer le consommateur. Savoir que les convictions personnelles passent au second plan doit ainsi inciter les chercheurs à étudier davantage l'impact des mécanismes situationnels.

Pour conclure ce premier point, cette recherche contribue à une meilleure compréhension du processus de décision adopté par le consommateur sollicité pour fournir des données à caractère personnel. En outre, en privilégiant une approche réaliste, nous apportons les conditions permettant de mieux appréhender ce phénomène. L'étude de facteurs novateurs et à forte portée managériale permet aussi de mettre à jour les limites actuelles de la littérature. Enfin, la mise en avant de la prépondérance de la situation de collecte conduit à une remise en question des recherches qui seraient uniquement axées sur l'étude des facteurs individuels.

[338] Seule l'hypothèse concernant l'impact du contexte sur l'intention de poursuivre la relation est réellement validée, les autres résultats allant dans le sens inverse de ce qui était attendu (i.e. effet supérieur du jeu-concours comparativement à celui de l'enquête).

2.2 Les apports empiriques de la recherche

Le respect de la vie privée et la réponse du consommateur face à une sollicitation de ses données personnelles sont quasi-exclusivement appréhendés sous l'angle anglo-saxon, voire américain. Dans cette recherche, nous nous intéressons plus spécifiquement au comportement du consommateur français et à ses spécificités (réglementation à l'échelle européenne et valeurs liées à la culture latine). Pouvoir confronter nos résultats avec ceux obtenus outre-Atlantique constitue donc une première avancée, dans la mesure où cela permet de consolider les conclusions établies dans ce domaine ou au contraire d'en déterminer les limites. Ainsi, le fait de trouver que les mentions de RVP ont une influence sur la réponse du consommateur des deux côtés de l'Atlantique montre, qu'en dépit de systèmes de régulation différents, Américains et Européens sont également attentifs à l'utilisation qui peut être faite de leurs données personnelles. De plus, la mise en évidence du processus de réponse étudié ici constitue une validation empirique de notre modèle et confirme l'intérêt du recours à une démarche processuelle. Ce cadre conceptuel peut, à son tour, être utilement comparé aux recherches existantes, afin d'en dégager les apports majeurs et les limites éventuelles. Le test de ce modèle dans un autre contexte culturel constitue ainsi une voie de recherche prometteuse.

Pour conclure ce second point, cette recherche apporte une contribution empirique à l'étude du comportement de réponse à une sollicitation de données personnelles, notamment sur Internet.

2.3 Les implications managériales de la recherche

Les implications managériales de ce travail sont nombreuses et permettent de répondre à certaines interrogations (et craintes) formulées par les professionnels concernant le dilemme entre besoin en informations et nécessaire respect de la vie privée des consommateurs. Cette partie donne donc lieu à un ensemble de recommandations à l'usage des responsables marketing et, plus spécifiquement, à tous ceux qui conçoivent des formulaires sur Internet.

A l'issue de notre recherche, nous sommes ainsi en mesure d'identifier les critères en fonction desquels les consommateurs décident de fournir (ou non) les données demandées (2.3.1) et les mécanismes de confiance et de réassurance à mettre en place pour que ceux-ci acceptent de se dévoiler (2.3.2).

Par ailleurs, les conséquences d'une sollicitation jugée abusive peuvent obliger les entreprises à revoir leurs politiques de sollicitation et nous proposons alors des moyens pouvant être utilisés pour se prémunir contre les risques de rupture de la relation (2.3.3).

2.3.1 L'identification des critères de décision des consommateurs

Les résultats de notre étude qualitative, confirmés par les expériences 1 et 2, soulignent l'existence de quatre critères majoritairement utilisés par les consommateurs pour former leur décision. Ceux-ci concernent respectivement la confidentialité, la sensibilité et la pertinence des données sollicitées, auxquelles s'ajoute la perception du gain net (valeur) à répondre. Par ricochet, ces critères permettent d'identifier les pratiques susceptibles de faire l'objet d'un rejet. Ainsi, une sollicitation de données jugées confidentielles, sensibles ou peu pertinentes eu égard au contexte accroît la probabilité d'engendrer un refus. De même, le fait d'entrevoir plus de risque à répondre que de bénéfices à en retirer peut également encourager l'individu à s'abstenir. Ces critères de décision doivent donc être connus des managers afin qu'ils puissent s'y adapter.

2.3.2 L'identification des mécanismes de confiance et de réassurance

Connaître les critères utilisés par les consommateurs pour former leur décision ne suffit pas. Il faut encore trouver des méthodes permettant de contourner d'éventuels mécanismes de blocage. Nos résultats conduisent à identifier plusieurs solutions visant à rassurer les individus et ainsi à accroître leur confiance dans la capacité de l'entreprise sollicitante à respecter ses engagements. Ces solutions sont détaillées ci-dessous.

Solution n° 1 : Adopter une politique de respect de la vie privée claire et cohérente

Le premier impératif en cas de sollicitation de données est de rassurer le consommateur quant à l'utilisation qui pourra être faite de ses données. Ceci passe, notamment, par une politique de respect de la vie privée respectueuse des lois en vigueur et empreinte d'une certaine éthique. En cela, afficher des indications claires et cohérentes sur chaque page comportant un formulaire à remplir s'avère essentiel, si ce n'est primordial. De plus, donner au consommateur la possibilité de choisir ce qui sera fait de ses données (transfert ou non à des tiers extérieurs par exemple) apparaît tout aussi fondamental. Ces éléments pourraient en outre être renforcés par l'apposition de tout sigle susceptible de rassurer le consommateur quant aux garanties qui lui sont offertes.

Solution n° 2 : Limiter la quantité et la sensibilité des informations demandées

Le deuxième impératif réside dans l'élaboration d'un formulaire limité aux données strictement nécessaires à la transaction en vigueur et qui ne soient pas considérées comme sensibles. Une manière efficace de respecter ces principes tout en y ajoutant une certaine souplesse consisterait à favoriser le recours à des champs facultatifs qui offrent la possibilité de collecter l'information de la part d'individus moins sensibles à ce type de critères tout en permettant aux personnes qui y sont attachées de s'abstenir de répondre. L'étude de l'efficacité de cette solution constitue une piste de recherche prometteuse qui sera détaillée ultérieurement.

Solution n° 3 : Accroître les bénéfices proposés en échange des données

Une autre solution susceptible de rassurer le consommateur et donc de l'inciter à répondre consiste à accroître les bénéfices proposés en échange des données. Au-delà d'une réflexion poussée sur le type d'avantages à privilégier compte tenu du contexte dans lequel les données sont sollicitées et du type de public auquel on s'adresse[339], nous estimons qu'une meilleure mise en avant de ces bénéfices pourrait s'avérer tout aussi efficace, si ce n'est plus. Cette dernière passe notamment par une description détaillée des privilèges pouvant être accordés en cas de réponse favorable et par une communication adéquate, susceptible d'attirer l'attention (i.e. couleurs « flashy », clignotement, taille plus importante, etc …).

Solution n° 4 : Réduire les coûts liés au remplissage du formulaire

Outre les risques liés à une utilisation abusive des données une fois celle-ci divulguées, nos résultats indiquent que les coûts liés à la fourniture des informations concernent principalement la charge cognitive engendrée par le remplissage du formulaire. Il convient alors de réfléchir à tout mécanisme susceptible de le faciliter (i.e. données préenregistrées, champs facultatifs, lisibilité des intitulés …), tout en insistant sur l'esthétique et le design des pages correspondantes (i.e. convivialité, clarté …).
Ces éléments de « confort » peuvent, en outre, faire l'objet de pré-tests afin de déterminer leur niveau optimal, compte tenu du type de produits/services concernés et de la clientèle ciblée.

[339] Les bénéfices proposés doivent en effet s'adapter au type de transaction étudié (les bénéfices monétaires sont ainsi limités à l'organisation de jeux-concours) et au profil de l'individu auquel on s'adresse (à titre d'exemple, les jeunes et les personnes plus âgées n'ont pas forcément les mêmes attentes en termes d'avantages recherchés, même si les résultats de l'étude descriptive présentée au chapitre 6 peinent à le démontrer).

Au vu des résultats de la recherche, il semble par ailleurs que les entreprises devraient également porter leur attention aux conséquences possibles d'une sollicitation qui serait jugée abusive et/ou trop intrusive. Nous avons en effet montré que plusieurs éléments pouvaient concourir à la volonté du consommateur de rompre la relation avec l'entreprise l'ayant sollicité. Parmi ces éléments, le fait d'avoir ou non accepté de remplir le formulaire, le contexte dans lequel les informations ont été sollicitées et, dans une moindre mesure, le niveau de familiarité avec la firme sollicitante (supplanté toutefois par la qualité perçue de la relation).
Il faut donc réfléchir activement aux moyens de se prémunir contre un tel risque de rupture.

2.3.3 Les moyens de se prémunir d'une rupture de la relation

Quel que soit le domaine étudié, les managers cherchent souvent à « manipuler » les perceptions et l'attitude des consommateurs afin de les faire réagir de manière conforme à leurs attentes. Notre recherche montre que les consommateurs ne sont pas dupes et que de telles pratiques peuvent s'avérer dangereuses voire, à terme, nuire à l'établissement de relations durables. Nos résultats permettent en outre d'envisager plusieurs solutions à mettre en place pour se prémunir d'un tel risque. Quatre pistes sont plus particulièrement envisagées.

Solution n° 1 : Réfléchir aux occasions permettant de solliciter des informations

Même si nous n'avons pas pu démontrer l'impact du contexte dans lequel les informations sont collectées sur l'évaluation de la sollicitation (plus précisément sur la valeur perçue à répondre), nos résultats confirment son influence sur le comportement de réponse (bien que dans un sens contraire à celui attendu) et sur l'intention de poursuivre la relation avec l'entreprise à l'origine de la demande. Que l'individu ait ou non répondu, nous trouvons donc que le simple fait d'avoir été sollicité et le contexte dans lequel cette demande s'est opérée sont susceptibles d'avoir une incidence sur son engagement dans la relation. Ceci doit donc inciter les managers à ne pas sur-solliciter leurs prospects et/ou clients au risque de les voir fuir. De la même manière, ils devront différencier les situations où l'on peut se permettre de solliciter des données des cas où la sollicitation n'apparaît pas opportune et où elle est susceptible d'occasionner un rejet.

Solution n° 2 : Adapter la demande à l'état de la relation entretenue avec le consommateur

Là encore, même si nous n'avons pu formellement démontrer l'impact de la familiarité sur les différentes variables du processus de réponse (évaluation et comportement[340] notamment), les résultats prouvent que cette variable ne doit pas pour autant être négligée. Nous trouvons en effet qu'elle exerce une influence sur l'intention de continuer à commercer avec l'entreprise quand la qualité perçue de la relation n'est pas contrôlée. Ceci indique qu'en termes de sollicitation, les managers doivent tenir compte de l'ancienneté de la relation avec le consommateur (i.e. ne pas sur-solliciter les prospects et les nouveaux clients). A terme, il serait également souhaitable de personnaliser la demande en fonction de ce que l'on connaît déjà de la personne plutôt que de fournir un formulaire identique à tout le monde.

Solution n° 3 : Montrer l'intérêt de répondre favorablement à la sollicitation

Comme indiqué précédemment, nos résultats montrent que le fait d'avoir répondu (ou non) à la sollicitation aurait une influence non négligeable sur la poursuite de la relation : un individu ayant accepté aurait ainsi une plus forte intention d'engagement. La firme a donc intérêt à mettre en valeur l'avantage qu'il y a à répondre, de manière à favoriser l'implication du consommateur et à réduire ainsi le risque que la sollicitation ne conduise à une rupture de la relation. Outre les bénéfices pouvant être proposés en échange des données, la solution passe aussi par une facilitation du passage à l'acte, grâce à tout mécanisme susceptible de provoquer la confiance et de réassurer l'individu au moment de sa prise de décision (cf. paragraphe 2.3.2).

Solution n° 4 : Prendre en compte les caractéristiques des consommateurs / visiteurs

Nous avons montré que certaines caractéristiques individuelles étaient susceptibles d'influencer la réaction des consommateurs face à une sollicitation de leurs données personnelles. Ainsi, le niveau de préoccupation pour le respect de la vie privée et le niveau d'implication auraient une influence certaine sur la décision de l'individu. Il serait donc utile, pour autant que cela puisse se faire[341], d'identifier suffisamment en amont le profil du visiteur, afin d'adapter la demande (le formulaire) aux caractéristiques de celui-ci.

[340] Nous trouvons toutefois qu'il existe un effet direct de la familiarité sur la réponse, bien qu'il soit contraire au sens attendu.

[341] La question qui se pose ici est autant technique (dispositif nécessaire pour le faire) que morale, dans le sens où un tel mécanisme d'identification ne doit pas pour autant porter atteinte au respect de la vie privée des personnes.

Plusieurs solutions « techniques » sont d'ores et déjà disponibles à cet égard :

- identifier le visiteur par son comportement de navigation grâce à l'envoi de cookies et/ou l'analyse de fichiers logs. Le niveau d'implication pourrait ainsi se mesurer par le temps passé sur chaque page du site. Ces éléments permettraient de déterminer rapidement (dès les premières secondes de visite) le profil de l'individu, grâce à quelques indicateurs clés.

- identifier le visiteur à l'aide de questions filtres posées dès le début du processus de recueil. Il s'agit ici de formuler quelques questions préalables auxquelles l'individu devrait répondre (par exemple, son genre ou son âge) et qui permettrait de déterminer son niveau de préoccupation pour le RVP et ainsi de l'affecter à une page formulaire adaptée à son « groupe de référence »[342].

Afin de déterminer le type de questions à poser, des tests[343] visant à analyser l'influence des variables sociodémographiques et expérientielles sur le niveau de préoccupation pour le RVP ont été menés dans nos différentes études (i.e. 2 à 5). Dans les expériences 1 à 3 (correspondant aux études 3 à 5), nous avons ainsi créé 4 groupes de taille proportionnée à partir du score de préoccupation, afin de pouvoir comparer les résultats avec ceux obtenus lors de l'étude 2 (cf. tableau 6.6)[344]. L'objectif est alors de voir si ces groupes se différencient sur des variables signalétiques et si les conclusions sont proches d'une étude à l'autre. Les résultats sont présentés ci-dessous (cf. tableaux D.4 et D.5).

Tableau D.4 Effectifs (en %) des groupes de préoccupation pour le RVP dans les études 2 à 5

Groupes de préoccupation [345]	Etude 2 (FING, Ch. 6)	Etude 3 (EXP1, Ch. 10)	Etude 4 (EXP2, Ch. 11)	Etude 5 (EXP3, Ch. 12)
Groupe 1 (peu préoccupés)	19,9%	22,9%	23,8%	20,6%
Groupe 2 (assez préoccupés)	24,5%	20,8%	24,6%	32,7%
Groupe 3 (préoccupés)	31,6%	25,1%	27,8%	22,7%
Groupe 4 (très préoccupés)	24,0%	31,2%	23,8%	24,0%

[342] Les résultats de notre étude descriptive (cf. chapitre 6) pourraient alors se révéler fort utiles, les quatre classes identifiées ayant montré une certaine variété, bien que parfois relative, en termes de caractéristiques individuelles.
[343] Il s'agit quasi exclusivement de tests de chi-deux, dont on a démontré les limites dans l'étude présentée au chapitre 6. Des tests multivariés seraient ici aussi les bienvenus, de par la richesse des informations obtenues.
[344] Pour rappel, la typologie réalisée pour l'étude 2 (cf. chapitre 6) identifiait 4 groupes différents sur la base des avantages recherchés en échange des données (et donc par prolongement sur leur niveau de préoccupation).
[345] Dans les expériences 1 à 3, les peu préoccupés ont un score de préoccupation assez faible (de 1 à 10 sur 21), ce score étant moyen (11 à 14) pour les « assez préoccupés », ainsi de suite. Pour l'étude 2, ces groupes correspondent aux quatre classes identifiées dans l'ordre suivant : bienveillants, négociateurs, désintéressés et réticents.

Tableau D.5 Tests de chi-deux sur le niveau de préoccupation pour le RVP

Variables signalétiques [346]	Etude 2	Etude 3 (EXP1)		Etude 4 (EXP2)		Etude 5 (EXP3)	
	Sig.	p	Sig.	p	Sig.	p	Sig.
Sexe (H, F)	ns	0,088	**0,10**	0,651	ns	0,049	**0,05**
Age	ns	-		0,449	ns	0,004	**0,01**
CSP (Profession)	ns	-		0,062	**0,10**	0,041	**0,05**
Niveau d'études	**0,01**	-		0,865	ns	0,235	ns
Ancienneté d'usage du Net	**0,01**	0,128	ns	0,778	ns	0,558	ns
Fréquence messagerie	-	0,249	ns	0,089	**0,10**	0,011	**0,01**
Fréquence de surf	**0,05**	0,867	ns	0,473	ns	-	
Expérience d'achat en ligne	**0,01**	0,926	ns	0,807	ns	-	

Les résultats issus de cette analyse sont décevants dans la mesure où aucune des variables signalétiques testées ne semble discriminer l'appartenance aux groupes dans toutes les études (elles sont au mieux significatives dans 2 études sur les 4). Ils peuvent néanmoins s'expliquer par le manque d'hétérogénéité au sein des échantillons d'une part et par la différence de profil de chaque échantillon d'autre part. Des études complémentaires seraient donc nécessaires pour trouver d'éventuelles variables discriminantes.

Quelle que soit la technique utilisée parmi celles proposées (i.e. identification par le comportement ou par question filtre) et dans la mesure où les résultats sont utilisables (obtention de variables discriminantes notamment), une fois le profil de l'individu identifié, il faudrait alors lui proposer un formulaire adapté, tant au niveau du « design » [347] que du type de questions posées. Concernant cette quatrième solution, nous voulons cependant insister sur le fait que nous ne recommandons pas en priorité l'adaptation des formulaires aux caractéristiques individuelles de chaque consommateur[348], mais plutôt que ceux-ci soient conçus de manière à favoriser son implication et à réduire ses préoccupations en matière de respect de la vie privée.

Pour conclure ce troisième point, nous avons indiqué aux managers différentes recommandations visant à accroître la motivation et à réduire les freins liés à la fourniture de données personnelles, tout en ayant un effet préventif sur les risques de rupture de la relation.

[346] Malgré notre volonté d'harmoniser la mesure de ces variables dans les différentes études, tel n'est pas toujours le cas (exemple : l'ancienneté d'usage est une variable métrique dans l'étude 2 et ordinale dans les 3 autres).
[347] Nous entendons ici ce terme au sens large pour désigner l'ensemble des éléments concourant à la présentation du formulaire (champs obligatoires/facultatifs, type de champs i.e. cases à cocher, …).
[348] Les solutions proposées ici étant difficiles et coûteuses à mettre en place, elles sont probablement réservées aux entreprises renommées et fortement implantées sur le réseau.

Les principales solutions proposées pour pallier les risques de refus de répondre et de désengagement sont ainsi résumées à travers la grille de décision suivante (cf. tableau D.6).

Tableau D.6 Solutions envisagées pour pallier les risques de refus et de rupture

Solutions envisagées pour réduire les risques de refus et accroître la confiance	Caractéristiques et opérationnalisation
Adopter une politique de respect de la vie privée claire et cohérente	- avoir des mentions de RVP clairement affichées (visibles) au niveau du formulaire - proposer des informations claires et détaillées - donner la possibilité de cocher (opt-in) ou décocher (opt-out) les cases correspondant à l'utilisation des données - adopter et mettre en avant des logos, sceaux et labels garantissant la sécurisation des procédures - proposer sur chaque page comprenant un formulaire un lien vers une charte de RVP claire et cohérente
Se limiter à des données peu sensibles et en nombre restreint	- proposer en priorité un formulaire court (le nombre optimal de champs à compléter reste à déterminer) - ne pas solliciter de données majoritairement considérées comme sensibles (exemple : n° de téléphone, revenus, ..) ou les mettre en champs facultatifs - favoriser le recours aux champs facultatifs
Proposer des bénéfices substantiels et mieux les mettre en avant	- proposer des avantages substantiels et adaptés au contexte de sollicitation et au public visé - insister sur ces bénéfices par une mise en valeur appropriée (couleurs, taille des caractères, ..)
Réduire les coûts liés au remplissage du formulaire	- faciliter le remplissage du formulaire (champs facultatifs ou pré-remplis, lisibilité des intitulés, …) - rendre le remplissage agréable en insistant sur l'esthétique et le design de la page (convivialité, clarté, ..)
Solutions envisagées pour réduire les risques de rupture	Caractéristiques et opérationnalisation
Se limiter à une sollicitation opportune des données	- ne solliciter des informations que lorsque cela est strictement nécessaire à la réalisation de la transaction (accepter l'idée de rendre des services « gratuits »)
Adapter la sollicitation à l'état de la relation client	- tenir compte de l'ancienneté de la relation : ne pas demander la même chose aux prospects et aux clients - personnaliser le formulaire en fonction des informations déjà détenues dans les bases de données
Montrer l'intérêt de répondre favorablement à la sollicitation	- mettre en avant les bénéfices proposés en échange - insister sur les mécanismes producteurs de confiance et de réassurance
Prendre en compte les caractéristiques des consommateurs	- identifier au préalable le profil de l'individu - adapter le formulaire au niveau de préoccupation pour le RVP

2.4 Les contributions méthodologiques

Les contributions méthodologiques, plus limitées que les apports théoriques et managériaux, concernent principalement la mise en œuvre de la démarche expérimentale (2.4.1) et, de manière plus accessoire, la procédure de développement d'échelles (2.4.2)

2.4.1 Les leçons tirées de la mise en œuvre du plan expérimental

Puisque notre expérimentation a fait l'objet d'une réflexion approfondie qui s'est soldée par l'adoption d'un design original (i.e. trois expériences successives, chacune focalisée sur une partie du modèle à tester) et qu'une des études s'est déroulée sur le terrain auprès de clients réels, nous pouvons faire part de notre expérience en avançant quelques propositions.

Proposition n°1 : Choisir un design réaliste et impliquant

L'objectif d'une expérience est de mettre des individus dans une situation fictive et de mesurer leurs réactions face à différents scénarios proposés. Pour que la validité (aussi bien interne qu'externe) des résultats soit assurée, il faut alors faire en sorte que la situation soit la plus proche possible de la réalité et suffisamment impliquante pour que les répondants n'adoptent pas un rôle factice et se comportent de la manière la plus naturelle qui soit. Ainsi, dans le cadre de notre expérience 1, nous avons remarqué (suite à plusieurs pré-tests) qu'il était préférable de faire remplir le formulaire aux participants de manière à mieux simuler la situation et à les impliquer davantage. Cette solution offre en outre l'avantage de permettre une manipulation plus efficace du ou des facteur(s) (dans notre cas du facteur Quantité) concerné(s). De plus, le degré de réalisme du scénario peut être pré-testé puis mesuré lors de l'expérience, par le biais d'une question simple en fin de questionnaire, comme nous l'avons fait pour l'expérience 1.

Proposition n°2 : Réfléchir avec soin aux moyens de manipuler les facteurs

Lors des pré-tests et des études finales, nous avons remarqué que certains consommateurs avaient du mal à se plonger dans les scénarii proposés, notamment lorsque les facteurs faisaient l'objet d'une manipulation par instruction. Dans ce cas en effet, cela oblige les répondants à lire ces instructions et à être particulièrement attentifs à tout ce qui est dit, ce qui est cognitivement beaucoup plus exigeant que le simple fait d'observer une image, comme c'est généralement le

cas quand il s'agit d'une manipulation par évènement. Nous en déduisons que dans une logique expérimentale, il faut privilégier ce dernier type de manipulation (i.e. par évènement) ou conjuguer les deux. Ainsi, dans le cas de la familiarité, nous avons ajouté au texte un logo correspondant à l'opérateur concerné. De plus, si la manipulation par instruction est la seule envisageable, il faut alors multiplier les occasions de rappeler au répondant les caractéristiques de la situation dans laquelle il se trouve, afin d'être certain que la manipulation sera effective. A ce propos, le pré-test systématique des manipulations et le recours à des « *manipulation checks* » en fin de questionnaire s'avèrent tout à fait indispensables.

Proposition n°3 : Préférer multiplier les expériences avec un petit nombre de facteurs

Nous n'inventons rien en indiquant qu'il est préférable de multiplier les expériences comprenant un petit nombre de facteurs manipulés plutôt que d'adopter une expérience unique. L'originalité du plan expérimental proposé ici vient cependant confirmer ce propos, tout en apportant des éclaircissements supplémentaires. Nous recommandons ainsi de choisir avec soin les facteurs manipulés conjointement, en étant particulièrement attentifs au principe d'orthogonalité, tout en maximisant le nombre total d'effets testés (qu'ils soient directs ou d'interaction)[349]. Ces deux règles, bien que parfois difficilement conciliables, doivent si possible être respectées de manière à optimiser les résultats obtenus et permettre les recoupements entre les différentes expériences. Dans notre cas, plusieurs effets d'interaction possibles ont ainsi pu être testés, hormis celui correspondant à un risque d'orthogonalité (i.e. RVP x Contexte).

Proposition n°4 : Favoriser le recours à des expériences sur le terrain

Dans le cadre de cette recherche, une des trois expériences réalisées a été menée sur le terrain auprès de clients d'une entreprise partenaire. Cette solution offre plusieurs avantages dont les plus importants sont liés à la représentativité de l'échantillon interrogé et à la limitation de biais inhérents à la logique expérimentale (notamment le risque d'artefact de la demande). Les résultats sont également d'autant plus intéressants qu'ils offrent la possibilité de mesurer le comportement réel plutôt que des intentions comportementales ou un comportement déclaré. Néanmoins, le recours à ce type d'échantillon hétérogène rend plus difficile la mise en œuvre de l'expérimentation et oblige le chercheur à suivre plusieurs précautions, notamment en ce qui concerne l'indépendance des observations au sein des différents traitements.

[349] Ces deux critères ont été expliqués au chapitre 8.

Proposition n°5 : Réduire au maximum la durée de l'expérimentation

Cette dernière proposition n'est pas spécifique à la logique expérimentale mais est encore plus prégnante dans ce cas, du fait de la lourdeur du dispositif mis en place et du temps dont doivent alors disposer les participants (accueil, réponse au questionnaire, débrief éventuel, etc, ..).
Là encore, nous ne sommes pas les premiers à dire que l'une des préoccupations majeures du chercheur et du praticien des études doit être de réduire au maximum le temps de passation du questionnaire. Dans nos expériences 1 et 2, les répondants passaient en moyenne 15 minutes à le remplir. En l'occurrence, certains ont semblé trouver cela trop long et, même s'ils n'ont pas abandonné avant la fin, leur agacement se remarque à travers leurs réponses (i.e. pas de réponses ou réponses en halo). A la limite, ce type de résultat est un moindre mal car on peut alors les identifier et les supprimer de l'échantillon final. Il est fort possible cependant, et ce, malgré nos efforts pour le vérifier, que des répondants aient été plus subtils et que la détection de leur lassitude n'est pas été réalisée, ce qui est fort dommageable pour les résultats. Dans cette logique de réduction du temps de passation, des instruments comportant un faible nombre d'items doivent donc être privilégiés pour éviter de lasser le répondant. Dans notre recherche, nous avons ainsi multiplié les échelles à 3 items maximum[350]. Un autre moyen de réduire le temps de passation est de favoriser au maximum la clarté et la lisibilité des énoncés, tout en privilégiant les mesures visant à faciliter le remplissage (ne pas trop changer de type d'échelles par exemple). Enfin, la présentation du questionnaire doit également être pensée avec soin.

2.4.2 Les leçons tirées du développement d'échelles

Même si l'objectif de cette thèse n'était pas axé en priorité sur la création d'échelles, l'ensemble des concepts présents dans notre modèle conceptuel ont fait l'objet d'un développement soigné, présenté au chapitre 9. Si les mesures obtenues ne sont pas parfaites, loin s'en faut, les instruments créés et utilisés pour tester nos hypothèses présentent des indices de fiabilité et de validité satisfaisants. Ils représentent donc autant d'éléments mis à la disposition de recherches futures qui pourraient être menées sur le sujet, notamment dans un contexte français.
L'ensemble des étapes suivies pour développer ces échelles et les résultats auxquels nous avons aboutis nous ont permis de déduire certains enseignements que nous présentons ci-dessous.

[350] Nous reviendrons toutefois sur les limites associées à cette « règle » dans le paragraphe suivant.

Proposition n°1 : Ne pas limiter le nombre d'items au départ

Lors de la phase initiale du développement des instruments de mesure, nous avons choisi de limiter le nombre d'items pour chaque échelle. Les raisons de ce choix ont été largement développées dans le chapitre correspondant (i.e. 9). Au final, même si les instruments obtenus ont démontré leur validité et leur fiabilité, le faible nombre d'items (deux) dévolus à certains concepts du modèle engendre quelques inconvénients dont le premier est l'impossibilité d'évaluer les qualités psychométriques des échelles prises une à une. Si le critère de parcimonie favorisé dans cette recherche est toujours recommandé, nous préconisons toutefois de ne pas toujours l'appliquer de manière stricte et systématique, de manière à conserver une certaine souplesse et à ne pas en retirer plus d'inconvénients que d'avantages.

Proposition n°2 : Limiter le nombre d'items inversés dans les instruments de mesure

Il est souvent conseillé, lors du développement d'instruments de mesure, de recourir à des items inversés afin d'éviter la lassitude des personnes interrogées et de limiter ainsi les effets de halo. Dans toutes nos échelles, nous avons donc utilisé ce type d'items afin de garantir une meilleure qualité psychométrique des instruments ainsi développés. Or, les résultats révèlent que ces items n'étaient pas toujours bien compris ou correctement pris en compte par les répondants puisqu'ils étaient quasi systématiquement les moins bien représentés et les plus à même de faire baisser l'alpha. Au final, certains d'entre eux ont été éliminés lors du développement, conduisant ainsi à une réduction, parfois « dramatique », du nombre d'items dévolus à chaque échelle (cf. paragraphe précédent). Nous n'en déduisons pas qu'il faille systématiquement supprimer les items inversés des échelles de mesure, mais plutôt qu'il faut réfléchir avec soin au statut qu'on leur donne et au caractère plus ou moins indispensable de leur présence. Ainsi, peut-être sont-ils nécessaires pour mesurer des construits complexes mais le sont-ils moins s'agissant de concepts étroits, comme ceux mobilisés dans le cadre de cette recherche ?

Proposition n°3 : Utiliser des échelles adaptées au concept mesuré et au public interrogé

Répondre à un questionnaire n'est pas quelque chose de facile. Les répondants ont parfois, pour ne pas dire souvent, des difficultés à donner leur avis à partir des réponses proposées. Il est donc nécessaire de trouver le type d'échelle adapté au concept mesuré et favorisant la réponse des participants. Dans certains cas, il semble ainsi plus approprié de demander un degré d'accord (échelle de Likert en 5 ou 7 points) alors que dans d'autres, une mesure d'intensité étant plus

adéquate, il faudra privilégier les échelles sémantiques différentielles. Dans le cadre de notre recherche, nous avons ainsi opté le plus souvent pour le format Likert, hormis pour l'échelle de valeur perçue pour laquelle le format sémantique différentiel s'est avéré préférable. Cette nécessaire adaptation ne doit cependant pas se faire au détriment de la facilité de remplissage du questionnaire, au risque de perturber le répondant.

Pour conclure ce quatrième point, le plan expérimental adopté et l'expérience menée sur le terrain auprès de plus de 20 000 clients d'une entreprise partenaire permettent de rendre compte des avantages et inconvénients de la méthode expérimentale, tout en en déduisant un certain nombre de recommandations. Dans le même ordre d'idée, la procédure de développement d'échelle peut être améliorée en suivant certaines des propositions que nous venons de formuler.

Synthèse de la deuxième section

Les apports théoriques de cette recherche reposent avant tout sur une compréhension élargie de la décision du consommateur de fournir (ou non) les données personnelles demandées et sur la validation d'un modèle processuel détaillant les étapes par lesquelles passe l'individu pour aboutir à la réponse. La mise en avant de la supériorité des variables situationnelles sur les convictions personnelles pour influencer ce processus constitue également une avancée majeure.

Les apports empiriques sont principalement liés à l'étude du comportement du consommateur dans le cadre français et, qui plus est, sur Internet.

Les implications managériales consistent à comprendre les critères majoritairement utilisés par les consommateurs pour former leur décision. Elles concernent aussi le développement de recommandations visant à identifier les mécanismes permettant d'accroître les motivations et de réduire les freins liés à la fourniture d'informations à caractère personnel. Cette recherche permet enfin de mettre en avant la nécessité de prendre en compte les risques liés à une sollicitation qui serait jugée trop abusive et d'identifier les moyens de les réduire.

L'originalité méthodologique de la recherche réside principalement dans la mise en œuvre d'un plan expérimental audacieux dont les résultats permettent d'avancer plusieurs consignes à respecter pour que la démarche soit la plus réussie et la plus valide possible.

3 Limites et voies de recherche

Notre travail comporte certaines limites, à partir desquelles nous présenterons des voies de recherche futures visant à les contourner et à en réduire l'impact. Celles-ci se rapportent principalement à la validité interne (3.1) et à la validité externe (3.2) de la recherche. D'autres pistes de recherche, de portée plus générale, seront abordées à la fin de cette section (3.3).
Dans tout ce qui suit, nous ne développerons toutefois que les limites les plus importantes ainsi que les pistes de recherches les plus prometteuses.

3.1 Limites et voies de recherche liées à la validité interne

Dans notre cas, le souci de validité interne a été prédominant puisque notre recherche repose sur la validation d'un modèle théorique visant à expliquer le comportement du consommateur face à une sollicitation de données personnelles. De plus, la mise en œuvre d'un plan expérimental a renforcé la nécessité de se prémunir face à différents biais susceptibles de perturber les résultats. Pourtant, malgré notre volonté de réduire les menaces liées à la validité interne de la recherche, certaines limites apparaissent à partir desquelles des recherches futures pourraient être menées.

Les principales limites liées à la validité interne qui ont pu être mises à jour dans le cadre de cette recherche se rapportent aux éléments suivants :
- la mise en situation des sujets expérimentaux (3.1.1)
- les biais inhérents au sujet de recherche et au dispositif expérimental (3.1.2)

Nous abordons tour à tour ces deux points en présentant les remèdes susceptibles d'être mis en œuvre dans le cadre de recherches futures.

3.1.1 La mise en situation des sujets expérimentaux

Nous incluons dans cette catégorie tout ce qui a trait à la manière dont le scénario et les tâches qui y étaient proposées sont susceptibles d'avoir exercé une influence sur les résultats obtenus.
Nous ne revenons pas, toutefois, sur les limites liées au recours à un questionnaire papier dans le cadre de l'expérience 1 puisque nous en avons déjà justifié le choix (cf. chapitre 10) et, que nous y avons partiellement répondu dans les études suivantes (i.e. expériences 2 et 3).

Une première limite à la validité interne de nos résultats a trait à l'absence de contrôle exercé sur certains éléments susceptibles d'influencer la réponse à une sollicitation de données sur Internet. Les variables liées à la navigation en ligne ont ainsi été omises de notre recherche, notamment celles correspondant au niveau technologique du site visité, à sa fonctionnalité/navigabilité, à sa qualité visuelle, à la rapidité de chargement des pages, etc, … Comme dans les expériences 1 et 2 les répondants n'étaient pas en situation réelle, il était difficile de les prendre en compte. De plus, alors que l'expérience 3 avait lieu sur le terrain, nous n'avons pas pu mesurer leur effet, certaines variables étant de fait identiques pour tous les répondants (par exemple l'esthétique du site) alors que d'autres auraient dû être contrôlées (notamment la rapidité de chargement).
Dans la mesure où des variables de ce type sont susceptibles d'exercer une influence non négligeable sur le comportement de réponse, de nouvelles études réalisées en conditions réelles devraient donc être menées afin de pouvoir évaluer leur impact potentiel.

Une seconde limite à la validité interne de cette recherche, commune à la majorité des travaux en comportement du consommateur, et plus encore à ceux fondés sur une logique expérimentale, tient au fait de demander aux répondants de se projeter dans une situation prédéterminée et de déterminer ensuite leurs attitudes et intentions comportementales. Cette démarche n'est pas sans poser problème, notamment si la situation proposée est « distante » de la personne qui l'évalue, car il y a des chances pour que cette personne se montre psychologiquement détachée et que l'évaluation soit alors faussée. Ce « biais » pourrait être encore renforcé dans les études portant sur le respect de la vie privée, comme à tout ce qui a trait à l'éthique en général, car on a alors souvent affaire à des « réponses émotionnelles » (Jones et Verstegen 1992), qui rendent difficile pour l'individu de prédire de façon juste ce qu'il ferait.

A titre d'application, cette dernière remarque nous amène à rebondir sur les résultats obtenus dans le cadre de notre expérience 2 concernant l'intention de mentir (cf. chapitre 11). Nous avions souligné dans ce chapitre, les précautions qu'il fallait prendre à propos des conclusions produites à cette occasion. Sans remettre en cause la validité des résultats trouvés, nous estimons en effet nécessaire de réfléchir de manière approfondie à la façon de mesurer de manière plus juste cette intention comportementale particulière. L'inclusion de variables nouvelles au sein du modèle (tels que le contrôle comportemental[351] et les obligations morales perçues) constituerait une première étape dans l'amélioration de la compréhension et de la mesure du mensonge. L'étude menée par Lwin et Williams (2003) à ce propos offre d'ailleurs une base de comparaison

[351] En référence au modèle TPB (théorie du comportement planifié), évoqué au chapitre 7.

intéressante. Bien qu'*a priori* paradoxale, une seconde option obligerait à inclure différents moyens de « contrôler » la véracité des réponses émises par les répondants au sujet du mensonge. Il s'agirait alors de mesurer ce comportement à différents endroits du questionnaire (début et fin), de différentes manières (mesure objective et subjective par exemple) ou encore à l'occasion de collectes différentes (espacées de quelques semaines) afin de pouvoir faire des recoupements et ainsi d'avoir une vision plus juste du phénomène.

Les limites à la validité interne de ce travail concernent aussi les différents biais liés au sujet de recherche en lui-même et au dispositif expérimental mis en œuvre dans ce cadre.

3.1.2 Les biais liés au sujet de recherche et au dispositif expérimental

Comme le soulignent la majorité des chercheurs travaillant sur le thème de la vie privée, le sujet étudié est en lui-même une première source de biais, qui peut être encore renforcée si l'on a recours à une logique expérimentale. Ainsi, comme l'indique Kelvin (1973), une étude expérimentale sur ce thème peut être considérée comme constituant une sorte d'invasion de la vie privée, ce qui peut conduire à biaiser les résultats. Les solutions permettant de pallier ce problème ne sont pas évidentes, sauf à éviter d'indiquer lors du recrutement et au début du questionnaire que l'étude porte sur le respect de la vie privée. Une autre solution consisterait à employer des méthodes de questionnement plus « indirectes », de nature qualitative, comme celles faisant appel aux techniques projectives ou aux protocoles.

Les autres limites liées au dispositif expérimental tiennent essentiellement à l'existence de plusieurs types de biais susceptibles de perturber les résultats. Nous ne reviendrons pas sur ceux relatifs au répondant d'une part et à la non-réponse d'autre part, puisque nous les avons déjà largement évoqués au chapitre 8[352] et que l'organisation de notre plan expérimental visait, notamment, à les maitriser tour à tour dans chacune des expériences menées. Même si nous ne pouvons affirmer que ce dispositif a permis d'en réduire totalement la portée, nous estimons et les résultats le confirment, qu'ils ont été au moins en partie maîtrisés, leur influence étant dès lors très faible et ce type de limite, de fait, réduite.

[352] Nous n'évoquons pas ici le biais de couverture puisqu'il n'est pas spécifique à la logique expérimentale.

Tel n'est pas forcément le cas concernant le biais de désirabilité sociale. En effet, puisque ce type de biais est provoqué par le fait même d'être questionné et que le sujet étudié amenait à introduire des questions personnelles et indiscrètes (voire socialement sensibles), on peut penser qu'il était au minimum présent dans nos expériences 1 et 2[353].

Puisque ce biais semble particulièrement délicat à contrôler, nous proposons, dans le cadre de recherches futures, de le mesurer afin d'évaluer son effet potentiel sur les réponses des participants et donc sur la validité des résultats obtenus. Nous abordons alors une nouvelle difficulté liée au choix d'un instrument de mesure adéquat. En effet, des chercheurs indiquent que la mesure de la désirabilité sociale pose problème, faute d'échelle simple, fiable et valide disponible dans la littérature (Herbert 2005 a). Ainsi, à notre connaissance, l'échelle la plus courte qu'on puisse trouver (échelle en 12 items du 16PF-Q de Cattell 1995[354]) est peu fiable (alpha très inférieur à 0,7). Malgré cette limite, nous estimons indispensable de mesurer ce biais de désirabilité sociale dans nos travaux futurs.

Pour conclure ce premier point, nous avons répertorié les principales limites liées à la validité interne de notre travail, qu'elles concernent la mise en situation des sujets expérimentaux, le sujet de recherche ou le dispositif expérimental. Nous pourrions alors mener des études complémentaires qui prennent en compte davantage de variables de contrôle (notamment celles liées à la navigation) et qui mesurent de manière plus fiable le comportement de mensonge. D'autres pistes proposant une mesure plus indirecte des préoccupations pour le respect de la vie privée et une évaluation du biais de désirabilité sociale sont également envisagées.

Nous évoquons maintenant les limites et voies de recherche liées à la validité externe.

3.2 Limites et voies de recherche liées à la validité externe

Le souci de validité externe, qui conduit à s'interroger sur le réalisme de la recherche et à évoquer la possibilité d'extrapoler les résultats à un domaine d'investigation plus vaste, entrait en second plan dans notre travail. Les voies de recherche sont donc nombreuses et concernent principalement le choix et l'opérationnalisation des variables étudiées et manipulées (3.2.1) ainsi que la robustesse et la généralisation des résultats obtenus (3.2.2).

[353] Sa présence dans le cadre de l'expérience 3 est tout aussi probable mais beaucoup moins problématique dans la mesure où son influence s'exerce au niveau des réponses données aux différentes questions, partie dont nous ne nous sommes pas préoccupés pour l'analyse des résultats.

[354] Cette échelle présente en outre l'avantage d'avoir été traduite et validée sur un échantillon français de 1000 individus par Rolland et Mogenet (1995).

3.2.1 Le choix et l'opérationnalisation des variables étudiées

Une première limite liée à la validité externe de cette recherche concerne le choix et l'opérationnalisation des facteurs manipulés dans le cadre de notre travail.

Le modèle conceptuel général présenté au chapitre 7 fait en effet état de quatre catégories de facteurs situationnels susceptibles d'influencer la réponse du consommateur à une sollicitation de ses données personnelles. Nous avons fait le choix d'étudier plus spécifiquement une variable correspondant à chacune d'elle, limitant ainsi notre champ d'étude à quatre facteurs spécifiques (i.e. mentions de RVP, familiarité, quantité, contexte). De plus, il existait, pour chaque facteur étudié, plusieurs moyens de l'opérationnaliser. Ainsi, par exemple, nous aurions pu étudier les mentions de RVP sous la forme de la présence/absence d'un lien conduisant à une charte de RVP depuis la page contenant le formulaire.

Dans une optique d'augmentation de la validité externe de notre recherche, les travaux futurs doivent donc envisager d'autres moyens d'opérationnaliser les facteurs étudiés ici et s'intéresser à l'influence de nouvelles variables situationnelles.

D'autres moyens d'opérationnaliser les facteurs manipulés

Nous aurions pu (et pourrions à l'avenir) trouver d'autres moyens d'opérationnaliser les facteurs manipulés dans le cadre de cette recherche (mentions de RVP, familiarité, quantité et contexte).

Ainsi, concernant les mentions de RVP, la littérature souligne que le format de « *Notice* » (informations données par les entreprises sur leurs pratiques) peut affecter les perceptions et le comportement du consommateur (Sprott, Hardesty et Miyazaki 1998). Des recherches sur ce thème seraient donc particulièrement bienvenues et concerneraient notamment la clarté, la longueur et la place de ces informations sur le site et/ou sur la page contenant le formulaire. De même, concernant l'aspect « choix », il serait intéressant de tester l'effet de la possibilité donnée au visiteur de choisir le type de tiers à qui seraient transférées les données. Milne (1997) a montré que les gens étaient plus aptes à autoriser le transfert de leurs données quand il était indiqué qu'elles seraient distribuées à des entreprises dont les produits et services pourraient être intéressants pour eux plutôt qu'à des entreprises quelconques. Nous pourrions tester de manière plus concrète ce résultat en étudiant l'effet de la communication (ou non) du type de tiers concernés ou du choix laissé (ou non) à l'individu du type de tiers à qui transmettre les données (par exemple par le biais de cases à cocher).

Les recherches concernant la familiarité apparaissent prometteuses mais la difficulté que nous avons eue à montrer l'impact d'un tel facteur indique la nécessité de réfléchir de manière plus approfondie à l'opérationnalisation de cette variable. Il se pourrait que le fait de comparer une entreprise inconnue jusqu'ici (familiarité nulle) à un fournisseur actuel (familiarité élevée), comme ce fut le cas dans l'expérience 1, ne soit pas la meilleure solution, les répondants affectés au premier cas ayant manifestement du mal à « s'approprier » le scénario proposé. De plus, cette absence de familiarité se double d'une absence de réputation, ce qui renforce la difficulté à se mettre dans la situation. Manipuler conjointement ces deux facteurs (réputation et familiarité) semble d'ailleurs une idée intéressante permettant, notamment, d'introduire un niveau intermédiaire, celui d'une entreprise connue mais avec laquelle on n'a jamais eu de contacts au préalable. Concernant la familiarité nulle, peut-être faudrait-il aussi envisager d'introduire l'idée de première rencontre mémorable[355] (ou de première transaction) avec ce fournisseur plutôt que celle de rencontre fortuite (ou liée au hasard) comme ce fut le cas dans notre scénario. Comparer un degré d'ancienneté de la relation (faible ou élevé) plutôt que d'avoir un niveau nul semble également plus efficace, les analyses menées pour l'expérience 3 montrant un impact de ce facteur (i.e. la familiarité) sur la probabilité de fournir des données (même s'il s'avère dans le sens inverse de celui prévu). On pourrait, à cet égard, envisager d'introduire un niveau intermédiaire (ancienneté moyenne), afin de tester l'éventualité d'un effet de seuil.

Il faudrait aussi réfléchir au sens donné au mot « familiarité ». Nous avons en effet, à l'instar d'autres chercheurs, repris la définition classique donnée par Alba et Hutchinson (1987), alors même que d'autres approches existent. On pourrait considérer, comme le font Siriex et Dubois (1999), qu'il s'agit du « poids de l'expérience passée avec la marque » et introduire alors, en plus de la notion d'ancienneté, celle de satisfaction vis-à-vis des expériences passées.

Enfin, prendre en compte l'idée de familiarité avec le site en plus ou à la place de la familiarité avec l'entreprise[356] nous semble également une piste intéressante puisque nous pensons que le fait de connaître le site et d'avoir l'habitude d'y aller pourrait avoir une influence sur la probabilité d'accepter de fournir des données personnelles.

En ce qui concerne le facteur quantité, dont on a montré qu'il avait un impact significatif sur le fait de fournir (ou non) des données, nous estimons que de nouveaux travaux doivent être menés dans deux directions précises : d'une part, en testant d'autres niveaux que ceux proposés ici (i.e. 5 et 20 données dans les expériences 1 et 2) et, d'autre part, en introduisant un niveau intermédiaire afin d'étudier la présence éventuelle d'un effet de seuil (effet non linéaire).

[355] Nous renvoyons les lecteurs intéressés par ce sujet aux travaux de Florence Benoit (2006).

[356] Ces deux variables sont bien différentes car on peut avoir des relatons régulières avec un fournisseur mais ne jamais avoir consulté son site Internet, l'inverse n'étant cependant pas vrai.

A propos du contexte (et des bénéfices proposés en échange des données), d'autres possibilités d'opérationnaliser ce facteur existent. Nous pourrions ainsi opposer l'inscription à un service personnalisé (exemple : adhésion à un club de fidélité) versus non personnalisé (exemple : envoi d'une brochure ou d'un catalogue) ou encore l'offre de gains monétaires (i.e. argent ou bons de réduction) versus non monétaires (services, gain de temps, …). L'analyse de l'impact de ce facteur en conjonction avec d'autres variables manipulées serait également bienvenue, notamment celle non étudiée ici entre le contexte et la familiarité.

Au-delà de l'opérationnalisation des facteurs étudiés dans le cadre de cette recherche, un autre moyen d'améliorer la validité externe est de réfléchir à de nouvelles variables situationnelles susceptibles d'être testées.

D'autres variables situationnelles à étudier

Il serait fastidieux de lister toutes les variables situationnelles qui restent à étudier. Nous focaliserons donc plus particulièrement notre attention sur certaines d'entre-elles, et notamment sur celles dont l'étude apparaît comme la plus prometteuse compte tenu à la fois des résultats obtenus ici et de l'apport théorique, empirique et managérial que cela occasionnerait.

Une première piste envisageable concerne le caractère obligatoire ou facultatif des champs à compléter. La littérature sur la réponse aux enquêtes montre que les répondants abandonnent plus facilement le questionnaire quand une question ou un module de questions est obligatoire plutôt qu'optionnel (Vehovar et al. 2000). Nous pouvons donc supposer qu'il en sera de même pour un formulaire de données personnelles. Toutefois, la principale interrogation des managers à ce sujet n'est pas tant de savoir s'il est préférable de rendre certains champs facultatifs (la réponse est assez évidente) mais plutôt de connaître la proportion de champs facultatifs à intégrer sur l'ensemble du formulaire, le type de données concernées (numéro de téléphone, revenus, mais aussi peut-être la situation de famille, les loisirs, …) et la place de ces champs au sein du formulaire (est-il préférable de tous les regrouper à la fin ou faut-il les intégrer au milieu des champs à caractère obligatoire ?). L'ensemble de ces questions pourrait faire l'objet d'une recherche spécifique visant à étudier séparément chaque élément puis à les croiser afin d'identifier à la fois les effets directs et les effets d'interaction.

Une autre piste intéressante concerne l'impact du design et de l'esthétique du formulaire sur la réponse du consommateur. Sont principalement concernées ici les questions relatives à la couleur, au style, au format (i.e. formulaire sur une ou plusieurs pages, format des champs de réponse, …) de l'instrument de collecte. Ce facteur est clairement important dans le cadre de la réponse à une enquête même s'il est difficile d'établir une règle stricte dans ce domaine tant les chercheurs peinent à s'accorder sur les résultats. Certains auteurs trouvent ainsi qu'il aurait un impact limité sur la décision initiale de répondre (ou non) mais un impact probable sur le taux d'abandon (Vehovar et al. 2000), tandis que d'autres montrent qu'utiliser un design agréable augmente le taux de réponse (Dillman et al. 1993). Au-delà du type d'élément concerné, se pose donc aussi la question des variables dépendantes à tester. Une étude comme celle menée dans le cadre de notre troisième expérience serait tout à fait adaptée à ce type de sujet.

Une dernière piste de recherche que nous proposons concerne l'ordre des questions au sein du formulaire et son impact sur différentes variables dépendantes, telles que le taux et la qualité de la réponse. Cette variable gagnerait aussi à être étudiée conjointement avec d'autres facteurs tels que la quantité (étudiée ici) et le niveau de sensibilité des données demandées (pas étudié).
Cette piste avait été envisagée dans le cadre de notre travail (plus précisément dans la dernière étude) mais n'a pu aboutir du fait des choix opérés quant à la conception du questionnaire et à l'opérationnalisation de ce facteur[357] au sein de cette expérience. Ainsi, le fait de ne pas rendre obligatoire la réponse aux questions (faculté de passer à la question suivante sans répondre à celle proposée) laissait supposer que l'impact du facteur « Ordre » serait, si ce n'est nul, au minimum très faible[358]. En effet, poser la question sur les revenus en 6ème ou en 15ème position, sachant que, dans les deux cas, la réponse est facultative, ne doit pas changer grand-chose.

L'analyse des effets de ce facteur dans l'expérience 3 (notamment de son impact sur la probabilité de répondre et d'abandonner)[359] semble confirmer cette hypothèse (cf. tableau D.7).

[357] Pour rappel, ce facteur a été manipulé dans l'expérience 3 sous la forme de la place de la question sur les revenus au sein du questionnaire (en 6ème position : niveau sensible ; en 15ème : niveau neutre).
[358] D'où notre décision de ne pas inclure ce facteur dans nos analyses.
[359] L'analyse a été menée sur le premier fichier (dit de requalification) et sans introduire de variable de contrôle. Nous procédons par le biais de régressions logistiques, la variable dépendante étant, dans les deux cas, binaire.

Tableau D.7 Effets de l'Ordre sur la communication de données et l'abandon (EXP3)

	VD : CDP (oui vs. non)				VD : Abandon (oui vs. non)			
	Effets directs		Tous effets		Effets directs		Tous effets	
	B	Sig.	B	Sig.	B	Sig.	B	Sig.
FAMILIARITE	0,236	**0,000**	0,492	**0,002**	- 0,094	**0,046**	- 0,152	0,560
QUANTITE	- 0,025	0,372	- 0,079	0,600	1,023	**0,000**	1,392	**0,000**
CONTEXTE	0,185	**0,000**	0,368	**0,015**	- 0,532	**0,000**	- 0,231	0,354
ORDRE	0,009	0,762	0,115	0,446	0,046	0,312	- 0,512	**0,040**
FAM x QUANTITE			0,063	0,267			- 0,112	0,257
FAM x CONTEXTE			- 0,148	**0,011**			0,049	0,605
FAM x ORDRE			- 0,076	0,181			0,111	0,239
QUANTITE x CONTEXTE			0,004	0,948			- 0,306	**0,001**
QUANTITE x ORDRE			- 0,021	0,708			0,163	**0,090**
CONTEXTE x ORDRE			0,022	0,699			0,090	0,328

Ainsi, l'ordre n'a aucun impact ni sur la communication de données (CDP) ni sur l'abandon. Par contre, quand nous introduisons les effets d'interaction entre les quatre facteurs manipulés, l'effet direct de l'ordre influence de manière significative l'abandon (p = 0,040). De plus, l'effet conjoint de ce facteur avec le facteur quantité apparaît significatif au seuil de 10% (p = 0,090). Cet effet est confirmé quand on s'intéresse au taux de réponse brut à la question sur les revenus. Ainsi, il apparaît que 6% des individus choisissent de ne pas répondre à cette question lorsqu'elle est en 6ème position, ce taux passant à 7,5% quand elle se trouve en 15ème position.

Le fait de poser une question sensible à un endroit ou à un autre du questionnaire ne serait donc pas neutre et pourrait influencer aussi bien le taux de réponse que le taux d'abandon.

Des études complémentaires sont donc nécessaires pour confirmer ou non ces hypothèses.

La validité externe de la recherche concerne aussi la généralisation des résultats obtenus.

3.2.2 La robustesse et la généralisation des résultats obtenus

Les limites liées à l'extrapolation des résultats et les voies de recherche qui en découlent concernent les choix relatifs au type d'échantillon interrogé, de secteur (i.e. de catégorie de produits/services) adopté (par exemple, la téléphonie mobile dans les expériences 1 et 2), de média étudié (Internet) et de population concernée (les internautes français).

Le type d'échantillon interrogé

Mener l'étude auprès d'un échantillon de convenance, comme ce fût le cas dans notre première expérience, réduit considérablement la validité externe de la recherche. Dans notre travail, cette limite est cependant compensée par la réalisation de deux études supplémentaires, l'une sur des panélistes, l'autre auprès de clients réels. Si ces deux échantillons se rapprochent des conditions de représentativité, ils n'y sont toutefois pas entièrement parvenus. De plus, la faible taille de l'échantillon dans l'expérience 2 (N = 252) et la spécificité liée au profil de clientèle dans l'expérience 3, empêchent de pouvoir généraliser complètement nos résultats à la population internaute en France, et encore moins à l'ensemble de la population française.
Ce constat amène à envisager de nouvelles études auprès d'échantillons plus importants et plus représentatifs. En outre, l'impossibilité de mesurer les variables médiatrices du processus de réponse dans l'expérience 3, rend la réalisation de travaux complémentaires indispensable.

La catégorie de produits/services adoptée

Pour s'assurer de la validité interne de la recherche, qui constituait la préoccupation majeure de ce travail, nous n'avons étudié dans les expériences 1 et 2 qu'une seule catégorie de produits/services, à savoir la téléphonie mobile. Ce choix a été justifié dans les chapitres correspondants (i.e. 10 et 11) et répondait à une volonté de pouvoir comparer et renforcer les résultats issus de ces deux études complémentaires. L'expérience 3 est singulière puisque le choix du secteur est lié au partenariat conclu à cette occasion avec l'entreprise Millemercis.
Ces deux univers (téléphonie mobile et cadeau dans le cas de Millemercis) constituent des cas particuliers, dont les spécificités sont en mesure d'avoir influencé les résultats obtenus. Ainsi, l'implication dans le type de produits/services concernés et le niveau de risque associé au dévoilement de soi sont autant d'éléments qui diffèrent d'un secteur à l'autre et qui peuvent, si ce n'est remettre en cause nos résultats, au moins empêcher leur extrapolation à des univers différents. Westin (1991) a ainsi montré que l'attitude des consommateurs vis-à-vis de la transmission de données personnelles variait selon le domaine d'activité, tandis que Spiekermann et al. (2001) trouvent que la catégorie de produit a une influence majeure sur la capacité de dévoilement. Enfin, Nwachukwu et al. (1997) prouvent que cet effet serait particulièrement présent sur Internet, les consommateurs s'avouant moins sensibles quand ils achètent un livre en ligne que lorsqu'ils y utilisent les services d'une banque.

Une voie de recherche repose donc sur l'exploration des effets identifiés dans notre travail à des catégories de produits/services différentes, i.e. de produits de recherche ou d'expérience, à forte ou faible fréquence d'achat, pour lesquels le consommateur est plus ou moins impliqué, *etc*…

Le média étudié

Dans le cadre de ce travail, Internet constituait un excellent terrain d'application, de par l'effet supposé de ce support sur le dévoilement de soi. Les recherches sur le e-commerce et le e-marketing soulignent en effet l'importance du respect de la vie privée pour l'établissement d'une relation de confiance entre un internaute et le site visité.

Si la majorité des travaux qui étudient la fourniture de données personnelles se contentent d'analyser ce comportement à travers un média spécifique, dans le domaine de la réponse à un sondage par contre, nombreuses sont les études qui s'évertuent à comparer les réponses à travers différents médias de collecte, avec des résultats pas toujours concordants d'ailleurs. Des travaux issus de cette littérature indiquent ainsi que l'administration par ordinateur pourrait accroître la véracité des réponses aux questions sensibles car elle augmenterait l'acceptation des individus de dévoiler des faits potentiellement embarrassants. Ce constat permet alors de supposer une différence de comportement selon que l'on se trouve en (ou hors) ligne et il n'y a *a priori* aucune raison pour que cet effet ne s'applique pas également aux données personnelles. Cette conclusion est cependant démentie par les résultats issus de notre étude descriptive (présentée au chapitre 6), qui montrent que le comportement habituel de partage de données personnelles se répercute à l'identique sur Internet (cf. tableau 6.7), sans que nous puissions toutefois affirmer qu'il s'agit là d'une vérité établie et sans pouvoir le vérifier à travers nos études ultérieures.

La principale question qui se pose au sujet des résultats que nous avons pu établir dans le cadre de cette recherche est donc la suivante : les effets trouvés seraient-ils similaires hors ligne ? Selon nous, répondre à cette interrogation constitue à la fois un challenge et une nécessité et plusieurs pistes sont alors envisagées pour y parvenir. La première consisterait à introduire le média en tant que facteur manipulé, conjointement à d'autres facteurs tels que le contexte ou la quantité. La seconde serait de mener une étude qualitative (par exemple par la méthode des protocoles) afin de mieux comprendre les freins et les motivations sous-jacents à chaque média. La troisième consisterait à analyser le comportement de réponse à un formulaire comportant des données personnelles en laissant le choix du média de recueil, et ce, auprès d'un double échantillon (internautes vs. non internautes). Cette étude pourrait prendre la forme d'une analyse conjointe et consisterait alors à comparer les préférences individuelles dans chacun des groupes.

La population concernée

Se pose ici la question de la validité des résultats obtenus dans un contexte plus large : francophone, européen voire mondial. La littérature laisse supposer que les valeurs individuelles et les comportements en découlant sont spécifiques à la culture et donc au pays dans lequel se déroule l'étude. Nous pourrions donc, dans un premier temps, voir si nos résultats se confirment s'agissant d'un public francophone (i.e. Canada, Belgique, Suisse, pays du Maghreb, ..), et ce, malgré les différences (réelles et supposées) de culture entre ces différents pays. L'extension à un public européen constitue une voie complémentaire dans la mesure où, même si cela ne se retrouve pas forcément ni dans les mentalités ni dans les faits, nous sommes censés partager avec les autres pays d'Europe une culture commune. Enfin, l'extension du modèle à d'autres régions du monde constituerait l'aboutissement ultime, bien que difficile, de cette recherche.
De telles études peuvent s'envisager sous deux angles. Tout d'abord, une réplication de nos études à de nouveaux publics et une comparaison des résultats avec ceux trouvés ici. Ensuite, l'introduction au sein de notre modèle de variables spécifiques à la dimension culturelle et le test de ce nouveau modèle dans différents pays « étrangers », afin d'identifier ce qui est commun et ce qui est différent et de pouvoir en expliquer les raisons.

Pour conclure ce deuxième point, la validité externe de cette recherche est essentiellement limitée par le choix et l'opérationnalisation des facteurs étudiés ici et par la (plus précisément le manque de) robustesse des résultats obtenus, notamment de par le type d'échantillon interrogé, la catégorie de produits/services adoptée, le média étudié et le type de population concernée. Plusieurs voies de recherche sont donc envisagées qui reposent principalement sur l'analyse d'autres moyens d'opérationnaliser les facteurs manipulés ou l'étude d'autres variables situationnelles (par exemple : le caractère facultatif ou obligatoire des champs à compléter, le design et l'esthétique du formulaire ou encore l'ordre des questions posées). Enfin, l'extension de nos résultats à de nouveaux contextes d'études (autres produits/services, autres médias, autres populations) s'avère tout aussi prometteuse.

Dans les deux premières parties de cette section, nous avons insisté sur les limites liées à la validité interne et externe de notre recherche, tout en présentant les voies de recherche permettant d'y remédier. D'autres pistes de travaux futurs, indépendantes de ces limites et de portée plus générale, sont à présent détaillées.

3.3 Voies de recherche complémentaires

Outre les voies de recherche ouvertes par les limites identifiées dans le cadre de notre travail, d'autres pistes susceptibles d'être explorées à l'avenir et qui apporteraient un enrichissement certain à la compréhension des mécanismes de décision explorés ici sont à présent détaillées.
Celles-ci peuvent être classées selon la probabilité de les voir se réaliser dans un futur proche.
Elles concernent ainsi le recours à des méthodologies plus sophistiquées (3.3.1), la prise en compte de nouvelles variables (3.3.2), l'enrichissement du cadre conceptuel afin d'aboutir à un modèle dit « intégrateur » (3.3.3) et, enfin, la réalisation d'études complémentaires (3.3.4).

3.3.1 Le recours à des méthodologies plus sophistiquées

Certains des résultats obtenus dans le cadre de ce travail doctoral mériteraient le recours à des méthodologies plus sophistiquées pour en retirer le maximum de renseignements et donc leur assurer une validité supérieure.
Ainsi, dans le cadre de notre étude qualitative, nous pourrions recourir à un logiciel d'analyse textuelle afin de mieux exploiter les discours des répondants et faire émerger de manière plus explicite les liens possibles entre les différentes catégories identifiées.
De même, la typologie réalisée dans le cadre de notre étude exploratoire descriptive pourrait faire l'objet d'une analyse de régression par classes latentes, afin de faire émerger les groupes de manière plus naturelle. De plus, le caractère discriminant de l'appartenance aux classes, notamment selon des critères sociodémographiques, gagnerait à faire l'objet d'une analyse multivariée, sous la forme d'une analyse des correspondances multiples.
Les analyses de variance et de régressions réalisées dans le cadre des expériences 1 et 2 pourraient également être remplacées par la méthode des équations structurelles, sous réserve de prendre les précautions nécessaires et de disposer du nombre de degré de libertés suffisant. Dans cette optique, on pourrait adapter cette technique à la logique expérimentale et inclure par conséquent, outre les variables clés du processus de réponse (perceptions, attitude et intentions comportementales), les facteurs manipulés et les construits correspondants aux « *Manipulation Checks* », suivant ainsi les recommandations de Bagozzi (1977) et de MacKenzie (2001). Une telle analyse, qui aurait surtout été intéressante dans le cadre de notre expérience 2, n'a pas pu être réalisée ici, faute de disposer de trois items par construit.

Enfin, concernant les résultats de l'expérience 3, nous pourrions envisager de réaliser plusieurs régressions logistiques simultanées sous la forme de modèles emboîtés. Dans le cadre de l'étude du comportement d'abandon, une analyse de survie serait également opportune.

3.3.2 La prise en compte de nouvelles variables

Nous ne parlerons pas ici des variables indépendantes de notre modèle (facteurs situationnels manipulés) dans la mesure où ce point a déjà été abordé précédemment (paragraphe 3.2.1).
Dans le même ordre d'idée, nous aurions pu évoquer la possibilité de prendre en compte d'autres variables dépendantes comme le fait de fournir des données erronées, la qualité ou la richesse des réponses fournies (i.e. nombre de mots employés dans la réponse à une question ouverte) ou encore l'intention (voire l'action) de revenir sur le site. Nous ne nous y attarderons toutefois pas puisque, bien que de nature à enrichir le modèle, elles n'apporteraient pas d'éclairage nouveau concernant le processus de décision du consommateur.

Nous avions indiqué, au début du chapitre 7, que l'impact de la sollicitation de données sur le comportement de réponse dépendait de l'habileté des consommateurs à traiter, percevoir et juger cette situation. La validité de cette proposition n'a pas été réellement abordée dans le cadre de notre recherche. Afin d'en évaluer la portée exacte, nous estimons qu'il serait nécessaire d'inclure de nouvelles variables de nature individuelle au sein du modèle, soit en tant que modérateurs, soit en tant que variables de contrôle.
La prise en compte d'autres variables individuelles susceptibles d'agir sur notre modèle se heurte toutefois à des problèmes de faisabilité. Nous ne présenterons donc ici que les variables qui nous semblent pouvoir apporter un éclairage intéressant aux résultats trouvés. Celles-ci concernent notamment les traits de personnalité et certaines variables de nature attitudinale.

Les traits de personnalité

Même si leur impact sur le dévoilement de soi est sujet à controverse, nous pensons que l'intégration de variables de personnalité au sein du modèle serait de nature à améliorer la compréhension du processus de décision du consommateur confronté à une sollicitation de ses données personnelles.

Selon Dibb et al. (1994), la personnalité concerne tous les traits et comportements qui font qu'une personne est unique. Dans leur théorie de la personnalité (*Trait Theory*), Costa et McCrae (1992) distinguent ainsi 5 traits principaux : l'extraversion, le neuroticisme, le caractère agréable, le caractère consciencieux et l'ouverture à l'expérience[360]. Nous nous intéresserons plus particulièrement au premier d'entre eux puisqu'il est déjà largement étudié dans la littérature. Ainsi, parce que les extravertis sont plus sociables et davantage tournés vers le monde extérieur (Carver et Scheier 1992, Gleitman 1995), on peut penser qu'ils seront plus enclins à faire confiance aux sites Internet en ce qui concerne leurs pratiques de gestion des données et qu'ils auront alors une plus forte tendance à percevoir la confidentialité de celles-ci.
Ce trait de personnalité peut donc être considéré comme une orientation générale vers le contrôle de ses données personnelles (Stone 1986) et l'ajout de cette variable au sein de notre modèle pourrait alors s'avérer intéressant.

Un autre trait de personnalité qui pourrait se révéler décisif dans le cadre de notre sujet de recherche concerne le locus de contrôle (en anglais « *locus of control* »). D'après Rotter (1966), ce concept peut être défini comme la croyance des individus selon laquelle les événements de leur vie sont contrôlés par des facteurs internes ou externes. Une personne à fort degré de contrôle interne attribuera ainsi les causes d'un événement à son propre effort, tandis que les personnes à fort degré de contrôle externe pensent généralement que les résultats sont dus à des variables extérieures comme le manque de chance ou l'intervention d'autres personnes.
Les personnes au locus interne croient donc en leurs propres capacités à trouver le comportement nécessaire pour contrôler les événements (Phillips et Gully 1997). A l'opposé, les « externes » ne se croient pas munis des capacités nécessaires pour résoudre les problèmes (Larson et al. 1990). Ils pourraient alors ne pas désirer avoir de contrôle sur les événements parce qu'ils pensent que cela est impossible.
Cette variable est donc liée à l'attitude du consommateur quant au pouvoir dont il estime disposer pour influencer les résultats d'une action et, en ce sens, elle pourrait avoir un impact décisif sur le processus de réponse à une sollicitation de données personnelles. Bien qu'ayant reçu assez peu d'attention dans la littérature en comportement du consommateur, nous estimons qu'elle serait fort utile à la compréhension du phénomène étudié ici.

En plus de ces deux traits de personnalité, nous pourrions également ajouter celui de la propension à faire confiance, en raison de son lien évident avec le sujet.

[360] En anglais : *extraversion, neuroticism, agreeableness, conscientiousness, openness to experience.*

Enfin, de manière plus anecdotique, nous pourrions intégrer le concept de « dissimulation de soi » (en anglais *self concealment*) qui explique pourquoi les gens cherchent à « protéger » (garder secret) les informations jugées intimes ou potentiellement menaçantes. Le fait de se montrer conciliant face aux sollicitations (en anglais *compliance*) est également employé par les chercheurs qui s'intéressent à la compréhension de la participation aux sondages et pourrait donc être étudié dans le cadre de la réponse à une sollicitation de données personnelles.

Les variables de nature attitudinale

Plusieurs variables de nature attitudinale nous semblent devoir être prises en compte à l'avenir, dans le cadre de la compréhension du processus de réponse à une sollicitation de données personnelles. Nous pensons notamment aux concepts d'aversion au risque, d'attitude face aux gains et aux pertes et d'attitude face à la technologie en général, et face à Internet en particulier.

D'autres variables individuelles

Plusieurs facteurs individuels, autres que les traits de personnalité et les variables attitudinales, pourraient également être ajoutés à notre modèle. Parmi eux, nous pensons plus spécifiquement à l'expertise d'Internet (dans le cadre de l'étude de ce média), au style cognitif et à l'orientation relationnelle des individus. Nous détaillons plus particulièrement les deux premiers.

L'expertise s'avère être une variable essentielle dans les recherches sur le comportement de l'internaute (Wallet Wodka 2003). Ce concept est ainsi souvent étudié comme un élément modérateur du comportement (Boulaire et Balloffet 1999, O'Cass et Fenech 2003). Cependant, il ne fait pas l'objet de mesures unifiées, certains chercheurs l'assimilant à la familiarité, d'autres à la fréquentation d'Internet, tandis que quelques uns distinguent expertise subjective et objective. Malgré ce problème de mesure, nous estimons que l'intégration de l'expertise d'Internet en tant que variable modératrice du processus de réponse constituerait un apport significatif à notre modèle.

Par ailleurs, nous pourrions nous interroger sur le rôle modérateur du besoin de cognition des consommateurs et élargir le champ de recherche autour des capacités cognitives des individus. Le style cognitif reflète en effet le processus d'acquisition, d'analyse, d'évaluation et d'interprétation des informations par l'individu au cours d'une prise de décision. Il y a donc des

chances pour que cette variable soit en mesure d'influencer le processus étudié ici et que son introduction soit susceptible de modifier certaines des relations qui ont été testées.

Une troisième voie de recherche complémentaire concerne l'enrichissement de notre cadre conceptuel afin d'aboutir à un modèle dit « intégrateur ».

3.3.3 Vers un modèle intégrateur

Nous distinguons cette piste de la précédente dans le sens où nous ne nous intéressons pas ici à l'ajout de variables « isolées » mais plutôt à l'enrichissement du cadre conceptuel, notamment à la lumière d'autres théories, afin d'avoir une vision plus complète du phénomène étudié. La première voie d'amélioration concerne la possibilité de compléter le modèle actuel au regard de la théorie du comportement planifié (modèle TPB) exposée au chapitre 7, et de le tester dans son intégralité. La seconde consiste à repenser le modèle à la lumière de nouvelles théories non évoquées jusqu'ici.

Compléter le modèle actuel et l'étudier dans son intégralité

Utiliser le modèle TPB plutôt que TRA comme ancrage théorique de cette recherche nous aurait permis d'introduire la notion de contrôle perçu, dont l'impact pourrait s'avérer important, compte tenu du sujet étudié ici[361]. L'ajout de cette variable et l'analyse de l'ensemble des liens susceptibles d'exister entre les différentes étapes du processus de réponse (notamment entre les intentions comportementales et le comportement) consisterait une première voie d'amélioration, somme toute substantielle, de notre modèle. A cela, nous pourrions ajouter l'étude de tous les liens pouvant exister entre l'exposition à la sollicitation et son évaluation[362], afin d'avoir une vision encore plus précise du phénomène. Enfin, en termes d'attitude, la littérature suggère que les individus peuvent avoir des réponses affectives différentes vis-à-vis : a) de la sollicitation elle-même, b) des individus (dans notre cas de l'entreprise) chargés de la collecte et c) des procédures utilisées pour la collecte (Stone et Stone 1990). Il serait donc intéressant de pouvoir différencier ces trois composantes au sein de notre modèle.

[361] Ce sujet a été évoqué au chapitre 7 et les raisons du choix du modèle TRA y ont été précisées.
[362] Dans notre recherche, seul l'effet du contexte et de la quantité sur la valeur perçue a par exemple été testé.

Repenser le modèle à la lumière d'autres théories

Une première piste consisterait à tester un modèle compréhensif du comportement de réponse d'un individu à une sollicitation de données personnelles en y incluant des théories psychosociologiques largement mobilisées dans les recherches portant sur la réponse à un sondage telles que celles de l'attribution, de l'auto-perception (*self perception*) (Hansen 1980) ou de la dissonance cognitive (Furse et Stewart 1982).

Une seconde piste vise à repenser le modèle de décision testé dans le cadre de cette recherche.
Au cours de ce travail en effet, nous nous sommes exclusivement intéressés au processus perceptuel (attention et interprétation du stimulus) et à la prise de décision en découlant. Dans la réalité, selon le comportement étudié, plusieurs processus de nature distincte interagissent pour former un jugement. Ainsi, dans le cadre de la thèse, nous avons mis de côté le processus de mémorisation (encodage, stockage et récupération de l'information), qui pourrait pourtant intervenir, notamment si la situation proposée se rapproche d'une expérience vécue précédemment par l'individu. Dans ce cas en effet, il est fort possible que l'expérience passée ait un impact plus important que les caractéristiques intrinsèques de la nouvelle situation et que la décision en découlant semble alors, de prime abord, contradictoire.
De plus, notre modèle se fonde essentiellement sur une approche rationnelle de la prise de décision. La littérature relative au processus de décision du consommateur s'inscrit en effet, principalement et originellement, dans la théorie classique de la décision qui stipule que les individus sont rationnels, disposent d'une information parfaite et maximisent leur utilité. Or, plusieurs recherches, menées notamment sur notre sujet d'étude, signalent les entorses, parfois nombreuses, à ce principe de rationalité[363]. Elles indiquent ainsi plusieurs éléments (i.e. asymétries d'informations, interprétations subjectives et distorsions psychologiques) susceptibles de conduire à une décision contraire à ce qu'on pouvait attendre.
Certains résultats obtenus dans le cadre de notre travail semblent d'ailleurs confirmer ces dires. Ceci conduit alors à penser que, dans certains cas (qu'il conviendra d'identifier), la prise de décision du consommateur à l'égard d'une sollicitation de données personnelles pourrait être altérée par rapport à une situation de choix rationnelle. En particulier, trois éléments relatifs à la prise de décision pourraient être concernés (cf. tableau D.8 ci-après).

[363] Nous avons notamment évoqué ce sujet dans le chapitre 4 de ce document (paragraphe 1.3).

Tableau D.8 Les éléments de la prise de décision susceptibles d'être « altérés »

L'objet de décision : la fourniture d'informations personnelles	La fourniture de données personnelles est un objet de décision complexe : Le formulaire est à la fois un élément objectif et subjectif qui agit sur des dimensions cognitives et/ou affectives. Il est également à la fois un stimulus qui doit être traité, une information sur la sollicitation, un critère d'évaluation et une contrainte puisqu'il faut le remplir.
Le preneur de décision : le consommateur	Le consommateur est un preneur de décision « faillible » : Les capacités de détection, de stockage et d'extraction des informations contenues dans un formulaire sont limitées. La facilité et la rapidité de traitement des informations dépendent aussi des conditions dans lesquelles se trouve l'individu quand il est sollicité.
La règle de décision	La règle de décision dépend des effets de contexte : Les motivations et les capacités des individus pour traiter les informations déterminent le recours à une analyse systématique ou heuristique de la sollicitation

Pour une meilleure compréhension du phénomène, toute tentative visant à évaluer l'effet des caractéristiques de la sollicitation sur le processus de réponse devrait donc prendre en compte ces différents éléments, voire les mesurer, afin d'identifier leur impact potentiel.

Dans cette optique, utiliser la « *leverage saliency theory* » proposée par Groves et al. (1992) pour expliquer la participation à un sondage pourrait s'avérer profitable. Groves et couper (1998) montrent ainsi que, face à ce type de sollicitation, les individus varient en termes d'attributs jugés pertinents pour prendre leur décision. Ces attributs dépendent notamment de leur expérience dans ce domaine et ne sont pas connus à l'avance par l'entreprise.

Pour mieux expliquer la décision de l'individu de participer (ou non), il faut donc tenir compte de deux critères complémentaires : 1) l'importance accordée à chaque attribut de la sollicitation (exemple : design, avantages, …) (ce que les auteurs appellent la « moyenne ») et 2) la saillance accordée à ces attributs durant la sollicitation. Chaque attribut peut donc avoir une « moyenne » positive ou négative, celle-ci reflétant alors une prédisposition favorable ou défavorable à participer. Dès lors, l'influence potentielle d'une caractéristique de la sollicitation serait fonction de l'importance accordée par le répondant à cette caractéristique, du fait que cette influence soit positive ou négative et de la manière avec laquelle cette caractéristique est communiquée et mise en avant au cours de la sollicitation (saillance).

Les principaux problèmes qui se posent dans l'application de cette théorie résident dans le fait qu'il est possible de trouver, au sein de l'échantillon, des gens accordant une importance positive à un attribut tandis que d'autres lui donnent un effet négatif. Il faut aussi noter l'existence d'un effet de seuil dans la mesure où l'effet d'un attribut à l'importance positive est plus faible quand la probabilité de coopération est proche de un (individu favorablement prédisposé). Il existerait

aussi un effet de contrepoids par lequel un attribut à l'importance positive (i.e. un avantage) pourrait contrer l'effet d'un attribut à l'importance négative (par exemple, le manque d'intérêt).

3.3.4 La réalisation d'études complémentaires

Des études complémentaires peuvent aussi être envisagées pour approfondir la connaissance des mécanismes aboutissant à la décision de l'individu de fournir ou non les données demandées. Ci-dessous, nous développerons plus particulièrement trois d'entre elles.

Une étude axée sur la question de la sensibilité des données

L'importance de la sensibilité des données demandées, largement développée dans la littérature, est confirmée par les résultats de notre recherche. Malgré cela, il manque encore d'études axées sur la compréhension de ce que les individus considèrent comme une donnée sensible et sur la mesure de ses effets sur le processus de décision. Notre recherche apporte des premiers éléments de réponse, qu'il conviendrait de confirmer et d'enrichir par des travaux complémentaires. Dans le prolongement de notre troisième expérience, une piste intéressante serait ainsi d'étudier l'effet de la présence d'une question sensible, de son type (i.e. n° de téléphone, revenus, …), de sa place au sein du formulaire et du caractère obligatoire ou non de la réponse sur différents éléments de la prise de décision (i.e. inscription, taux de réponse et abandon notamment).

Une étude longitudinale

Une autre piste prometteuse concerne la possibilité d'avoir une vision longitudinale du processus de dévoilement de soi, plutôt que de réaliser des études en coupe instantanée comme ce fut le cas ici. La littérature signale en effet que le dévoilement a plus de chance d'avoir lieu si les sollicitations se font graduellement en suivant l'évolution de la relation, depuis les échanges occasionnels, jusqu'à une relation plus intime (Altman et Taylor 1973, Derlega et al. 1993). De plus, si l'autre requiert une divulgation trop tôt dans la relation, les principes de réciprocité et de transfert ont alors des chances de ne pas fonctionner (Collins et Miller 1994, Cozby 1973).
Les résultats de notre recherche confirment l'effet potentiellement dommageable d'une sollicitation jugée abusive et/ou inappropriée et encouragent l'entreprise à se demander ce qu'il est possible de réclamer à chaque client compte tenu de la relation qu'elle entretient avec lui. C'est donc ici toute la stratégie de collecte et de management des informations clients qui doit

être (re)pensée de manière à obtenir des informations de qualité (non erronées) et en nombre suffisant, tout en répondant aux craintes et aux exigences des consommateurs.
Comment construire une vision progressive du client ? Comment définir le chemin de la connaissance du client ? Telles sont les principales interrogations qui se posent à ce sujet.
Malgré les difficultés liées à sa mise en place, une étude longitudinale menée en partenariat avec une entreprise auprès d'un échantillon de clients suivis sur plusieurs mois (depuis leur qualité de prospect à celui de client « engagé »), permettrait de répondre en partie à ce type de questions.

Une étude axée sur la perception éthique de l'échange

Une dernière piste envisagée concerne la possibilité d'étudier la réponse à une sollicitation de données personnelles sous l'angle éthique. La littérature souligne en effet qu'au fur et à mesure que l'on passe d'un échange transactionnel à un échange relationnel, l'éthique prend de plus en plus d'importance, supplantant alors les protections légales (Gundlach et Murphy 1993).
Les modèles de comportement éthique se basent généralement sur l'idée que « faire de l'éthique est une bonne chose »[364], autrement dit que le consommateur choisira en priorité de commercer avec un vendeur honnête. Malgré son importance soupçonnée, la perception par le consommateur d'un comportement (non) éthique de l'entreprise a été assez peu étudiée jusqu'ici, que ce soit dans la littérature sur l'éthique ou en comportement du consommateur. Or, selon Whalen, Pitts et Wong (1991), il est raisonnable de penser que, lors du processus de décision, le consommateur ajoute des jugements éthiques aux autres caractéristiques évaluées (comme les coûts et bénéfices perçus). De même, le consommateur pourrait considérer le comportement éthique passé du vendeur comme critère d'évaluation du comportement éthique futur et comme critère de décision dans les intentions d'établir ou de continuer une relation avec lui.
Dans le cas qui nous intéresse ici, ce jugement éthique pourrait donc être un critère primordial dans l'évaluation de la sollicitation et avoir une incidence majeure sur l'intention de poursuivre la relation avec l'entreprise ayant sollicité les données. En s'appuyant sur la littérature en éthique managériale, Whalen et al. (1991) développent à cet égard le construit d'attribution éthique (correspondant à la perception du comportement éthique du vendeur) et proposent de l'intégrer en tant que partie prenante du processus de décision. Nous pensons qu'un tel construit serait tout à fait approprié à notre modèle.

[364] En anglais : « *ethical business is good business* » (Abratt et Sacks 1988)

Pour conclure ce troisième point, nous avons exposé un certain nombre de pistes de recherche, indépendantes des limites qui ont pu être identifiées précédemment.

Celles-ci concernent notamment le recours à des méthodologies plus sophistiquées (par exemple : régression par classes latentes et modèle d'équations structurelles) pour analyser, de manière approfondie, certains résultats trouvés dans le cadre de ce travail doctoral.

Dans une optique d'amélioration de notre modèle, nous recommandons aussi l'intégration de nouvelles variables de nature individuelle, comme certains traits de personnalité (i.e. degré d'ex-traversion, locus de contrôle) ou le style cognitif. Nous souhaiterions aussi enrichir ce modèle par la prise en compte de théories complémentaires à celles mobilisées ici, telles que la théorie du comportement planifié (Ajzen 1991) ou la « *leverage saliency theory* » (Groves et al. 1992).

Enfin, nous proposons la réalisation d'études complémentaires axées sur la sensibilité des données et sur une vision longitudinale et éthique de l'échange.

Synthèse de la troisième section

Outre les contraintes dues à la nature expérimentale du processus de validation du modèle, les limites liées à la validité interne de cette recherche concernent à la fois la mise en situation des sujets expérimentaux et l'existence de biais inhérents au sujet étudié et au dispositif utilisé. Celles-ci constituent autant de pistes d'amélioration en vue de travaux futurs, notamment à travers la prise en compte de nouvelles variables de contrôle et la mesure du comportement de mensonge et du biais de désirabilité sociale.

Notre recherche a permis de valider l'existence d'un processus de réponse à une sollicitation de données personnelles et souligne l'impact majeur des variables situationnelles sur ce processus.

Pour tester la robustesse du modèle, ce travail devrait cependant être étendu à d'autres facteurs situationnels (ou à une opérationnalisation différente de ceux étudiés ici), d'autres catégories de produits/services, d'autres médias et d'autres populations.

Par ailleurs, dans l'optique d'affiner les résultats de la recherche, le recours à des méthodologies sophistiquées, à de nouvelles variables individuelles et à des théories différentes pourrait être pertinent. Enfin, il serait intéressant de se pencher sur les problématiques de sensibilité des données et de comportement éthique de l'entreprise, voire d'adopter une vision dynamique de l'échange de données personnelles en menant une étude longitudinale.

Conclusion de la discussion

Cette dernière partie de la thèse a permis de discuter des résultats de la recherche et d'en déduire les principaux enseignements.

Au regard des recherches passées en matière de dévoilement de soi, de réponse à un sondage et de préoccupation pour le respect de la vie privée, nous sommes en mesure de dire que la réponse du consommateur à une sollicitation de données à caractère personnel fait l'objet d'un processus de décision de nature essentiellement cognitive, qui comprend différentes phases depuis l'exposition à la collecte jusqu'au comportement de réponse en passant par l'évaluation de la sollicitation, l'attitude et les intentions comportementales.

Les résultats permettent aussi d'avancer que l'impact des facteurs situationnels sur ce processus de réponse est prépondérant et dépasse, de loin, celui lié aux variables individuelles.
Ces constats sont, dans l'ensemble, cohérents avec ceux issus de la littérature et apportent, pour certains d'entre eux, des éclairages complémentaires. Nous montrons ainsi qu'une sollicitation jugée abusive peut avoir des conséquences dommageables sur la relation avec l'entreprise à l'origine de la requête.

Les apports de ce travail sont principalement d'ordre académique et managérial et, dans une moindre mesure, de portée empirique et méthodologique. Des conclusions théoriques liées au processus de décision adopté par l'individu en réponse à une sollicitation de ses données personnelles et aux facteurs susceptibles de l'influencer ont ainsi été avancées.
Des conclusions plus appliquées ont aussi été proposées concernant les critères majoritairement utilisés par les consommateurs pour évaluer la sollicitation et les solutions à mettre en place, pour les rassurer d'une part, et éviter tout risque de rupture de la relation, d'autre part.

Les limites de cette recherche concernent d'abord la validité interne. Toutefois, elles sont surtout liées à la validité externe, ce qui conforte l'idée que la robustesse des effets trouvés doit être testée ultérieurement. Elles représentent donc autant de voies de recherche pour le futur.
Enfin, des pistes de travaux complémentaires se dessinent également, notamment en vue de compléter et de solidifier le modèle testé ici. La réalisation d'études axées sur la sensibilité des données et sur une vision longitudinale et éthique de l'échange apparaît tout aussi prometteuse.

CONCLUSION GÉNÉRALE

« Là où il y a conclusion, il y a critique. »

Rémy de Gourmont

INTRODUCTION GÉNÉRALE

↓

PARTIE I – COMPRENDRE
Collecte de données clients et Respect de la vie privée
Les apports de la littérature

- Chapitre 1 : Les enjeux liés à la collecte de données clients
- Chapitre 2 : Préambule sur la notion de vie privée
- Chapitre 3 : Dévoilement de soi et divulgation de données personnelles
- Chapitre 4 : Sollicitation de données personnelles et réactions des consommateurs : synthèse de la littérature

↓

PARTIE II – EXPLORER et BATIR
Sollicitation de données personnelles et réactions des consommateurs
Des études exploratoires à la méthodologie de la recherche

- Chapitre 5 : Processus de réponse face à une sollicitation de données personnelles : exploration qualitative (étude 1)
- Chapitre 6 : La gestion des données personnelles chez le consommateur-internaute : exploration descriptive (étude 2)
- Chapitre 7 : Formulation du cadre conceptuel général de la recherche
- Chapitre 8 : Méthodologie de la recherche
- Chapitre 9 : Choix et développement des instruments de mesure

↓

PARTIE III – EXPLIQUER
Attitude et réponses du consommateur face à une sollicitation de données personnelles en ligne
Analyse empirique des résultats

- Chapitre 10 : Expliquer l'attitude du consommateur face à la communication de ses données personnelles (étude 3)
- Chapitre 11 : Sollicitation de données personnelles et intentions comportementales de réponse (étude 4)
- Chapitre 12 : Réponses du consommateur face à une sollicitation de ses données personnelles (étude 5)

↓

DISCUSSION ET MISE EN PERSPECTIVE DES RESULTATS DE LA RECHERCHE

↓

CONCLUSION GÉNÉRALE

En matière de réponse du consommateur face à une sollicitation d'informations personnelles, les interrogations des professionnels et les chantiers de recherche sont nombreux. En particulier, les managers veulent comprendre la façon dont les individus sont susceptibles de réagir face aux collectes de données dont ils peuvent faire l'objet et s'interrogent sur les méthodes à utiliser pour qu'ils acceptent plus facilement de délivrer ces renseignements.

Les réponses que les chercheurs ont pu jusqu'à présent apporter restent encore limitées puisqu'on ne sait toujours pas de manière explicite ce que le consommateur considère comme une invasion de sa vie privée ni comment il gère la diffusion de ses données personnelles. De plus, les explications théoriques avancées sont encore pauvres alors même que la multiplication des occasions de recueillir ce type d'informations, notamment depuis l'arrivée d'Internet, et les préoccupations croissantes pour le respect de la vie privée rendent nécessaires une meilleure compréhension et conceptualisation du processus de décision du consommateur face à ce type de sollicitation. Ces interrogations sont prégnantes en France au vu du nombre restreint (voire inexistant) de travaux empiriques qui traitent de ce sujet.

Afin de dresser un bilan de notre travail, il est utile de revenir sur notre questionnement de départ et d'expliquer la démarche que nous avons suivie pour tenter d'y répondre.
La problématique de la recherche était la suivante : **Comment expliquer la réponse du consommateur face à une sollicitation de ses données personnelles ?**
Nous nous intéressions ainsi plus particulièrement à deux points principaux :
Qu'est-ce qui influence la décision de fournir (ou non) des données à une entreprise ?
Comment (i.e. en fonction de quels critères) cette décision se prend elle ?

Cette problématique a donné lieu à plusieurs questions de recherche exposées dans l'introduction et auxquelles nous avons cherché à répondre à travers l'ensemble des études menées.

L'approche que nous avons utilisée s'inscrit dans un paradigme hypothético-déductif classique. Sur la base d'une revue de la littérature et de deux études exploratoires, nous avons eu recours aux théories qui nous semblaient pertinentes par rapport à nos questions de recherche, afin de bâtir le cadre conceptuel et de proposer des hypothèses pertinentes. Nous avons alors conçu, mis au point et appliqué une méthodologie pour tester ces hypothèses.
Nous synthétisons ci-après cette démarche et les réponses que nous avons pu apporter.

La démarche de recherche suivie

Tout au long de ce travail doctoral, nous avons poursuivi un triple objectif : comprendre, tout d'abord, explorer/bâtir ensuite et expliquer, enfin, le processus de décision du consommateur qui se trouve sollicité pour fournir des données à caractère personnel ainsi que les facteurs susceptibles d'influencer ce processus. Nous détaillons ci-après chacun de ces objectifs.

1. Comprendre les fondements théoriques à l'origine de la réponse du consommateur

Dans un premier temps, après avoir identifié les enjeux liés à la collecte de données clients, notre travail s'est attaché à comprendre les fondements théoriques de la réponse du consommateur face à une sollicitation de ses données personnelles. Nous avons alors cherché à réaliser un état de l'art pertinent et mené pour cela une triple revue de la littérature.

Nous avons commencé par étudier la notion de vie privée et, en particulier, de préoccupation pour le respect de la vie privée, qui constitue l'un des concepts clés de cette recherche.
Nous avons poursuivi notre travail par une analyse des mécanismes et des facteurs à l'origine du dévoilement de soi, dans le domaine interpersonnel.
Enfin, pour avoir une vision globale de la réponse à une sollicitation de données personnelles, nous avons souhaité ne pas restreindre l'étude du phénomène à son seul processus. Il semblait alors important de prendre en compte l'ensemble des facteurs susceptibles de l'influencer. Le chapitre 4 a ainsi été l'occasion de réaliser une revue de la littérature sur l'impact de ces variables et de faire le point sur les connaissances établies sur le sujet. Ces facteurs sont principalement de deux ordres (situationnels et individuels), les facteurs environnementaux pouvant être considérés comme plus accessoires.
La prise en compte des éléments situationnels relève presque de l'obligation. Puisque nous étudions la réaction du consommateur face à une sollicitation, il est normal que les variables qui caractérisent cette situation jouent un grand rôle. Toutefois, la littérature fait peu référence à ce type de facteurs, préférant s'intéresser aux variables individuelles. Nous avons alors décidé de nous focaliser plus particulièrement sur l'étude de ces facteurs situationnels.

2. Explorer (afin de connaître les critères mobilisés par le consommateur pour prendre sa décision) et Bâtir le cadre conceptuel de la recherche

Pour répondre à ce deuxième objectif, deux études exploratoires ont tout d'abord été menées.

L'étude qualitative a permis de répertorier les critères majoritairement utilisés par les consommateurs pour évaluer la sollicitation et prendre leur décision. Il s'agit de la confidentialité, de la sensibilité et de la pertinence perçues des données sollicitées ainsi que de la valeur perçue à répondre. Nous avons également esquissé un modèle provisoire qui vise à souligner le caractère processuel de la réponse de l'individu à une sollicitation de données et à identifier l'impact respectif des facteurs individuels, situationnels et environnementaux.

Une étude descriptive s'en est suivie, afin de comprendre l'incidence des préoccupations de vie privée sur les comportements de partage de données personnelles. Il s'agissait aussi et surtout d'étudier le type d'avantages susceptibles de favoriser la divulgation avec, pour but, de dresser une typologie des internautes français en la matière. Quatre groupes d'individus aux attitudes bien distinctes ont ainsi pu être identifiés, qui se différencient sur des éléments sociodémographiques, expérientiels et comportementaux. Enfin, l'étude fait apparaître que la familiarité avec l'organisme à l'origine de la requête et le contexte dans lequel les informations sont demandées, seraient des éléments sur lesquels les individus s'appuient, afin de décider s'ils fourniront ou non les renseignements requis.

Au regard de la littérature et de ces deux études exploratoires, nous avons alors pu répertorier les principaux facteurs susceptibles d'influencer le processus de décision du consommateur. Nous avons, en particulier, identifié les éléments suivants :
- variable individuelle : la préoccupation pour le respect de la vie privée
- variables situationnelles : la politique de respect de la vie privée adoptée par l'entreprise, la nature de la relation avec l'entreprise sollicitante, les caractéristiques du dispositif de collecte, et enfin, les circonstances dans lesquelles intervient la sollicitation.

Un premier modèle conceptuel, de portée générale, a alors été proposé. Celui-ci s'appuie principalement sur la théorie de l'action raisonnée et, de manière plus accessoire, sur les théories de l'échange social, de la motivation ou de l'espérance attendue et de l'engagement. Ce faisant, nous avons adopté une approche théorique qui permet de considérer que la réponse de l'individu résulte d'un processus essentiellement cognitif de traitement de l'information. Cette démarche permet d'expliquer le comportement de réponse en le schématisant sous la forme d'une séquence linéaire où se succèdent différentes phases (évaluation, attitude, intention comportementale et réponse), et ce, même s'il est tout à fait possible que le processus ne se déroule pas de façon aussi figée dans la réalité.

La méthodologie choisie pour tester les hypothèses de la recherche a ensuite été présentée. Un deuxième modèle conceptuel, dit modèle général à tester, a alors été proposé en tenant compte de la méthodologie utilisée (i.e. méthode expérimentale) et des facteurs situationnels étudiés. Pour garantir la validité interne de la recherche, nous avons ainsi opté pour un design original qui repose sur trois expériences successives, chacune focalisée sur une partie du modèle. Cette solution présente en outre l'avantage de limiter alternativement chacun des biais identifiés. Une première expérience, conduite sur des étudiants, visait ainsi à expliquer l'attitude du consommateur face à la communication de données personnelles. La deuxième, réalisée sur des panélistes, s'attachait à mesurer les intentions comportementales de réponse. La troisième a été menée sur le terrain, auprès de clients réels, grâce au concours de la société Millemercis.

Au total, quatre facteurs situationnels ont été manipulés au sein du dispositif expérimental :
- les mentions de respect de la vie privée (RVP) figurant sur la page du formulaire à remplir
- la familiarité avec l'entreprise à l'origine de la sollicitation
- la quantité de données demandées, correspondant au nombre de champs sur le formulaire
- le contexte de sollicitation (et le type d'avantages proposés en échange des données)

L'opérationnalisation des variables présentes dans le modèle a ensuite consisté à développer puis valider un ensemble d'échelles de mesure. Ces instruments ont été élaborés puis testés au cours de plusieurs études menées auprès d'experts et de consommateurs afin de s'assurer de leurs propriétés psychométriques (validité et fiabilité) et de leur qualité opérationnelle.

Enfin, pour s'assurer de la validité de l'expérimentation et vérifier l'adéquation des données aux tests statistiques pressentis, les données collectées ont fait l'objet d'une batterie de tests préliminaires. Les résultats de ces tests (indépendance des observations, normalité, égalité des cellules expérimentales, homogénéité des variances) s'étant révélés satisfaisants dans nos trois expériences, nous avons pu procéder à chaque fois à l'analyse des données.

3. Expliquer l'influence des différents facteurs sur le processus de décision

Comme indiqué précédemment, pour tester nos hypothèses de recherche, trois expériences successives ont été menées, auprès d'échantillons différents et dans des contextes variés. Celles-ci ont été précédées de deux études exploratoires, l'une qualitative, l'autre descriptive. Nous synthétisons dans le tableau ci-après (tableau C.1) l'objectif, l'échantillon, la démarche suivie et les principales conclusions auxquelles chacune de ces études a permis d'aboutir.

Tableau C.1 Récapitulatif des études réalisées dans le cadre de ce travail doctoral (*)

N°	Type d'étude	Objectifs	Caractéristiques de l'échantillon	Démarche	Principaux résultats
1	ETUDE QUALITATIVE EXPLORATOIRE Entretiens semi-directifs	- comprendre les freins et motivations des internautes français face à la divulgation de leurs données personnelles - recueillir des verbatims en vue du développement d'échelles	9 internautes de profil différent au niveau sociodémographique et de l'utilisation d'Internet	- élaboration du guide d'entretien - recrutement des interviewés - entretiens et retranscription - analyse de contenu - re-bouclage avec la littérature - élaboration d'un modèle conceptuel provisoire	- les facteurs individuels et situationnels n'agissent pas au même niveau du processus de réponse - 4 éléments sur laquelle se fonde l'évaluation : confidentialité, sensibilité, pertinence, gain net à répondre
2	ETUDE DESCRIPTIVE EXPLORATOIRE (Etude FING)	- décrire les pratiques des internautes en matière de divulgation de données personnelles - Etablir une typologie	1364 internautes français recrutés sur les sites fing.org et secuser.com (échantillon technophile)	- élaboration du questionnaire - recrutement des participants - analyse des réponses : typologie et analyses univariées (tests de différences de moyennes)	4 classes d'individus qui diffèrent quant à leur manière de gérer leur identité en ligne - des différences attitudinales et comportementales entre les groupes
3	ETUDE QUANTITATIVE Expérimentation 1 (EXP 1)	- mesurer l'impact des facteurs situationnels sur l'évaluation de la sollicitation et l'attitude face à la communication des données	232 étudiants de Master	- élaboration des scénarios et du questionnaire - passation du questionnaire - analyse des réponses : MANCOVA et régressions	- le début du processus de réponse est validé - la politique de RVP et la quantité de données demandées ont une influence
4	ETUDE QUANTITATIVE Expérimentation 2 (EXP 2)	- mesurer les intentions comportementales de réponse - étudier l'effet médiateur de l'attitude	252 panélistes recrutés par Panel on the web (échantillon représentatif des internautes français)	- élaboration des scénarios et du questionnaire - passation du questionnaire - analyse des réponses : ANCOVA et régressions	- la suite du processus de réponse est validée - le contexte de sollicitation n'a pas d'effet sur la valeur perçue à répondre
5	ETUDE QUANTITATIVE Expérimentation 3 (EXP 3)	- mesurer le comportement de réponse dans un contexte réel	Plus de 20 000 clients du site Millemercis ayant cliqué sur le lien fourni dans l'email de recrutement	- élaboration du questionnaire - définition des cibles et « sélection » de l'échantillon - analyse des réponses : régressions logistiques et linéaires	- l'effet de la familiarité et du contexte de sollicitation est difficile à interpréter - la quantité a bien un effet sur le comportement de réponse

(*) Dans ce tableau n'apparaissent ni les études « exploratoires » ayant servi à purifier les instruments de mesure (analyses factorielles exploratoires) ni les études nécessaires au pré-test des instruments de collecte (questionnaires), des manipulations de facteurs, des instructions et des scénarios dans leur ensemble.

Les trois études expérimentales qui ont été menées se sont donc attachées à montrer les liens entre les différentes étapes du processus de réponse d'une part, et à identifier l'influence des facteurs individuels et situationnels sur les variables correspondant à ces étapes, d'autre part.

Plusieurs méthodes statistiques ont ainsi été utilisées pour tester nos différentes hypothèses de recherche, tout en tenant compte des facteurs exogènes introduits en tant que co-variables : une MANCOVA, une ANCOVA, et plusieurs régressions linéaires et régressions logistiques.

A l'issue des analyses, nous sommes en mesure de valider l'existence d'un processus de réponse à une sollicitation de données et l'enchaînement existant entre les différentes phases.
Notre recherche permet ainsi de confirmer le rôle majeur de l'attitude qui est un médiateur du lien entre l'évaluation de la sollicitation et les intentions comportementales de réponse, conformément à ce que la théorie de l'action raisonnée prédit.

Concernant l'impact des facteurs situationnels sur les différents éléments du processus de réponse, nous validons les hypothèses concernant l'effet des mentions relatives à la politique de respect de la vie privée adoptée par l'entreprise et de la quantité de données demandées.
De plus, bien que les résultats n'aillent pas toujours dans le sens attendu, ils montrent l'existence possible d'un lien direct entre deux autres facteurs (i.e. la familiarité avec l'entreprise et le contexte de sollicitation) et le comportement de réponse. En revanche, nous n'avons pas observé d'influence de ces deux éléments (i.e. familiarité et contexte) sur les variables liées à l'évaluation de la sollicitation (sensibilité, pertinence et valeur perçues). Nous prouvons enfin l'impact de ces facteurs (notamment du contexte de sollicitation) et du comportement de réponse sur l'intention du consommateur de poursuivre la relation avec l'entreprise qui l'a sollicité. Ce résultat permet alors de répondre à une des interrogations qui ont motivé ce travail de thèse : la sollicitation peut-elle conduire à une rupture de la relation entre le consommateur et l'entreprise à l'origine de la requête ? La réponse est oui.

Dans une même démarche de compréhension du phénomène, nous avons également testé les liens pouvant exister entre les variables qui caractérisent le répondant (variables individuelles) et son comportement. Nous avons ainsi pu vérifier empiriquement l'impact de la préoccupation pour le respect de la vie privée à différents niveaux du processus de réponse. Nous avions aussi introduit dans nos analyses des variables de contrôle dont l'effet a pu être testé. Il apparaît alors que le comportement habituel exerce une influence significative, alors que celle liée aux variables sociodémographiques et expérientielles est plus aléatoire.

L'examen des données empiriques a donc permis d'enrichir la vision théorique de la réponse à une sollicitation de données personnelles et de mieux comprendre les liens qui existent entre les facteurs qui caractérisent la sollicitation et les différentes étapes du processus de réponse.

Les contributions de la recherche

Les résultats obtenus ont permis d'identifier plusieurs contributions liées aux explications théoriques avancées et aux effets observés sur le processus de décision du consommateur.
Les apports de la recherche sont également d'ordre empirique, managérial et méthodologique.

Suite à nos analyses, nous pouvons ainsi avancer les « propositions » théoriques suivantes :

- Face à une sollicitation de ses données personnelles, le consommateur adopte un processus de décision majoritairement cognitif, où se succèdent différentes phases depuis l'exposition à la sollicitation jusqu'au comportement de réponse

- Les facteurs situationnels auraient un impact majeur sur ce processus, notamment ceux liés à la politique de RVP de l'entreprise et à la quantité de données demandées

- Même si certains facteurs individuels, notamment la préoccupation pour le RVP et le comportement de réponse habituel, exercent une influence notable à différents niveaux du processus, celle des facteurs situationnels est largement prépondérante

Cette recherche contribue donc à enrichir et à compléter les travaux déjà menés sur le sujet. De plus, c'est l'une des premières études réalisées dans le contexte français, ce qui constitue un apport empirique certain. Enfin, nos résultats soulignent l'importance pour les managers de connaître les critères utilisés par leurs clients pour évaluer la sollicitation des données. Nous sommes alors en mesure de proposer un ensemble de solutions à mettre en place pour rassurer les consommateurs, d'une part, et éviter tout risque de rupture de la relation, d'autre part.

Les limites et les voies de recherche

Malgré nos efforts pour en réduire la portée, ce travail comporte un ensemble de limites, principalement liées à la validité interne et à la validité externe de cette recherche.

Les limites à la validité interne concernent surtout la mise en situation des sujets expérimentaux et l'existence de biais inhérents au sujet étudié et au dispositif utilisé. Elles constituent toutefois autant de pistes d'améliorations possibles, en particulier à travers la prise en compte de nouvelles variables de contrôle et la mesure du biais de désirabilité sociale.
Les limites à la validité externe constituent également des voies de recherche pour le futur, notamment afin de tester la robustesse des effets empiriques qui ont pu être mis en évidence. Notre travail gagnerait ainsi à être étendu à d'autres facteurs situationnels testés (ou à une opérationnalisation différente de ceux étudiés ici), à d'autres catégories de produits/services, à d'autres médias (hors Internet), voire à d'autres populations (francophones ou non).
Enfin, des pistes de recherches complémentaires se dessinent aussi, en vue de compléter et de solidifier le modèle testé ici. La réalisation d'études axées sur la sensibilité des données et sur une vision longitudinale et éthique de l'échange apparaît également prometteuse.

La communauté académique et les professionnels ne sont pas les seuls bénéficiaires des résultats auxquels ce travail a permis d'aboutir. Nous pensons en effet que les recherches sur le comportement du consommateur sont au service des sciences sociales, voire sociétales. A cet égard, nos conclusions peuvent être utiles aussi bien au législateur qu'à l'ensemble des acteurs concernés par la collecte et l'utilisation de données sur les consommateurs. Nous pensons plus particulièrement aux organismes de contrôle (comme la CNIL), aux organisations professionnelles (comme la FEVAD), aux différents groupes de pression (activistes, avocats et autres défenseurs de la vie privée) ainsi qu'aux tiers certificateurs.

Cette recherche contribue à poser les bases d'une réflexion sur les déterminants de la réponse du consommateur face à une sollicitation de ses données personnelles et sur le processus de décision y afférant. Nous espérons que de nombreux travaux futurs viendront l'enrichir.

BIBLIOGRAPHIE

Aaker D. et Keller K. (1990), Consumer evaluations of brand extensions, *Journal of Marketing*, 54, 1, 27-41

Abratt R. et Sacks D. (1988), The marketing challenge: towards being profitable and socially responsible, *Journal of Business Ethics*, 7, 497-507

Acquisti A. (2004), Privacy in electronic commerce and the economics of immediate gratification, *Electronic Commerce Conference*, New York, USA, May

Acquisti A. et Grossklags J. (2003), Losses, gains and hyperbolic discounting: an experimental approach to information security attitudes and behavior, *2nd Annual Workshop on Economics and Information Security*, UC Berkeley, May

Acquisti A. et Grossklags J. (2004), Privacy and rationality: evidence from survey data, *Workshop on Economics and Information Security*

Adams J. (1963), Toward an understanding of inequity, *Journal of Abnormal and Social Psychology*, 67, 5, 422-436

Ajzen I. (1977), Intuitive theories of events and the effects of base-rate information on prediction, *Journal of Personality and Social Psychology*, 35, 303-314

Ajzen I. (1987), Attitudes, traits, and actions: dispositional prediction of behavior in social psychology, *Advances in Experimental Social Psychology*, 20, 1-63

Ajzen I. (1991), The Theory of Planned Behaviour, *Organizational Behavior & Human Decision Processes*, 50, 2, 179–211

Ajzen I. et Fishbein M. (1980), *Understanding Attitudes and Predicting Social Behavior*, Prentice-Hall Inc, Englewood Cliffs, New Jersey

Ajzen I. et Madden T. (1986), Prediction of goal-directed behavior: attitudes, intentions and perceived behavioral control, *Journal of Experimental Social Psychology*, 22, 453-474

Alba J. et Hutchinson J. (1987), Dimensions of consumer expertise, *Journal of Consumer Research*, 13, 411-454

Albaum G., Evangelista F. et Medina N. (1998), Role of response behaviour theory in survey Research: a cross-national study, *Journal of Business Research*, 42, 115-125

Alder G. (1998), Ethical issues in electronic performance monitoring: a consideration of deontological and teleological perspectives, *Journal of Business Ethics*, 17, 729-743

Alexander P. (1998), Attitudes toward information privacy: differences among and between faculty and students, *AIS Americas Conference*, Baltimore, Maryland, August 14-16, 3-6

Allen N. et Meyer J. (1990), The measurement and antecedents of affective continuance and normative commitment to the organization, *Journal of Occupational Psychology*, 63, 1-18

Alliger G. et Dwight S. (2001), Invade or evade? The trade-off between privacy invasion and item fakability, *Applied HRM Research*, 6, 2, 95-104

Altman I. (1975), *The Environment and Social Behavior:Privacy, Personal Space, Territory and Crowding*, Monterey, CA: Brooks/Cole Pub. Co., Inc.

Altman I. (1976), Privacy: a conceptual analysis, *Environment and Behavior*, 8, 1, 7-29

Altman I. (1977), Privacy regulation: culturally universal or culturally specific?, *Journal of Social Issues*, 33, 3, 66-84

Altman I. et Taylor D. (1973), *Social Penetration: The Development of Interpersonal Relationships*, New York: Holt, Rinehart and Wilson.

Amine A. (1990), Essai comparatif de deux échelles de mesure de l'implication, *Congrès de l'Association Française du Marketing*, La Baule, 512-539

Amine A. (1993), Etude empirique des liaisons entre l'implication du consommateur et son activité de recherche d'information, Cahier de recherche DMSP n°214, Université Paris Dauphine

Anderson E. et Weitz B. (1992), The use of pledges to built and sustain commitment in distribution channels, *Journal of Marketing Research*, 24, 18-34

Andrade E., Kaltcheva V. et Weitz B. (2002), Self disclosure on the web: the impact of privacy policy, reward and company reputation, *Advances in Consumer Research*, 29, 350-353

Archer R., Hormuth S. et Berg J. (1982), Avoidance of self disclosure : an experiment under conditions of self awareness, *Personality and Social Psychology bulletin*, 8, 1, 122-128

Archer R., Berg J., Runge T. (1980), Active and passive observers: attraction to self-disclosing others, *Journal of Experimental Social Psychology*, 16, 130-145

Aronson E., Wilson, T. et Brewer M. (1998), Experimentation in Social Psychology, dans *The Handbook of Social Psychology*, Gilbert D., Fiske S. et Lindzey G., Oxford University Press, 99-147

Armstrong J. et Overton T. (1977), Estimating non response bias in mail surveys, *Journal of Marketing Research*, 14, 396-402

Archer R. (1979), Role of Personality and the Social Situation, dans *Self-Disclosure: Origins, Patterns, and Implications of Openness in Interpersonal Relations*, Chelune and Associates (eds.), San Francisco: Jossey Bass

Aron A., Aron E., Tudor M. et Nelson G. (1991), Clause relationships as including other in the self, *Journal of Personality and Social Psychology*, 60, 2, 241-253

Awad N. et Krishnan M. (2006), The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization, *MIS Quarterly*, 30, 1, 13-28

Bates A. (1964), Privacy, a useful concept ?, *Social Forces*, 42, 4, 430-434

Bagozzi R. (1975), Marketing as exchange, *Journal of Marketing*, *39*, October, 32-39

Bagozzi R. (1977), Structural equation models in experimental research, *Journal of Marketing Research*, 14, 209-226

Bagozzi R. (1989), La renaissance de la recherche marketing sur les attitudes, *Recherche et Applications en Marketing*, 4, 2, 61-84

Bagozzi R. (1995), Reflections on relationship marketing in consumer markets, *Journal of the Academy of Marketing Science*, 23, 4, 272-277

Bandura A., Adams N., Hardy A. et Howells G. (1980), Tests of the generality of self-efficacy theory, *Cognitive Therapy and Research*, 4, 39-66

Bardin L. (1998), *L'Analyse de Contenu*, Collection Le Psychologue, Presses Universitaires de France, Paris

Baron R. et Kenny D. (1986), The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations, *Journal of Personality and Social Psychology*, 51, 6, 1173-1182

Basi R. (1999), www response rates to socio-demographic items, *Journal of the Market Research Society*, 41, 4, 397-401

Bauer H., Grether M. et Leach M. (2002), Building customer relations over the Internet, *Industrial Marketing Management*, 31, 155-163

Beauvois J.-L. et Joule R. (1981*), Soumission et Idéologies : Psychosociologie de la Rationalisation*, Presses Universitaires de France, Paris

Belk R. (1974), An exploratory assessment of situational effects in buyer behavior, *Journal of Marketing Research*, 11, May, 156-163

Belk R. (1975), Situational variables and consumer behavior, *Journal of Consumer Research*, 2, 157-164

Belk R. (1988), Possessions and the extended self, *Journal of Consumer Research*, 15, September, 139-168

Belleil A. (2001), *e-privacy : le marché des données personnelles*, Dunod, Paris

Benassi P. (1999), The platform of privacy preferences, *Communications of the ACM*, 42, 2, 56-59

Bendapudi N. et Berry L. (1997), Customers' motivations for maintaining relationships with service providers, *Journal of retailing*, 73, 1, 15-37

Benn S. (1971), *Privacy, Freedom, and Respect for persons*, dans Pennock and Chapman, 1-26

Bennett C. (1967), What price privacy ?, *American Psychologist*, 22, 371-376

Bennett R. et Gabriel H. (2001), Reputation, trust, supplier commitment : the case of shipping company / seaport relations, *Journal of Business and Industrial Marketing*, 16, 6, 424-438

Benoit Moreau F. (2006), La première rencontre mémorable entre marque et consommateur et son influence sur la relation : exploration par une approche qualitative phénoménologique, Cahier de Recherche DMSP n°354, Université Paris Dauphine

Bentler P. (1990), Comparative fit indexes in structural models, *Psychological Bulletin*, 238–246

Berdie D. (1973), Questionnaire length and response rate, *Journal of Applied Psychology*, 58, October, 278-280

Berelson B. (1952), *Content analysis in communication research*, New York: Hafner Press, Macmillan Publishing

Berendt B., Gunther O. et Spiekermann S. (2003), Privacy in e-commerce: stated preferences vs. actual behavior, *Communications of the ACM*, 48, 4, 101-106

Berg J. et Archer R. (1983), The disclosure-liking relationship: effects of self-perception, order of disclosure and topical similarity, *Human Communication Research*, 10, winter, 269-281

Berry J., Verity J., Kervin K. et DeGeorge G. (1994), Database marketing: a potent new tool for selling, *Business Week*, September 5, 56-62

Bercheid E. (1977), Privacy: A hidden variable in social psychology, *Journal of Social Issues, 33*, 85-101

Berkowitz L. et Donnerstein E. (1982), External validity is more than skin deep, *American Psychologist*, 37, 3, 245-257

Bhaskar R. (1975), *A realist theory of science*, Leeds, England: Leeds

Bibas S. (1994), A contractual approach to data privacy, *Harvard Journal of Law and Public Policy*, 17, 2, 591-612

Bickart B. et Schmittlein D. (1999), The distribution of survey contact and participation in the United States: Constructing a survey-based estimate, *Journal of Marketing Research, 36*, 2, 286-294

Bies R. (1993), Privacy and procedural justice in organizations, *Social Justice Research*, 6, 69-86

Blalock H. et Wilken P. (1979), *Intergroup processes: A micro-macro perspective*, New York: Free Press

Blanchette J.-F. et Johnson D. (2002), Data retention and the panoptic society: the social benefits of forgetfulness, *The Information Society*, 18, 33-45

Blattberg T., Buesing P., Peacok et Sen S. (1978), Identifying the deal prone segment, *Journal of marketing research*, 25, August, 369-377

Blattberg R. et Deighton J. (1991), Interactive marketing: exploiting the age of addressability, *Sloan Management Review*, Fall, 5-14

Blau P. (1964), *Exchange and Power in Social Life*, New York: Wiley

Bloch P. (1981), An exploration into the scaling of consumers' involvement with a product class, *Advances in Consumer Research*, 8, 61-65

Bloom P., Milne G. et Adler R. (1994), Avoiding misuse of new information technologies: legal and societal considerations, *Journal of Marketing*, 58, 98-110

Bloustein E. (1964), Privacy as an aspect of human dignity: a reply to Dean Prosser, *New York University Law Review*, 39, 962-1007

Boatright, J. (2000), *Privacy, Ethics and the Conduct of Business*, Upper Saddle River, NJ: Prentice Hall, 159-183

Bok S. (1984), *Secrets: On the Ethics of Concealment and Revelation*, Oxford University Press, Oxford

Bollen K. (1989), Structural equations with latent variables, New York: John Wiley & Sons

Bolton R. et Drew J. (1991), A longitudinal analysis of the impact of service changes on customer attitudes, *Journal of Marketing*, 55, 1, 1-10

Bosnjak M. et Tuten T. (2001), Classifying response behaviors in web-based surveys, *Journal of Computer Mediated Communication*, 6, 3

Boulaire C. et Balloffet P. (1999), Freins et motivations à l'utilisation d'Internet : une exploration par le biais des métaphores, *Recherche et Applications en Marketing*, 14, 1, 23-39

Bozzo C. (2000), Modélisation du comportement d'achat en milieu industriel : fidélité, rétention, inertie, Thèse de Doctorat en Sciences de Gestion, IAE d'Aix en Provence

Brackett L. et Carr B. (2001), Cyberspace advertising vs. other media: consumer vs. mature students' attitudes, *Journal of Advertising Research*, September-October, 23-32

Breckler S. (1984), Empirical validation of affect, behaviour and cognition as distinct components of attitude, *Journal of Personality and Social Psychology*, 47, 1191-1205

Bressoles G. (2003), L'instauration d'un climat de confiance en ligne : le rôle des signes de qualité des sites internet marchands, *Actes du 3ème Congrès sur les Tendances du Marketing en Europe*, Venise.

Brockner J. et Higgins T. (2001), Regulatory focus theory: implications for the study of emotions at work, *Organizational Behavior and Human Decision Processes*, 86, 1, 35-66

Brown B. (1987), Territoriality, dans *Handbook of Environmental Psychology*, Stokols et Altman eds., New York: John Wiley & Sons, 505-533

Buchholz R. (1989), *Business Ethics*, Englewood Cliffs, NJ

Buchholz R. et Rosenthal S. (2002), Internet privacy: individuals rights and the common good, *SAM Advanced Management Journal*, winter, 34-40

Burgoon J. (1982), Privacy and communication, *Communication Yearbook 6*, Michael Burgoon ed., Beverly Hills: Sage

Buttler R. (1973), Effects of signed and unsigned questionnaires for both sensitive and non sensitive items, *Journal of Applied Psychology*, 57, 3, 348-349

Calciu M. (2003), Expérimentation et aide à la décision en marketing sur internet, *Congrès de l'Association Française du Marketing*, Tunis

Calder B., Phillips L. et Tybout A. (1981), Designing research for application, *Journal of Consumer Research*, 8, 197-207

Calder B., Phillips L. et Tybout A. (1982), The concept of external validity, *Journal of Consumer Research*, 9, 240-244

Calder B., Phillips L. et Tybout A. (1983), Beyond external validity, *Journal of Consumer Research*, 10, 112-114

Camp L. (1999), Web security and privacy: an American perspective, *The Information Society*, 15, 249-256

Campbell A. (1997), Relationship marketing in consumer markets: a comparison of managerial and consumer attitudes about information privacy, *Journal of Direct Marketing*, 11, 3, 44-57

Campbell D. et Stanley J. (1963), *Experimental and quasi-experimental designs for research*, Chicago: RandMcNally

Cannell C., Oksenberg L. et Converse J. (1977), Striving for response accuracy: experiments in new interviewing techniques, *Journal of Marketing Research*, 14, 306-315

Carver C. et Scheier M. (1992), *Perspective on Personality*, Allyn and Bacon

Cattell R. (1995), *Handbook for the Sixteen Personality Factors Questionnaire*, Champaign, Illinois, IPAT

Caudill E. et Murphy P. (2000), Consumer online privacy: legal and ethical issues, *Journal of Public Policy and Marketing*, 19, 1, 7-19

Celsi R. et Olson J. (1988), The role of involvement in attention and comprehension processes, *Journal of Consumer Research*, 15, September, 210-224

Cespedes F. et Smith J. (1993), Database marketing: new rules for policy and practice, *Sloan Management Review*, 7-22

Chaiken S. (1980), Heuristic vs. systematic information processing and the use of source vs. message cues in persuasion, *Journal of personality and social psychology*, 39, 5, 762-766

Chaikin A. et Derlega V. (1974), Variables affecting the appropriateness of self-disclosure, *Journal of Consulting and Clinical Psychology*, 42, 4, 588-593

Champion D. et Sear A. (1969), Questionnaire response rate : a methodological analysis, *Social Forces*, 47, 335-339

Chang M. (1998), Predicting unethical behavior : a comparison of Theory of Reasoned Action and Theory of Planned Behavior, *Journal of Business Ethics*, 17, 1825-1834

Chapman F. (1985), The ruckus over medical testing, *Fortune*, August 19, 57-58

Charters D. (2002), Electronic monitoring and privacy issues in business-marketing: the ethics of the DoubleClic experience, *Journal of Business Ethics*, 35, 243-254

Chebat J.-C. et Cohen A. (1993), Response speed in mail surveys : beware of shortcuts, *Journal of Marketing Research*, 5, 2, 20-26

Chellappa R. (2001), Contrasting expert assessment of privacy with perceived privacy: implications for public policy, www.ebizlab.usc.edu

Chellappa R. et Sin R. (2005), Personalization versus privacy: an empirical examination of the online consumer's dilemma, *Information Technology and Management*, 6, 2-3, 181-202

Chelune G. (1975), Self-disclosure: an elaboration of its basic dimensions, *Psychological Reports*, 36, 79-85

Cheskin Research (2000), Trust in the wired Americas, http://www.cheskin.com/think/studies/trustIIrpt.pdf.

Cheung C. et Lee M. (2001), Trust in internet shopping: instrument development and validation through classical and modern approaches, *Journal of Global Information Management*, 9, 3, 23-32

Childers T. et Ferrell O. (1979), Response rates and perceived questionnaire length in mail surveys, *Journal of Marketing Research*, 16, 429-431

Childers T. et Ruekert R. (1982), The meaning and determinants of cooperation within an inter organizational marketing network, dans *Marketing Theory: Philosophy of Science Perspectives*, Ron Bush and Shelby Hunt eds.

Childers T. et Skinner S. (1979), Gaining respondent cooperation in mail surveys through prior commitment, *Public Opinion Quarterly*, 43, 4, 558-561

Childers T. et Skinner S. (1985), Theoretical and empirical issues in the identification of survey respondents, *Journal of the Market Research Society*, 27, 1, 39-53

Childers T. et Skinner S. (1996), Toward a conceptualization of mail survey response behavior, *Psychology and Marketing*, 13, 2, 185-209

Chircu A., Davis G. et Kauffman R. (2000), The role of trust and expertise in the adoption of electronic commerce intermediaries, papier de recherche, MISRC

Chumpitaz Cacerès R. et Vanhamme J. (2003), Les processus modérateurs et médiateurs : distinction conceptuelle, aspects analytiques et illustrations, *Recherche et applications en Marketing*, 18, 2, 67-100

Churchill G. (1979), A paradigm for developing better measures of marketing constructs, *Journal of Marketing Research*, 16, 64-73

Cockroft S. et Clutterbuck (2001), Attitudes towards information privacy, *12th Australasian Conference on Information System*, December

Collins N. et Miller L. (1994), Self-disclosure and liking: a meta-analytic review, *Psychological Bulletin*, 116, 3, 457-475

Colombo R. (2000), A model for diagnosing and reducing non response bias, *Journal of Advertising Research*, 85-93

Comley P. (2000), Pop-up Surveys : what works, what doesn't work and what will work in the future, *ESOMAR Net Effects Internet Conference*, Dublin, April, http://www.virtualsurveys.com/papers/popup_paper.htm

Connolly T. (1976), Some conceptual and methodological issues in expectancy models of work performance motivation, *Academy of Management Review*, 179-186

Constant D., Kiesler S. et Sproull L. (1994), What's mine is ours, or is it ? a study of attitudes about information sharing, *Information Systems Research*, 5, 4, 400-421

Cook T. et Campbell D. (1979), *Quasi-Experimentation, Design & Analysis Issues for Field Settings*, Houghton Mifflin: Boston

Cook B. et Coupey E. (1998), Consumer behavior and unresolved regulatory issues in electronic marketing, *Journal of Business Research*, 41, 231-238

Cook K. et Emerson R. (1978), Power, dependence and commitment in exchange networks, *American Sociological Review*, 43, October, 721-739

Cook K. et Whitemeyer J. (1992), Two approaches to social structure: exchange theory and network analysis, *Annual Review of Sociology*, 18, 109-127

Corcoran K. (1988), The relationship of interpersonal trust to self-disclosure when confidentiality is assured, *Journal of Psychology*, 122, 193-195

Cossette P. (1994), Cartes cognitives et organisations, Paris: Eska

Costa P. et McCrae R. (1992), *Revised NEO Personality Inventory (INEO PI-R) and NEO Five-factor Inventory (NEO_FFI): professional manual*, Odessa, Fl.

Couper M. (2000), Web surveys: a review of issues and approaches, *Public Opinion Quarterly*, 64, 464-494

Cox D. et Rich S. (1964), Perceived risk and consumer decision-making, the case of telephone shopping, *Journal of Marketing Research*, 1, November, 32-39

Cozby P. (1973), Self disclosure: a literature review, *Psychological Bulletin*, 79, 2, 73-91

Cranor L., Reagle J. et Ackerman M. (1999), Beyond concern: understanding net users' attitudes about online privacy, AT&T Labs, Research Technical Report, 99.4.3

Cristau C. et Strazzieri A. (1996), Implication et leadership d'opinion : la valeur prédictive de trois échelles d'implication durable, *Congrès de l'Association Française de Marketing*, Poitiers, 141-160

Cronin J. et Taylor S. (1992), Measuring service quality: a reexamination and extension, *Journal of Marketing*, 56, 55-68

Cronbach L. (1975), Beyond the two disciplines of scientific psychology, *American Psychologist*, 30, 116-127

Cronbach L. (1988), Five perspectives on validity argument, dans *Test Validity*, Wainer et Braun eds., Hillsdale Erlbaum, 3-17

Culnan M. (1993), How did they get my name ? an exploratory investigation of consumer attitudes toward secondary information use, *MIS Quarterly*, 17, 3, 341-363

Culnan M. (1995), Consumer awareness of name removal procedures: implications for direct marketing, *Journal of Direct Marketing*, 9, 2, 10-19

Culnan M. (1999), Georgetown Internet Privacy Policy Survey: Report to the Federal Trade Commission, http://www.msb.edu/faculty/culnanm/gippshome.html]

Culnan M. (2000), Protecting privacy online: is self regulation working ?, *Journal of Public Policy et Marketing*,19, 1, 20-26

Culnan M. et Armstrong P. (1999), Information privacy concerns, procedural fairness and impersonal trust: an empirical investigation, *Organization Science*, 10, 1, 104-115

Culnan M. et Bies R. (2003), Consumer privacy: balancing economic and justice considerations, *Journal of Social Issues*, 59, 2, 323-342

Curran P., West S. et Finch J. (1996), The robustness of test statistics to non normality and specification error in confirmatory factor analysis, *Psychological Methods*, 1, 1, 16-29

Curry D. (1993), The new marketing research systems: how to use strategic database information for better marketing decisions, *Journal of Marketing Sciences*, New York: John Wiley

Curtis R. et Miller K. (1986), Believing another likes or dislikes you: behaviors making the beliefs come true, *Journal of Personality and Social Psychology*, 51, August, 284-290

Dalton D. et Metzger M. (1992), Toward candor, cooperation and privacy in applied business ethics research: the randomized response technique (RRT), *Business Ethics Quarterly*, 2, 207-221

Daniel W. (1975), Nonresponse in sociological surveys: a review of some methods for handling the problem, *Sociological Methods and Research*, 3, 291-307

Darpy D. (1999), La procrastination du consommateur : une contribution à l'explication du report d'achat, Thèse de Doctorat en Sciences de Gestion, Université Paris Dauphine

Darpy D. et Volle P. (2003), *Comportements du consommateur - Concepts et Outils*, Dunod, Paris

Davis F. (1989), Perceived usefulness, perceived ease of use and user acceptance of information technology, *MIS Quarterly*, 13, 3, 319-340

Davis F. (1993), User acceptance of information technology: system characteristics, user perceptions and behavioral impacts, International Journal of Man-Machine Studies, 38, 475-487

Davis J. (1994), Good ethics is good for business: ethical attributions and response to environmental advertising, *Journal of Business Ethics*, 13, 873-885

Davis J. (1997), Property rights to consumer information; a proposed policy framework for direct marketing, *Journal of Direct Marketing*, 11, Summer, 32-43

Davis F., Bagozzi et Warshaw (1989), User acceptance of computer technology: a comparison of two theoretical models, *Management Science*, 35, 8, 982-1003

Day G. (1969), Buyer attitudes and brand choice behaviour, *Journal of Advertising Behaviour*, 9, 5, 29-35

Debenedetti S. (2001), Rôle et impact de l'accompagnement du visiteur du lieu culturel : le cas de la sortie au musée d'art, Thèse de Doctorat en Sciences de Gestion, Université Paris Dauphine

De Leeuw E., Hox J. et Snijkers G. (1995), The effect of computer-assisted interviewing on data quality: a review, *Journal of The Market Research Society*, 37, 4, 325-344

DeMaio T. (1980), Refusals: Who, where and why, *Public Opinion Quarterly*, 221-233

Dennis P. (1998), Marketing one to one and its dependence on knowledge discovery in databases, *The Journal of Consumer Marketing*, 15, 5, 468-480

Derbaix C. (1983), Perceived risk and risk relievers: an empirical investigation, *Journal of Economic Psychology*, 3, 19-38

Derlega V. (1988), Self-disclosure: inside or outside the mainstream of social psychological research?, *Journal of Social Behavior and Personality*, 3, 27-34

Derlega V. et Chaikin A. (1976), Norms affecting self disclosure in men and women, *Journal of Consulting and Clinical Psychology*, 44, 3, 376-380

Derlega V. et Chaikin A. (1977), Privacy and self-disclosure in social relationships, *Journal of Social Issues*, 33, 3, 102-115

Derlega V. et Grzelak J. (1979), Appropriateness of self-disclosure, dans *Self-disclosure: Origins, patterns, and implications of openness in interpersonal relationships*, Chelune (Ed.), San Francisco: Jossey-Bass

Derlega V., Harris M. et Chaikin A. (1973), Self-disclosure reciprocity, liking and the deviant, *Journal of Experimental Social Psychology*, 9, 277-284

Derlega V., Metts S., Petronio S. et Margulis S. (1993), *Self-disclosure*, Newbury Park: Sage

Derlega V., Wilson M. et Chaikin A. (1976), Friendship and disclosure reciprocity, *Journal of Personal and Social Psychology*, 34, 4

Desai C., Fletcher K. et Wright G. (2001), Drivers in the adoption and sophistication of database marketing in the services sector, *The Service Industry Journal*, 21, 4, 17-32

Desai C., Wright G. et Fletcher K. (1998), Barriers to successful implementation of database marketing : a cross industry study, *International Journal of Information Management*, 18, 4, 265-276

Deutskens L., Ruyter K. et Wetzels M. (2005), Are online access panels the future of marketing research? A motivational perspective on respondent participation in online panels, *34th European Marketing Academy Conference*, Milan, Italy: Bocconi University.

Dibb S., Pride W. et Ferrell O. (1994), *Marketing Concepts and Strategies*, Houghton Mifflin Company

Didellon L. et Valette-Florence P. (1996), L'utilisation des indices d'ajustement dans les modèles d'équations structurelles : Présentation et recommandation d'usage, XIIIème Journées Nationales des IAE, Toulouse, 1, 111-126.

Dimond R. et Hellkamp D. (1969), Race, sex, ordinal position or birth and self disclosure in high school students, *Psychological Reports*, 25, 235-238

Dillman D., Sinclair M. et Clark J. (1993), Effects of questionnaire length, respondent friendly design, and a difficult question on response rates for occupant addressed census mail surveys, *Public Opinion Quarterly*, 57, 289-304

Dillman D. (1978), *Mail and telephone surveys; the total design method*, New York: John Wiley & Sons

Dillman D. (1991), The design and administration of mail surveys, *Annual Review of Sociology*, 17, 225-249

Dillman D. (1998), Influence of plain vs. fancy design on response rates for web surveys, *Proceedings of The Survey Research Methods Section*, American Statistical Association

Dillman D. (1999), Influence of type of question on skip pattern compliance in self administered questionnaires, Proceedings of the Survey Research Methods Section, American Statistical Association

Dillman D. et Bowker D. (2001), The Web questionnaire challenge to survey methodologists, dans U.-D. Reips & M. Bosnjak Eds., *Dimensions of Internet science*, Lengerich: Pabst Science Publishers, 159-178

Dinev T. et Hart P. (2002), Internet privacy concerns and trade-off factors: empirical study and business implications, *International Conference On Advances In Infrastructure for E-Business*, L'Aquila, Italy

Dinev T. et Hart P. (2003), Internet privacy concerns and trade off factors: empirical study and business implications, *Academy of Management Proceedings*, Seatle, 63-73

Dommeyer C. et Gross B. (2003), What consumers know and what they do: an investigation of consumer knowledge, awareness, and use of privacy protection strategies, *Journal of Interactive Marketing*, 17, 2, 34-51

Dommeyer C. et Moriarty E. (2000), Comparing two forms of an e-mail survey: embedded vs. attached, *International Journal of Market Research*, 42, 1, 39-50

Donaldson T. et Dunfee T. (1994), Toward a unified conception of business ethics: integrative social contracts theory, *Academy of Management Review*, 19, 2, 252-284

Doney P. et Cannon J. (1997), An examination of the nature of trust in buyer – seller relationships, *Journal of Marketing*, 61, April, 35-51

Doster J. et Strickland B. (1967), Perceived childrearing practices and self-disclosure patterns, *Journal of Consulting and Clinical Psychology*, 33, 382

Dunfee T., Smith C. et Ross W. (1999), Social contracts and marketing ethics, *Journal of Marketing*, 63, 14-32

Dwyer F., Schurr P. et Oh S. (1987), Developing buyer – seller relationships, *Journal of Marketing*, 51, 11-27

Eddy E., Stone D. et Stone E. (1999), The effects of information management policies on reactions to human resource information systems : an integration of privacy and procedural justice perspectives, *Personnel Psychology*, 52, 2, 335-359

Edney J. et Buda M. (1976), Distinguishing territoriality and privacy: two studies, *Human Ecology*, 4, 4, 283-296

Edwards R. (1993), An education in interviewing: placing the researcher and the research, dans *Researching Sensitive Topics*, C. Renzetti and M. Lee eds., Newsbury Park, CA: Sage Publications, 45-57

Ehrenberg A. (1974), Repetitive advertising and the consumer, *Journal of Advertising Research*, 14, 2, April, 25-34

Eisenhardt K. (1989), Building theories from case study research, *Academy of Management Review*, 14, 4, 532-550

Emerson R. (1962), Power-dependence relations, *American Sociological Review*, 27, 1, 31-41

Emerson R. (1972). Exchange Theory, dans *Sociological Theories in Progress*, J. Berger et al. eds., Boston: Houghton-Mifflin.

Essig L. et Winter J. (2003), Item nonresponse to financial questions in household surveys: an experimental study of interviewer and mode effects, http://www.sfb504.uni-mannheim.de/publications/dp05-18.pdf

Evans M., O'Malley L. et Patterson M. (1998), Relationship marketing and privacy issues: building bonds or barriers, *Journal of Database Marketing***,** 6, 1, 34-47

Evans M., Patterson M., O'Malley L. (2001), The direct marketing – direct consumer gap : qualitative insights, *Qualitative Market Research*, 4, 1, 17-24

Evrard Y., Pras B. et Roux E. (2003), *Market : études et recherches en marketing*, Dunod, Paris

Falk D. et Wagner P. (1985?), Intimacy of self-disclosure and response process as factors affecting the development of interpersonal relationships, *The Journal of Psychology*, 125, 5, 557-570

Farag N. et Krishnan M. (2003), An empirical evaluation of information features and the willingness to be profiled online for personalization, http://misrc.umn.edu/workshops/2003/spring/farag_030703.pdf

Farissi Sainz M-K. (1997), La théorie d'engagement et de confiance du marketing relationnel, *Congrès de l'Association Française du Marketing*, Toulouse, 376-402

Ferber R. (1977), Research by convenience, *Journal of Consumer Research*, 4, 57-58

Fernandes M. et Randall D. (1992), The nature of social desirability response effects in ethics research, *Business Ethics Quarterly*, 2,2, 183-205

Filion F. (1975-1976), Estimating bias due to non response in mail surveys, *The Public Opinion Quarterly*, 39, 4, 482-492

Fishbein M. (1967), *Readings in Attitude Theory and Measurement*, New York: John Wiley and Sons, Inc

Fishbein M. et Ajzen I. (1975), *Belief, attitude, intention and behaviour: An introduction to theory and research*, Reading, Mass.: Addison-Wesley

Fitzgerald M. (1963), Self-disclosure and expressed self-esteem, social distance and areas of the self-revealed, *The Journal of Psychology*, 56, 405-412

Flaherty D. (1984), The need for a privacy protection study commission, *Government Information Quarterly*, 1, 235-258

Fleck-Dousteyssier N. (2006), Effets du parrainage sur le capital de marque : le rôle de la congruence, Thèse de Doctorat en Sciences de Gestion, Université Paris Dauphine

Fletcher K., Wheeler C. et Wright J. (1991), Database marketing: a channel, a medium or a strategic approach, *International Journal of Advertising*, 10, 117-127

Fletcher K. et Wright G. (1995), Organizational, strategic and technical barriers to successful implementation of database marketing, *International Journal of Information Management*, 15, 2, 115-126

Fletcher K. et Peters L. (1996), Issues in customer information management, *Journal of the Market Research Society*, 38, 2, 145-160

Fletcher K., Wheeler C. et Wright J. (1994), Strategic implementation of database marketing: problems and pitfalls, *Long Range Planning*, 27, 1, 133-141

Fletcher K. (2000), Consumer power and privacy: the changing nature of CRM, *International Journal of Advertising*, 22, 249-272

Fornell C. et Larker D. (1981), Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research*, 18, 39-50

Fournier C. (1999), Sensibilité du consommateur à la protection des données à caractère personnel et aux problèmes d'environnement : deux antécédents de l'attitude générale face au mailing, *Congrès de l'Association Française du Marketing*, Strasbourg, 137-157

Fox R., Crask M. et Kim J. (1988), Mail survey response rate: a meta analysis of selected techniques for inducing response, *Public Opinion Quarterly*, 52, 467-491

Fox C., Levitin A. et Redman T. (1994), The notion of data and its quality dimensions, *Information Processing and Management*, 30, 1, 9-19

Foxman E. et Kilcoyne P. (1993), Information technology, marketing practice and consumer privacy: ethical issues, *Journal of Public Policy and Marketing*, 12, 1, 106-126

Frankel K. (1990), Women and computing, *Communications of the ACM*, 33, 11, 34-45

Frick A., Bächtinger M. et Reips U-D. (1999), Financial incentives, personal information and drop-out rate in online studies, dans *Current Internet science: Trends, techniques, results*, U-D. Reips et al. Eds., http://www.dgof.de/tband99/pdfs/a_h/frick.pdf

Fried C. (1968), Privacy : a moral analysis, *Yale Law Journal*, 77, 1, 475-493

Friedman H. et Goldstein L. (1975), Effect of ethnicity of signature on the rate of return and content of a mail questionnaire, *Journal of Applied Psychology*, 60, 770-771

Friedman W. (2004), Perspectives on privacy, *The Journal of American Academy of Business*, 244-250

Friedman B., Kahn P. et Howe D. (2000), Trust online, *Communications of the ACM*, 43, 12, 34-41

Frisou J. (2000), Confiance interpersonnelle et engagement : une réorientation behavioriste, *Recherche et Applications en Marketing*, 15, 1, 63-79

Fry J. (1996), The move beyond the banner demands better measurements, *Wall Street Journal Interactive*, http://www.wsj.com

Furse D. et Stewart D. (1982), Monetary incentives versus promised contribution to charity: New evidence on mail survey response, *Journal of Marketing Research*, 19, 375-380

Fusilier M. et Hoyer W. (1980), Variables affecting perceptions of invasion of privacy in a personnel selection situation, *Journal of Applied Psychology*, 65, 5, 623-626

Galan J. et Vernette E. (2000), Vers une quatrième génération : les études de marché on line, *Décisions Marketing*, 19, 39-52

Gallager K., Parsons J. et Foster K. (2001), A tale of two studies: replicating advertising effectiveness and content evaluation in print and on the web, *Journal of Advertising Research*, July-August, 71-81

Ganassali S. et Moscarola J. (2002), Protocoles d'enquête et efficacité des sondages par Internet, *1ère Journée Nantaise de Recherche sur le e-marketing*, 57-73

Gandy O. (1993), *The Panoptic Sort: A Political Economy of Personal Information*, Boulder: Westview Press.

Gandy O. (2003), Public opinion surveys and the formation of privacy policy, *Journal of Social Issues*, 59, 2, 283-299

Ganesan S. (1994), Determinants of long-term orientation in buyer-seller relationships, *Journal of Marketing*, 58, 1-19

Ganesan S. et Hess R. (1997), Dimensions and levels of trust: implications for commitment to a relationship, *Marketing Letters*, 8, 4, 439-448

Gangadharbatla H. (2003), Online data collection, http://www.ciadvertising.org/sa/spring_03/391k/harsha/paper/about.html

Garbarino E. et Johnson M. (1999), The different roles of satisfaction, trust and commitment in customer relationships, *Journal of Marketing*, 63, 70-87

Garbarino E. et Strahilevitz M. (2004), Gender difference in the perceived risk of buying online and the effects of receiving a site recommendation, *Journal of Business Research*, 57, 768-775

Gauzente C. et Ranchhod A. (2002), Ethical marketing for competitive advantage on the Internet, *Academy of Marketing Science Review*, 10, http://www.amsreview.org/amsrev/forum/gauzente10-01.html

Gauzente C., Benetteau J. et Dubreuil E. (2002), Respect de la vie privée sur Internet : une étude des pratiques des sites français, *1ère Journée Nantaise de Recherche sur le e-marketing*, 117-133

Gauzente C. (2003), Vie privée et internet : la pratique des sites français, *Décisions Marketing*, 30, 7-17

Gauzente C. (2004), Web merchants' privacy and security statements: how reassuring are they for consumers? A two-sided approach, *Journal of Electronic Commerce Research*, 5, 3, 181-197

Georgetown Internet Privacy Policy Survey (1999), Report to the Federal Trade Commission, www.msb.edu/faculty/culnanm/gippshome.html

Georgia Institute of Technology (1998), GVU's 10th WWW User Survey, http://www.gvu.gatech.edu/user_surveys/survey-1998-10/

Gervey B. et Lin J. (2000), Obstacles on the Internet, *Advertising Age [Special Internet edition]*, 71, 16, 12-22

Geyskens I., Steenkamp J., Sheer L. et Kumar N. (1996), The effects of trust and interdependence on relationship commitment: a trans-atlantic study, *International Journal of Research in Marketing*, 13, 303-317

Geyskens I., Steebkamp J., Kumar N. (1998), Generalizations about trust in marketing channels relationships using meta-analysis, *International Journal of Research in Marketing*, 15, 223-248

Gilroy D. et Desai H. (1986), Computer anxiety: sex, race and age, *International Journal of Man-Machine Studies*, 25, 711-719

Glazer R. (1991), Marketing in an information-intensive environment: strategic implications of knowledge as an asset, *Journal of Marketing*, 55, October, 1-19

Gleitman H. (1995), *Psychology*, Norton & Company, Inc.

Goffman E. (1959), *The presentation of self in everyday life*, Garden City:Doubleday

Goodwin C. (1991), Privacy: recognition of a consumer right, *Journal of Public Policy and Marketing*, 10, 1, 149-166

Goodwin K. (1992), A conceptualisation of motives to seek privacy for nondeviant consumption, *Journal of Consumer Psychology*, 1, 3, 261-284

Gordon M., McKeage K. et Mark A. (1998), Relationship marketing effectiveness: the role of involvement, *Psychology and Marketing*, 15, 5, 443-459

Goritz A. (2004), The impact of material incentives on response quantity, response quality, sample composition, survey outcome and cost in online access panels, *International Journal of Market Research*, 46, 3, 327-345

Goudy W. (1976), Non response effects on relationships between variables, *The Public Opinion Quarterly*, 40, 3, 360-369

Gouldner A. (1960), The norm of reciprocity: a preliminary statement, *American Sociological Review*, 25, 161-178

Goyder J. (1986), Surveys on surveys: limitations and potentialities, *Public Opinion Quarterly*, 50, 27-41

Goyder J. et McKenzie L. (1985), The decline in survey response: a social values interpretation, *Sociology*, 19, 55-71

Graeff T. et Harmon S. (2002), Collecting and using personal data: consumers' awareness and concerns, *Journal of Consumer Marketing*, 19, 4, 302-319

Green K. (1996), Sociodemographic factors and mail survey response, *Psychology and Marketing*, 13, 2, 171-184

Greenberg J. (1986), Determinants of perceived fairness of performance evaluations, *Journal of Applied Psychology*, 71, 2, 340-342

Greenberg J. (1990), Organizational justice : yesterday, today and tomorrow, *Journal of Management*, 16, 2, 399-432

Greenberg J. (1987), A taxonomy of organizational justice theories, *The Academy of Management Review*, 12, 1, 9-22

Greenberg J. et McCarty C. (1990), The interpersonal aspect of procedural justice: a new perspective on pay fairness, *Labour Law Journal*, 580-586

Groves R., Cialdini R. et Couper M. (1992), Understanding the decision to participate on the survey, *Public Opinion Quarterly*, 56, 475-495

Groves R. et Couper M. (1998), *Nonresponse in Household Interview Surveys*, New York: Wiley

Groves R., Singer E. et Corning A. (2000), Leverage-saliency theory of survey participation, *Public Opinion Quarterly*, 64, 299-308

Gruen T. (1995), The outcome set of relationship marketing in consumer markets, *International Business Review*, 4, 4, 447-469

Gundlach G., Achrol R. et Mentzer J. (1995), The structure of commitment in exchange, *Journal of Marketing*, 59, January, 78-92

Gundlach G. et Murphy P. (1993), Ethical and legal foundations of relational marketing exchanges, *Journal of Marketing*, 57, October, 35-46

Gur R. et Sackeim H. (1979), Self-deception: a concept in search of a phenomenon, *Journal of Personality and Social Psychology*, 37, 2, 147-169

Gurau C., Ranchhod A. et Gauzente C. (2003), To legislate or not to legislate : a comparative exploratory study of privacy-personalisation factors affecting French, UK and US web sites, *Journal of Consumer Marketing*, 20, 7, 652-654

Gurviez P. (1998), La confiance du consommateur dans la marque, conceptualisation, mesure et management, *Congrès de l'Association Française du Marketing*, 73-96

Gutierrez S., Cillan J. et Izquierdo C. (2004), The consumer's relational commitment: main dimensions and antecedents, *Journal of Retailing and Consumer Services*, 11, 351-367

Hagel J. et Rayport J. (1997), The coming battle for customer information, *Harvard Business Review*, 75, 1, 53-55

Halaby C. et Weakliem D. (1989), Worker control and attachment to the firm, *American Journal of Sociology*, 95, 549-591

Han P. et MacLaurin A. (2002), Do consumers really care about online privacy?, *Marketing Management*, January-February, 353-38

Hann I. et al. (2002), On line information privacy: measuring the cost-benefit trade-off, *23td International Conference on Information Systems*, http://www.comp.nus.edu.sg/~ipng/research/privacy_icis.pdf

Hann et al. (2005), Analysing online information privacy concerns: an information processing theory approach, papier de recherche, http://www.comp.nus.edu.sg/~lung/privacy_conjoint.pdf

Hann I., Hui K., Lee T. et Png I. (2002), The value of online information privacy: evidence from the USA and Singapore

Hann H., Hui K., Lee T. et Png I. (2003), Overcoming online information privacy concerns: a comparison of privacy policies, convenience and promotions, http://www.comp.nus.edu.sg/~lung/privacy.html

Hanna R. et al. (2005), Do Internet-based surveys increase personal self disclosure, *Database Marketing and Customer Strategy Management*, 12, 4, 342-356

Hansen R. (1980), A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior, *Journal of Marketing Research*, 17, February, 77-83

Harris L. (1995), Equifax-Harris mid-decade consumer privacy survey, http://www.mindspring.com/~mdeeb/equifax/cc/parchive/svry95/survy95a.html

Hart P. et Dinev T. (2002), Measuring antecedents and trade-offs to internet privacy, *Academy of Management Meeting*, Colorado

Harvey L. (1987), Factors affecting response rates to mailed questionnaires: a comprehensive literature review, *Journal of the Market Research Society*, 29, 341-353

Heberlein T. et Baumgartner R. (1978), Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature, *American Sociological Review*, 43, 4, 447-462

Heberlein T. et Baumgartner R. (1981), The effectiveness of the Heberlein-Baumgartner models for predicting response rates to mailed questionnaires: European and US examples, *American Sociological Review*, 46, 3, 363-367

Heide J. et John G. (1992), Do norms matter in marketing relationships? *Journal of Marketing*, 56, April, 32-44

Helme-Guizon A. et Amato S. (2004), Favoriser l'achat et la fidélité des internautes : les apports de la théorie psychosociale de l'engagement, *Décisions Marketing*, 34, 53-66

Henderson S. et Snyder C. (1999), Personal information privacy: implications for MIS managers, *Information & Management*, 36, 213-220

Hennig-Thurau T., Gwinner K. et Gremler D. (2002), Understanding relationship marketing outcomes: an integration of relational benefits and relationship quality, *Journal of Service Research*, 4, 3, 230-247

Hennig-Thurau T. et Klee A. (1997), The impact of customer satisfaction and relationship quality on customer retention: a critical reassessment and model development, *Psychology and Marketing*, 14, 8, 737-764

Herbert M. (2003), Erreur de mesure liée au comportement du sujet expérimental : les artefacts de la demande en comportement du consommateur : état de l'art et perspectives, *Congrès de l'Association Française du Marketing*, Tunis

Herbert M. (2005 a), Le comportement de réponse de l'individu en situation de questionnement : le biais du répondant en comportement du consommateur, Thèse de Doctorat en Gestion, Université Paris Dauphine

Herbert M. (2005 b), Petit abrégé pour mieux comprendre la notion de méthode expérimentale et ses enjeux méthodologiques, Cahier de recherche DMSP n°340, Université Paris Dauphine

Hershel R. et Andrews P. (1997), Ethical implications of technological advances on business communications, *Journal of Business Communication*, 34, 167-170

Higgins E. (1987), Self-discrepancy: a theory relating self and affect, *Psychological Review*, 94, 3, July, 319-340

Higgins E. (1997), Beyond pleasure and pain, *American Psychologist*, 52, 12, December, 1280-1300

Higgins E. (1998), Promotion and Prevention: Regulatory Focus as a Motivational Principle, dans *Advances in Experimental Psychology*, Vol. 30, Zanna ed., San Diego, CA:Academic Press, 1-46

Hill R. (1995), Researching sensitive topics in marketing: the special case of vulnerable populations, *Journal of Public Policy and Marketing*, 14, 1, 143-148

Hinde S. (2002), The perils of privacy, *Computers and Security*, 21, 5, 424-432

Hine C. et Eve J. (1998), Privacy in the marketplace, *The Information Society*, 14, 253-262

Hirschman A. (1970), *Exit, Voice and Loyalty: responses to declines in firms, organizations and states*, Cambridge, MA: Harvard University Press

Hochstim J. (1967), A critical comparison of three strategies of collecting data from households, *Journal of the American Statistical Association*, 62, 319, 976-989

Hoekstra J. et Huizingh E. (2000), The interaction data framework: measuring the suitability of the customer database for relational strategies, *Journal of Database Marketing*, 7, 3, 232-245

Hoffman D. et Novak T. (1996), Marketing in hypermedia computer mediated environments: conceptual foundations, *Journal of Marketing*, 60, July, 50-68

Hoffman D. et Novak T. (1997), A new marketing paradigm for electronic commerce, *The Information Society*, 13, 1, 43-54

Hoffman D., Novak T. et Peralta M. (1999), Information privacy in the marketplace: implications for the commercial uses of anonymity on the web, *The Information Society*, 15, 2, 129-140

Hoffman D., Nowak T. et Peralta M. (1999), building consumer trust online, *Communications of the ACM*, 42, 4, April, 80-85

Hoffman D., Nowak T. et Schlosser (2000), Consumer control in online environment, www.elab.vanderbilt.edu

Hoffman D., Novak T. et Chatterjee P. (1995), Commercial scenarios for the web: opportunities and challenges, *Journal of CMC*, 1, 3

Homans G. (1961), *Social Behaviour: Its Elementary Forms*, London: Routledge and Kegan Paul

Hood T. et Back K. (1971), Self-disclosure and the volunteer: a source of bias in laboratory experiments, *Journal of Personality and Social Psychology*, 17, 2, 130-136

Hornback K. (1971), Toward a theory of involvement propensity for collective behaviour, *Sociological Forces*, 4, Summer, 61-77.

Horne D. et Horne D. (1997), Privacy : a paranoid's view, *Advances in Consumer Research*, 24, 351-354

Houston F. et Gassenheimer J. (1987), Marketing and Exchange, *Journal of Marketing*, 51, October, 3-18

Howell D. (1998), *Méthodes Statistiques en Sciences Humaines*, DeBoeck Université, Paris

Hoyle R. et Panter A. (1995), Writing about structural equation models, dans *Structural Equation Modelling*, R. H. Hoyle (Ed.), Thousand Oaks, CA: Sage, 158-176

Hu L. et Bentler P. (1999), Cut-off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, *Structural Equation Modelling*, *6*, 1-55

Huber J., Payne J. et Puto C. (1982), asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis, *Journal of Consumer Research*, 1, June, 3-42

Hugues A. (1994), *Strategic database marketing: the masterplan for starting and managing a profitable, customer-based marketing program*, Chicago: Richard D. Irwin

Hui K., Lee T. et Teo H. (2004), Information privacy and online participation : a field experiment, papier de recherche, online

Iacobucci D. (2001), Should I treat a participant attribute as a blocking factor? *Journal of Consumer Psychology*, Special Issue: Methodological and Statistical Concerns of the Experimental Researcher, 10, 1 et 2, 18-20

Iacobucci D. (2001), Unequal Cell Sizes (Missing Data), *Journal of Consumer Psychology*, Special Issue: Methodological and Statistical Concerns of the Experimental Researcher, 10, 1 et 2, 18-20

Igbaria M. et Parasuraman S. (1989), A path analytic study of individual characteristics, computer anxiety and attitudes toward microcomputers, *Journal of Management*, 15, 3, 373-388

Ingargiola S. (2002), Realizing privacy in the information age: an exercise in uncovering the values shaping the online privacy debate, Doctoral Dissertation, Georgetown University, Washington DC

Inness J. (1992), *Privacy, Intimacy and Isolation*, Oxford: Oxford University Press

Jacobs P. (1997), Privacy: what you need to know, *InfoWorld*, 19, 44, 111-112

Jarvenpaa S. et Tractinsky N. (1999), Consumer trust in an Internet store: a cross-cultural validation, *Journal of CMC*, 5, 2

Jarvis S. (2001), Maybe this year stricter internet privacy laws may emerge, *Marketing News*, 35, 9, 14-15

Jensen C., Potts C. et Jensen C. (2005), Privacy practices of Internet users: self-reports versus observed behavior, *International Journal of Human Computer Studies*, Special issue on HCI Research in Privacy and Security, October

Johnson E., Bellman S. et Lohse G. (2002), Defaults, framing and privacy : why opting in-opting out, *Marketing Letters*, 13, 1, 5-15

Johnson J. (1992), A theory of the nature and value of privacy, *Public Affairs Quarterly*, 6, 3, 271-288

Jones M. (1991), Privacy: a significant marketing issue for the 1990s, *Journal of Public Policy and Marketing*, 10, 1, 133-148

Jones R. (1999), *Méthodes de Recherche en Sciences Humaines*, DeBoeck Université, Paris

Jones E. et Archer R. (1976), Are there special effects of personalistic self-disclosure?, *Journal of Experimental Social Psychology*, 12, 180-193

Jones T. et Verstegen L. (1992), A moral approbation model of ethical decision making in organizations, Annual Meeting of the Society for Economics, Irvine, CA

Joreskog K. (1969), A general approach to confirmatory maximum likelihood factor analysis, *Psychometrika*, 34, 183-202

Joreskog K. (1971), Statistical analysis of sets of congeneric tests, *Psychometrika*, 36, 2, 109-133

Joreskog K. et Sobom D. (1984), LISREL V7 User's Guide, Indiana: Scientific Software

Joule R. et Beauvois J. (1998) *La soumission librement consentie*, Presses universitaires de France, Paris

Jourard S. (1959), Healthy personality and self-disclosure, *Mental Hygiene*, 43, 4,499-507

Jourard S. (1964), *The transparent self: self-disclosure and well-being*, Princeton, N.J., Van Nostrand

Jourard S. (1966), Some psychological aspects of privacy, *Law and Contemporary Problems*, 31, 307-318

Jourard S. (1971), *The transparent self*, New York: Van Nostrand

Jourard S. (1974), *Healthy personality*. New York: MacMillan Co

Jourard S. et Lasakow P. (1958), Some factors in self-disclosure, *Journal of abnormal and Social Psychology*, 56, 91-98

Kanuk L. et Berenson C. (1975), Mail surveys and response rates: a literature review, *Journal of Marketing Research*, 12, November, 440-453

Kassarjian H. (1971), Personality and consumer behavior: a review, *Journal of Marketing Research*, 8, November, 409-418

Katz J. et Tassone A. (1990), Public opinion trends: privacy and information technology, *Public Opinion Quarterly*, 24, 1, 125-143

Kehoe C., Pitkow J. et Morton K. (1997), *GVU's 8th WWW User Survey*, http://www.gvu.gatech.edu/user_surveys

Kehoe C., Pitkow J. et Rogers J. (1998), *GVU's 9th WWW User Survey*, http://www.gvu.gatech.edu/user_surveys

Kelley H. (1973), The processes of causal attribution, *American Psychologist*, 107-128

Kelly E. et Rowland H. (2000), Ethical and online privacy issues in electronic commerce, *Business Horizons*, 3-11

Kelvin P. (1973), A social-psychological examination of privacy, *British Journal of Social and Clinical Psychology*, 12, 248-261

Kenhove P., Wijnen K. et De Wulf K. (2000), The influence of topic involvement on mail survey response behavior, Papier de recherche n° 00/94, November

Kennedy M., Ferrell L. et Le Clair D. (2001), Consumer's trust of salesperson and manufacturer: an empirical study, *Journal of Business Research*, 51, 73-86

Keppel G. (1991), *Design and analysis: a researcher's handbook*, Englewood Cliffs, Prentice-Hall, New Jersey

Kerlinger F. (1986), *Foundations of Behavioral Research*, Holt, Rinehart and Winston, New York

Khalil O. et Harcar T. (1999), Relationship marketing and data quality management, *SAM Advanced management Journal*, 26-33

Kiesler C. (1971), *The psychology of commitment: experiments linking behaviour to beliefs*, New-York: Academic Press

Kiesler S., Siegel J. et McGuire T. (1984), Social psychological aspects of computer-mediated communication, *American Psychologist*, 39, 10, 1123-1134

Kiesler S. et Sproull L. (1986), Response effects in the electronic survey, *Public Opinion Quarterly*, 50, 402-413

Killingsworth S. (1999), Minding your own business: privacy policies in principle and practice, *Journal of Intellectual Property Law*, 7, fall, 57-82

Kimmel A. (1996), *Ethical issues in behavioural research: A study*, Cambridge, MA: Blackwell

King W. et Miles E. (1995), A quasi-experimental assessment of the effect of computerizing non-cognitive pencil-and-paper measurements: a test of measurement equivalence, *Journal of Applied Psychology*, 80, 643-651

Klitsch J. (1997), Database put the direct in direct marketing, *Marketing Health Services*, 4-7

Klopfer P. et Rubenstein D. (1977), The concept of privacy and its biological basis, *Journal of Social Issues*, 33, 52-65

Knapp F. et Heidingsfelder M. (2001), Drop-out analysis: Effects of the survey design, dans *Dimensions of Internet science*, U.-D. Reips & M. Bosnjak Eds., Lengerich: Pabst Science Publishers, 221-230

Koller M. (1998), Risk as a determinant of trust, *Basic and Applied Social Psychology*, 9, 4, 265-276

Korgaonkar P. et Wolin L. (1999), A multivariate analysis of web usage, *Journal of Advertising Research*, 39, March, 53-68

Koslowski S. (2000), Attitudes and behavior : a marriage made in heaven ?, *Journal of Database Marketing*, 7, 3, 246-253

Kotler P. (1984), *Marketing Management: Analysis, Planning, and Control*, Prentice Hall, Englewood Cliffs, New Jersey

Krishnamurthy S. (2001), A comprehensive analysis of permission marketing, *Journal of CMC*, 6, 2

Krysan M., Schuman H., Scott L. et Beatty P. (1994), Response rates and response content in mail versus face to face surveys, *Public Opinion Quarterly*, 58, 381-399

Kuhfeld W., Tobias R. et Garratt M. (1994), Efficient design with marketing research applications, *Journal of Marketing Research*, 31, 545-557

Kumar N., Scheer L. et Steenkam J. (1995), The effects of perceived interdependence on dealer attitudes, *Journal of Marketing Research*, 32, August, 348-356

Lacoeuilhe J. (2000), L'Attachement à la marque : proposition d'une échelle de mesure, *Recherche et Application en Marketing*, 15, 4, 61-78

Lambert K. et Brittan G. (1970), *An Introduction to the Philosophy of Science*, Englewood Cliffs, NJ: Prentice-Hall

Lambin J-J. (1990), *La Recherche Marketing*, Paris, McGraw - Hill

Lancelot Miltgen C. (2002), Vie privée et Internet : analyse de l'influence des caractéristiques individuelles et situationnelles sur les attitudes et les comportements des internautes face à la collecte et au traitement de leurs données personnelles à des fins commerciales, Mémoire de DEA, Université Paris Dauphine

Lancelot Miltgen C. (2003), Vie privée et Internet : influence des caractéristiques individuelles et situationnelles sur les attitudes et les comportements des internautes face à la collecte de données personnelles, *Congrès de l'Association Française du Marketing*, Tunis

Lancelot Miltgen C. et Gauzente C. (2006), Vie privée et partage de données personnelles en ligne : une approche typologique, *Congrès de l'Association Française du Marketing*, Nantes

Lancelot Miltgen C. et Volle P. (2005), Préoccupation des clients pour le respect de la vie privée et réponse à une sollicitation de données personnelles, *Congrès de l'Association Française du Marketing*, Nancy

LaPiere R. (1934), Attitudes vs. action, *Social Forces*, 13, 2, 230-237

Larson L., Piersel W., Imao R. et Allen S. (1990), Significant predictors of problem-solving appraisal, *Journal of Counseling Psychology*, 37, 482-490

Laufer R., Proshansky H. et Wolfe M (1976), Some Analytic Dimensions of Privacy, dans *Environmental Psychology: People and Their Physical Settings*, Proshansky, Ittelson and Rivlin, New York, Holt, Rinehart and Winston, 206-217

Laufer R. et Wolfe M. (1977), Privacy as a concept and a social issue: a multidimensional developmental theory, *Journal of Social Issues*, 33, 22-42

Laurenceau J. et Barrett L. (1998), Intimacy as an interpersonal process: the importance of self disclosure, partner disclosure and perceived partner responsiveness in interpersonal exchanges, *Journal of Personality and Social Psychology*, 74, 5, 1238-1251

Laurent G. et Kapferer J-N. (1986), Les profils d'implication, *Recherche et Applications en Marketing*, 1, 1, 41-58

Lawler E. et Yoon J. (1996), Commitment in exchange relations: test of a theory of relational cohesion, *American Sociological Review*, 61, 1, 89-108

Lefcourt H. (1966), Internal versus external control of reinforcement: a review, *Psychological Bulletin*, 65, 4, 206-220

Lemley M. (2000), Rational ignorance at the patent office, Cahier de recherche, Berkeley Olin Program in Law and Economics

Levin F. et Gergen K. (1969), Revealingness, ingratiation and the disclosure of self, *77th Annual Convention of the American Psychological Association*, 447-448

Lewin K. (1948), Some socio-psychological differences between the United States and Germany, dans *Resolving social conflicts: selected papers on group dynamics*, Lewin G. ed., New York: Harper

Lewington J. (1999), The evolution of database marketing systems to attain competitive advantages, *Journal of Database Marketing*, 6, 3, 203-221

Liebermann Y. et Stashevsky S. (2002), Perceived risks as barriers to internet and e-commerce usage, *Qualitative Market Research*, 5, 4, 291-300

Linsky A. (1975), Stimulating responses to mailed questionnaires: a review, *Public Opinion Quarterly*, 39, 82-101

Liu C. et al. (2005), Beyond concern: a privacy-trust-behavioral intention model of electronic commerce, *Information and Management*, 42, 289-304

Lohse G., Bellman S. et Johnson E. (2000), Consumer buying behavior on the internet: finding from panel data, *Journal of Interactive Marketing*, 14, March, 15-29

Long G. et al. (1999), Relationship marketing and privacy: exploring the thresholds, *Journal of Marketing Practice*, 5, 1, 4-20

Lord R. et Smith J. (1983), Theoretical, information processing and situational factors affecting attribution theory models of organizational behaviour, *Academy of Management Review*, 8, 1, 50-60

Lord R. et Maher K. (1990), Alternative information processing models and their implications for theory research and practice, *Academy of Management Review*, 15, 1, 9-28

Lubin B., Levitt E. et Zuckerman M. (1962), Some personality differences between responders and non-responders to a survey questionnaire, *Journal of Consulting Psychology*, 26, 2, 192

Luo X. (2002), Trust production and privacy concerns on the internet: a framework based on relationship marketing and social exchange theory, *Industrial Marketing management*, 31, 111-118

Lutz R. (1991), The role of attitude theory in marketing, dans *Perspectives in Consumer Behavior*, Kassarjian H. & Robertson T. eds., 317-339

Lwin M. et Williams J. (2003), A model integrating the multidimensional developmental theory of privacy and theory of planned behavior to examine fabrication of information online, *Marketing Letters*, 14, 4, 257-272

Lynch J. (1982), On the external validity of experiments in consumer research, *Journal of Consumer Research*, 9, 3, 225-239

Lynch J. (1999), Theory and external validity, *Academy of Marketing Science*, 3, Summer, 367-376

Lynn S. (1978), Three theories of self-disclosure exchange, *Journal of Experimental Social Psychology*, 14, 466-479

Lynch P. et Beck J. (2001), Profiles of Internet buyers in 20 countries: evidence for region-specific strategies, *Journal of International Business Studies*, 32,725-748.

Lysaker R. (1989), Data collection methods in the US, *Journal of the Market Research Society*, 31, 4, 477-488

Mabley K. (2000), Privacy versus personalization, papier de recherche, http://www.cyberdialogue.com/library/pdfs/wp-cd-2000-privacy.pdf

MacGraw H. (1999), Managing the privacy revolution, *Direct Marketing*, 62, 36-38

Machleit K. et Mantel S. (2001), Emotional response and shopping satisfaction moderating effects of shopper attributions, *Journal of Business Research*, 54, 97-106

MacInnis D. et Jaworski B. (1989), Information processing from advertisements: Toward an integrative framework, *Journal of Marketing*, 53, 4, 1-23

MacKenzie S. (2001), Opportunities for improving consumer research through latent variable structural equation modelling, *Journal of Consumer Research*, 28, 159-166

MacNeil I. (1980), *The New Social Contract: An Inquiry into Modern Contractual Relations*, New Haven, CT: Yale University Press

Madden T., Scholder Ellen P. et Ajzen I. (1992), A comparison of the theory of planned behavior and the theory of reasoned action, *Personality and Social Psychology Bulletin*, 18, 1, 3-9

Malhotra N. (1993), *Marketing Research: An Applied Orientation*, Prentice-Hall International Edition, New-Jersey

Malhotra N. (2004), Analyse typologique, *Etudes marketing avec SPSS*, Pearson Education, Paris

Margulis S. (1977), Conceptions of privacy: current status and next steps, *Journal of Social Issues*, 33, 5-10

Margulis S. (2003), On the status and contribution of Westin's and Altman's theories of privacy, *Journal of Social Issues*, 59, 2, 411-429

Margulis S. (2003), Privacy as a social issue and behavioral concept, *Journal of Social Issues*, 59, 2, 243-261

Manfreda K., Batagelj Z. et Vehovar V. (2002), Design of web survey questionnaires: three basic experiments, *Journal of Computer Mediated Communication*, 7, 3

Marshall N. (1974), Dimensions of privacy preferences, *Multivariate Behavioral Research*, July, 255-271

Martin C. (1994), The impact of topic interest on mail survey response behaviour, *Journal of the Market Research Society*, 36, 4, 327-338

Mason R. (1986), Four ethical issues of the information age, *MIS Quarterly*, 10, 1, 5-12

Masterson S. (2001), A trickle-down model of organizational justice: relating employee' and customers' perceptions of and reactions to fairness, *Journal of Applied Psychology*, 86, 4, 594-604

Matheson K. et Zanna M. (1988), Self-awareness in computer-mediated communication, *Computers in Human Behavior*, *4*, 221-233

Matheson K. et Zanna M. (1989), Persuasion as a function of self-awareness in computer-mediated communication, *Social Behaviour, 4*, 99-111

Mauss M. (1990), *The Gift: forms and functions of exchange in archaic societies*, London: Routledge

Maxham J. et Netemeyer R. (2003), Firms reap what the sow: the effects of shared values and perceived organizational justice on customers' evaluations of complaint handling, *Journal of Marketing*, 67, 1, 46-63

Mayer T. (2002), Privacy and confidentiality research and the U.S. Census Bureau : recommendations based on a review of the literature, Research Report Series, February

Maynard M. et Taylor C. (1996), A comparative analysis of Japanese and US attitudes toward direct marketing, *Journal of Direct Marketing*, 10, 1, 34-44

McArthur L. (1972), The how and what of why: some determinants and consequences of causal attribution, *Journal of Personality and Social Psychology*, 22, 2, 171-193

McCloskey H. (1980), Privacy and the right to privacy, *Philosophy*, 55, 17-38

McCrohan K. (2001), Information technology, privacy and the public good, *Journal of Public Policy and Marketing*, 8, 266-278

McDaniel S. et Rao C. (1981), An investigation of respondent anonymity's effect on mailed questionnaire response rate and quality, *Journal of Market Research Society*, 23, 150-160

McDonagh E. et Rosemblum A. (1965), A comparison of mailed questionnaires and subsequent structured interviews, *Public Opinion Quarterly*, 29, 1, 131-136

McDonald C. (1991), Sponsorship and the image of the sponsor, *European Journal of Marketing*, 25, 11, 30-38

Mc Grath J. et Brinberg D. (1983), External validity and the research process: a comment on the Calder/Lynch dialogue, *Journal of Consumer Research*, 10, 115-124

McKim R. (2001), Privacy notices: what they mean and how marketers can prepare for them, *Journal of Database Marketing*, 9, 1, 79-84

McMellon C., Schiffman L. et Sherman E. (1997), Consuming cyber seniors: some personal and situational characteristics that influence their online behavior, *Advances in Consumer Research*, 24, 517-521

Mehta R. et Sivadas E. (1995), Comparing response rates and response content in mail versus electronic mail surveys, *Journal of the Market Research Society*, 17, 4, 429-439

Mehta R. et Sivadas E. (1995), Direct marketing on the internet: an empirical assessment of consumer attitudes, *Journal of Direct Marketing*, 9, 3, 21-32

Meltzer M. (2000), E-mining: myth and magic, CRM Forum Resources, http://www.crm-forum.com

Merton R., Fiske M. et Kendall P. (1990), *The focused interview: a manual of problems and procedures*, London: Collier MacMillan

Metcaf L., Frear C. et Krishnan R. (1992), buyer-seller relationships: an application of the IMP Interaction model, *European Journal of Marketing*, 26, 2, 27-46

Meyer J. et Allen N. (1991), A three-component conceptualization of organizational commitment, *Human Resource Management Review*, 1, 1, 61-89

Meyers-Levy J. et Sternthal B. (1991), Gender differences in the use of message cues and judgments, *Journal of Marketing Research*, 28, February, 63–70

Middleton H. (2000), Data warehouse or marketing database ? a case study, *Journal of Database Marketing*, 7, 4, 381-384

Milberg S., Burke S., Smith J. et Kallman E. (1995), Values, personal information, privacy and regulatory approaches, *Communication of the ACM*, 38, 12, 65-74

Miles M. et Huberman A. (1991), *Qualitative Data Analysis*, Sage Publications, Newbury Park, CA

Miligan G. (1980), An examination of the effects of six types of error perturbation on fifteen clustering algorithms, *Psychometrika*, 45, 325-341

Miller A. (1971), *The Assault on Privacy: Computers, Data Banks and Dossiers*, New York: Mentor Books

Miller F., Smith E. et Uleman J. (1981), Measurement and interpretation of situational and dispositional attributions, *Journal of Experimental Social Psychology*, 17, 81-95

Miller D. et Ross M. (1975), self-serving biases in the attribution of causality: fact or fiction ?, *Psychological Bulletin*, 82, 2, 213-225

Miller J. et Arning R. (2003), How companies can benefit by addressing privacy issues, *The CPA Journal*, May, 14

Milne G. (1997), Consumer participation in mailing lists: a field experiment, *Journal of Public Policy and Marketing*, 16, 2, 298-309

Milne G. (2000), Privacy and ethical issues in database / interactive marketing and public policy: A research framework and overview of the special issue, *Journal of Public Policy and Marketing*, 19, 1, 1-6

Milne G., Beckman J. et Taubman M. (1996), Consumer attitudes toward privacy and direct marketing in Argentina, *Journal of Direct Marketing*, 10, 1, 22-33

Milne G. et Boza M. (1998), A business perspective on database marketing and consumer privacy practices, *Marketing Science Institute*, Papier de Recherche n°98-110

Milne G. et Boza M. (1999), Trust and concern in consumers' perceptions of marketing information management practices, *Journal of interactive Marketing*, 13, 1, 5-24

Milne G. et Culnan M. (2002), Using the content of online privacy notices to inform public policy: a longitudinal analysis of the 1998-2001 U.S. web surveys, *The Information Society*, 18, 345-359

Milne G. et Culnan M. (2004), Strategies for reducing online privacy risks: why consumers read (or don't read) online privacy notices, *Journal of Interactive Marketing*, 18, 3, 15-29

Milne G. et Gordon M. (1994), A segmentation study of consumers' attitudes toward direct mail, *Journal of Direct Marketing*, 8, 2, 45-52

Milne G. et Gordon M. (1993), Direct mail privacy efficiency trade-offs within implied social contracts framework, *Journal of Public Policy and Marketing*, 12, 2, 206-219

Milne G. et Rohm A. (2000), Consumer privacy and name removal across direct marketing channels: exploring opt-in and opt-out, *Journal of Public Policy et Marketing*, 19, 2, 238-249

Mimouni A. (2002), Les programmes de fidélisation : approche par les bénéfices consommateurs et lien avec la qualité de la relation, Mémoire de DEA, Université Paris Dauphine

Miniard P., Bhatla S., Lord K., Dickson P. et Unnava H. (1991), Picture-based persuasion processes and the moderating role of involvement, *Journal of Consumer Research*, 18, June, 92-107

Mitra K., Reiss M. et Capella L. (1999), An examination of perceived risk, information search and behavioral intentions in search, experience and credence services, *The Journal of Services Marketing*, 13, 3, 208-228

Miyazaki A. et Fernandez A. (2000), Internet Privacy and security: an examination of online retailers disclosures, *Journal of Public Policy and Marketing*, 19, 1, 54-61

Miyazaki A. et Fernandez A. (2001), Consumer perceptions of privacy and security risks for online shopping, *The Journal of Consumer Affairs*, 35, 1, 27-44

Miyazaki A. et Krishnamurthy S. (2002), Internet seals of approvals: effects on online privacy policies and consumers perceptions, *The Journal of Consumer Affairs*, 36, 1, 28-35

Mizerski R., Golden L. et Kernan J. (1979), The attribution process in consumer decision making, *Journal of Consumer Research*, 6, September, 123-140

Moller K. et Halinen A. (2000), Relationship marketing theory: its roots and direction, *Journal of Marketing Management*, 16, 29-54

Moon Y. (2000), Intimate exchanges: using computers to elicit self disclosure from consumers, *Journal of Consumer Research*, 26, 323-339

Moore W. et Benbasat I. (1991), Development of an instrument to measure the perceptions of adopting an information technology innovation, *Information System Research*, 2, 3, 192-222

Moore J. et McDonald S. (1987), The Census community awareness program : an evaluation of the potential and actual effectiveness of CCAP based on evidence from the 1986 Los Angeles Census test, Report prepared for the bureau of the Census, February

Moorman C., Deshpande R. et Zaltman G. (1993), Factors affecting trust in market research relationship, *Journal of Marketing*, 57, 1, 81-103

Moorman C., Zaltman G. et Deshpande R. (1992), Relationships between providers and users of market research : the dynamics of trust within and between organizations, *Journal of Marketing Research*, 29, August, 314-328

Morgan R. et Hunt S. (1994), The commitment-trust theory of relationship marketing, *Journal of Marketing*, 58, 20-38

Morris L. et Pharr S. (1992), Invasion of privacy: a dilemma for marketing research and database technology, *Journal of Systems Management*, 10-43

Muchielli R. (1982), *L'analyse de contenu des documents et des communications*, Éditions ESF, Paris

Mueller C. et Price J. (1990), Economic, psychological and sociological determinants of voluntary turnover, *Journal of Behavioral Economics*, 9, 321-335

Nantel J. et Elie C. (2002), A study of the major factors influencing the perception consumers have about how their privacy is being invaded while surfing the net, *International Conference of Electronic Commerce Research*

Nass C., Steuer J., et Tauber E. (1994), Computers are social actors, *CHI Conference*, Boston, MA, 72-77

Newell P. (1995), Perspectives on privacy, *Journal of Environmental Psychology*, 15, 87-104

N'Goala G. (1998), Epistémologie et théorie du marketing relationnel, *Congrès de l'Association Française du Marketing*, Bordeaux, 473-495

Nguyen D., Gray G. et Kira D. (2000), Trust and quality assurance in business-to-consumer electronic commerce: enhancing consumer acceptance and participation, *Proceedings of the Pacific Asia Conference on Information Systems*, June

Nissenbaum H. (1998), Protecting privacy in an information age: The problem of privacy in public, *Law and Philosophy*, 17, 559-596

Nowak G. et Phelps J. (1992), Understanding privacy concerns, *Journal of Direct Marketing*, 6, 4, 28-39

Nowak G. et Phelps J. (1995), Direct marketing and the use of individual level consumer information: determining how and when privacy matters, *Journal of Direct Marketing*, 9, 3, 46-60

Nuchols R. (1964), Personal interview versus mail panel survey, *Journal of Marketing Research*, February, 11-16

Nunnally J. (1978), *Psychometric theory*, New York, McGraw Hill

Nwachukwu S. et al. (1997), Ethics and social responsibility in marketing: an examination of the ethical evaluation of advertising strategies, *Journal of Business Research*, 39, 107-118

Nyshadham E. (2000), Privacy policies of air travel web sites: a survey and analysis, *Journal of Air Transport Management*, 6, 143-152

O'Cass A. et Fenech T. (2003), Web retailing adoption: exploring the nature of Internet users web retailing behaviour, *Journal of Retailing and Consumer Services*, 10, 1, 81-94

O'Connor P. (2003), What happens to my information if I make a hotel booking online: an analysis of online privacy policy use, content and compliance by the international hotel companies, *Journal of Services Research*, 3, 2, 5-28

O'Donogue T. et Rabin M. (1999), Doing it now or later, *American Economic Review*, 89, 103-124

Oliva T., Oliver R. et MacMillan I. (1992), A catastrophe model for developing service satisfaction strategies, *Journal of Marketing*, 56, July, 83-95

Oliver R. et Swan J. (1989), Equity and disconfirmation perceptions as influences on merchant and product satisfaction, *Journal of Consumer Research*, 16, December, 372-383

Olivero N. et Lunt P. (2004), Privacy versus willingness to disclose in e-commerce exchanges : the effect of risk awareness on the relative role of trust and control, *Journal of Economic Psychology*, 25, 2, 243-263

Olson J. et Dover P. (1978), Cognitive effects of deceptive advertising, *Journal of Marketing Research*, 15, February, 29-38

Olsson U., Foss T., Troye S., et Howell R. (2000), The Performance of ML, GLS and WLS estimation in structural equation modelling under conditions of misspecification and nonnormality, *Structural Equation Modelling*, 7, 4, 557-595

O'Malley L., Patterson M. et Evans M. (1997), Intimacy or intrusion? The privacy dilemma for relationship marketing in consumer markets, *Journal of Marketing Management*, 13, 541-559

O'Neil D. (2001), Analysis of internet users' level of online privacy concerns, *Social Science Computer Review*, 19, 1, 17-31

Orne M. (1962), On the social psychology of the psychological experiments: with particular reference to demand characteristics and their implications, *American Psychologist*, 17, 776-783

Orne M. (1969), Demand characteristics and the concept of quasi-control, 147-177, dans *Artifact in Behavioural Research*, Rosenthal et Rosnow (1969), New York: Academic Press

Palmer J., Bailey J., Faraj S. et Smith R. (2000), The role of intermediaries in the development of trust on the www : the use and prominence of trusted third parties and privacy statements, *Journal of CMC*, 5, 3

Papageorgis D. (1968), Warning and Persuasion, *Psychological Bulletin*, 70, 271-282

Pardun C. et Lamb L. (1999), Corporate websites in traditional print advertisements, *Internet Research*, 9, 2, 93-99

Park C. et Mittal B. (1985), A theory of involvement in consumer behavior, *Research in Consumer Behavior*, 1, 201-231

Parker L. (1992), Collecting data the e-mail way, *Training and Development*, July, 52-54

Patterson M., O'Malley L. et Evans M. (1997), Database marketing: investigating privacy concerns, *Journal of Marketing Communications*, 3, 151-174

Paulhus D. et Reid D. (1991), Enhancement and denial in socially desirable responding, *Journal of Personality and Social Psychology*, 60, 2, 307-317

Pavlou P. (2003), Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model, *International Journal of Electronic Commerce*, 7, 3, 69-103

Pavlou P. et Chellappa R. (2001), The role of perceived privacy and perceived security in the development of trust in electronic commerce transaction, *ISR*

Pearl D. et Failey D. (1985), Testing for the potential for non response bias in sample surveys, *The Public Opinion Quarterly*, 49, 4, 553-560

Pedersen D. (1979), Dimensions of privacy, *Perceptual and Motor Skills*, 48, 1291-1297

Pedersen D. (1982), Personality correlates of privacy, *Journal of Psychology*, 112, 11-14

Pedersen D. (1988), Correlates of privacy regulation, *Perceptual and Motors Skills*, 66, 595-601

Pedersen D. (1997), Psychological functions of privacy, *Journal of Environmental Psychology*, 17, 147-156

Pedersen D. (1999), Model for types of privacy by privacy functions, *Journal of Environmental Psychology*, 19, 397-405

Pedersen D. et Breglio V. (1968), The correlation of two self-disclosure inventories with actual self-disclosure: a validity study, *The Journal of Psychology*, 68, 291-298

Pedersen D. et Breglio V. (1968), Personality correlates of actual self-disclosure, *Psychological Reports*, 22, 495-501

Pedersen D. et Frances S. (1990), Regional differences in privacy preferences, *Psychological Reports*, 66, 731-736

Pedersen D. et Higbee K. (1969), Personality correlates of self-disclosure, *The Journal of Social Psychology*, 78, 81-89

Peeples D. (2002), Installing consumer confidence in e-commerce, *SAM Advanced Management Journal*, Autumn, 26-31

Pels J. (1992), Identification and management of key clients, *European Journal of Marketing*, 26, 5, 5-21

Peltier J., Schibrowsky J. et Davis J. (1998), Using attitudinal and descriptive database information to understand interactive buyer-seller relationships, *Journal of Interactive Marketing*, 12, 3, 32-45

Perdue B. et Summers J. (1986), Checking the success of manipulations in marketing experiments, *Journal of Marketing Research*, 23, November, 317-326

Perrien J., Cheron E. et Zins M. (1983), *Recherche en marketing : méthodes et décisions*, G. Morin ed., Québec

Peterson R. (1995), Une méta-analyse du coefficient alpha de Cronbach, *Recherche et Applications en Marketing*, 10, 2, 75-88

Peterson R. (1995), Relationship marketing and the consumer, *Journal of the Academy of Marketing Science*, 23, 4, 278-281

Peterson R., Albaum G. et Ridgway N. (1989), Consumers who buy from direct sales companies, *Journal of Retailing*, 65, 2, 273-287

Peterson R. et Wilson W. (1992), Measuring customer satisfaction; fact and artifact, *Journal of the Academy of Marketing Science*, 20, 1, 61-71

Petrison L., Blattberg R. et Wang P. (1997), Database marketing: past, present and future, *Journal of Direct Marketing*, 11, 4, 109-125

Petrison L. et Wang P. (1993), From relationships to relationship marketing: applying database technology to public relations, *Public Relations Review*, 19, 3, 235-245

Petrison L. et Wang P. (1993), Relationship issues in creating the customer database, *Journal of Direct Marketing*, 7, 4, 54-62

Petrison L. et Wang P. (1995), Exploring the dimensions of consumer privacy: an analysis of coverage in British and American medias, *Journal of Direct Marketing*, 9, 4, 19-37

Petronio S. (1991), Communication boundary management: a theoretical model of managing disclosure of private information between marital couples, *Communication Theory*, 1, 311-335

Petty R. (2000), Marketing without consent: consumer choice and costs, privacy and public policy, *Journal of Public Policy and Marketing*, 19, 1, 42-53

Petty R. et Cacioppo J. (1986). The elaboration likelihood model of persuasion, dans *Advances in experimental social psychology*, L. Berkowitz Ed., 19, New York: Academic Press, 123-205

Phelps J., D'Souza G. et Nowak G. (2001), Antecedents and consequences of consumer privacy concerns : an empirical investigation, *Journal of Interactive Marketing*, 15, 4, 4-17

Phelps J., Gonzenbach W. et Johnson E. (1994), Press coverage and public perception of direct marketing and consumer privacy, *Journal of Direct Marketing*, 8, 2, 9-22

Phelps J., Nowak G. et Ferrell E. (1999), Marketers' information practices and privacy concerns : how willing are consumers to provide personal information for shopping benefits ?, *Marketing Science Institute*, Papier de Recherche n°99-112

Phelps J., Nowak G. et Ferrell E. (2000), Privacy concerns and consumer willingness to provide personal Information, *Journal of Public Policy et Marketing*, 19, 1, 27-41

Phillips J. et Gully S. (1997), The role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal setting process, Journal of Applied Psychology, 82, 5, 792-802

Pitkow J. et Kehoe C. (1997), Federal Trade Commission Public Workshop on consumer privacy, http://www.gvu.gatech.edu/user_survey

Pitkow J., Kehoe C. et Morton K., GVU's WWW User Surveys (1997), http://www.cc.gatech.edu/gvu/user_surveys/survey_1997

Pitta D., Franzak F. et Laric M. (2003), Privacy and one to one marketing: resolving the conflict, *Journal of Consumer Marketing*, 20, 7, 616-628

Pitta D. (1998), Marketing one-to-one and its dependence on knowledge discovery in databases, *Journal of Consumer Marketing*, 15, 5, 468-480

Plog S. (1965), The disclosure of self in the United States and Germany, *Journal of Social Psychology*, 65, 193-203

Plouffe C., Vandenbosch M. et Hulland J. (2001), Intermediating technologies and multigroup adoption : a comparison of consumer and merchant adoption intentions toward a new electronic payment system, *Journal of Product Innovation Management*, 18, 2, 65-81

Ployhart R. et Ryan A. (1997), Toward an explanation of applicant reactions : an examination of organizational justice and attribution frameworks, *Organizational Behavior and Human Decision Processes*, 72, 3, 308-335

Polin R. (2000), La culture proprement européenne au sein de la culture occidentale, dans *La protection de la vie privée dans la société de l'information*, Pierre Tabatoni (ed.), Cahier des Sciences Morales et Politiques, Presses Universitaires de France, tome 2, 1-10

Post R. (2001), Three Concepts of Privacy, *Georgetown Law Journal*, 2087

Pras B. (2001), Gestion et vie privée : au croisement de plusieurs libertés, *Revue Française de Gestion*, Juin-Juillet-Août, 62-65

Prosser W. (1960), Privacy, *California Law Journal*, 48, 2, 383-423

Punch L. (1996), When Big Brother goes too far, *Credit Card Management*, October, 23

Rachels J. (1975), Why privacy is important, *Philosophy and Public Affairs*, 4, 4, 323-333

Ravald A. et Gronroos C. (1996), The value concept and relationship marketing, *European Journal of Marketing*, 30, 2, 19-30

Regan P. (1995), *Legislating privacy: technology, social values and public policy*, Chapel Hill: University of North Carolina Press

Reihheld F. et Schefter P. (2000), E-loyalty: your secret weapon on the web, *Harvard Business Review*, July-August, 4

Richards J. (1997), The legal potholes on the information superhighway, *Journal of Public Policy and Marketing*, 16, 2, 319-326

Rieunier S. (2000), L'influence de la musique sur le comportement du consommateur, Thèse de Doctorat en Sciences de Gestion, Université de Paris Dauphine

Rigaux F. (2000), Libre circulation des données et protection de la vie privée dans l'espace européen, dans *La protection de la vie privée dans la société de l'information*, Pierre Tabatoni (ed.), Cahier des Sciences Morales et Politiques, Presses Universitaires de France, tome 2, 29

Roberts M. (1997), Expanding the role of the direct marketing database, *Journal of Direct Marketing*, 11, 4, 26-35

Roberts J. et Gregor T. (1971), Privacy: a cultural view, dans *Privacy*, Pennock and Chapman, New York: Atherton Press, 199-225

Robins L. (1963), The reluctant respondent, *The Public Opinion Quarterly*, 27, 2, 276-286

Robinson R. et Pearce J. (1983), The impact of formalized strategic planning on financial performance in small organizations, *Strategic Management Journal*, 4, July-September, 197-207

Rogers J. (1989), Consumer response to advertising mail, *Journal of advertising Research*, 18-24

Rogers J. (1996), Mail advertising and consumer behaviour, *Psychology and Marketing*, 13, 2, 211-233

Rolland J. et Mogenet M. (1995), *Manuel d'utilisation du NEO-PI-R version française*, EFPA, Paris

Roscoe A., Lang D. et Sheth J. (1975), Follow-up methods, questionnaire length and market differences in mail surveys, *Journal of Marketing*, 39, April, 20-27

Rosen N. (1960), Anonymity and attitude measurement, *The Public Opinion Quarterly*, 24, 4, 675-679

Rosenbaum B. (1973), Attitude toward invasion of privacy in personal selection process and job applicant demographic and personality correlates, *Journal of Applied Psychology*, 58, 333-338

Rosenthal R. et Rosnow R. (1969), *Artifact in behavioral research*, Academic Press: New York

Rossiter J. (2002), The COARSE procedure for scale development in marketing, *International Journal of Research in Marketing* 3, 4, 179-184

Roster C. et al. (2004), A comparison of response characteristics from web and telephone surveys, *International Journal of Market Research*, 46, 3, 359-373

Rotemberg M. (1993), Communications privacy: implications for network design, *Communications of the ACM*, august, 61

Rotter J. (1966), Generalized expectancies for internal versus external control of reinforcement, *Psychological Monographs*, 80, 1, 609-633

Rotter J. (1967), A new scale for the measurement of interpersonal trust, *Journal of Personality and Social Psychology*, 651-665

Rotter J. (1980), Interpersonal trust, trustworthiness and gullibility, *American Psychologist*, 35, 1, 1-7

Roussel P., Durrieu F., Campoy E. et El Akremi A. (2002), *Méthodes d'Equations Structurelles*, Recherche et Applications en Gestion. Economica

Russel D. (1982), The causal dimension scale: a measure of how individuals perceive causes, *Journal of Personality and Social Psychology*, 42, 6, 1137-1145

Rustemli A. (1988), The effects of personal space invasion on impressions and decisions, *The Journal of Psychology*, 122, 2, 113-118

Rustemli A. (1992), Crowding effect of density and interpersonal distance, *The Journal of Social Psychology*, 132, 1, 51-58

Rustemli A. et Kokdemir D. (1993), Privacy dimensions and preferences among Turkish students, *Journal of Social Psychology*, 133, 6, 807-815

Salerno F. (2001), Web-marketing : obtenir la confiance du consommateur, *Revue Française de Gestion*, Juin-Juillet-Aout, 66-80

Sawyer A. (1975), Demand artefacts in laboratory experiments in consumer research, *The Journal of Consumer Research*, 1, 4, 20-30

Sawyer A. et Ball A. (1981), Statistical power and effect size in marketing research, *Journal of Marketing Research*, 18, August, 275-290

Schein V. (1977), Individual privacy and personnel psychology: the need for a broader perspective, *Journal of Social Issues*, 33, 154-167

Schkade D. et Kleinmuntz D. (1994), Information displays and choice processes: differential effects of organization, form and sequence, *Organizational behavior and Human Decision Process*, 57, march, 319-337

Schoeman F. (1984), Privacy: Philosophical Dimensions of the Literature, dans *Philosophical Dimensions of Privacy: An Anthology*, Schoeman ed, New York: Cambridge University Press

Schoenbachler D. et Gordon G. (1997), Understanding consumer database marketing, *The Journal of Consumer Marketing*, 14, 1, 5-20

Schoenbachler D. et Gordon G. (2002), Trust and customer willingness to provide information in database driven relationships marketing, *Journal of Interactive marketing*, 16, 3, 2-16

Schultz D. (1990), Plowing new research ground with new plowshares, *Journal of Direct Marketing*, 4, 3, 4- 6

Schwartz B. (1968), The social psychology of privacy, *The American Journal of Sociology*, 73, 6, 741-752

Schwartz B. (2000), Charting a privacy research agenda: responses, agreements, and reflections, *Connecticut Law Review*, 929

Scott C. (1961), Research on mail surveys, *Journal of The Royal Statistical Society*, 124, 143-205

Shapiro S. (1987), The social control of impersonal trust, *American Journal of Sociology*, 93, 623-658

Sharma S. (1996), Confirmatory factor analysis, dans *Applied Multivariate Techniques*, John Wiley & Sons, Inc., 144-184

Sharp L. and Frankel J. (1983), Respondent burden: a test of some common assumptions, *Public Opinion Quarterly*, 47, 36-53

Sheehan K. (1999), An investigation of gender differences, online privacy concerns and resultant behaviors, *Journal of Interactive Marketing*, 13, 4, 24-38

Sheehan K. (2001), E-mail survey responses rates : a review, *Journal of Computer Mediated Communication*, 6, 2

Sheehan K. (2002), Toward a typology of internet users and online privacy concerns, *The Information Society*, 18, 21-32

Sheehan K. et Hoy M. (1999), Using e-mail to survey internet users in the United States responses: methodology and assessment, *Journal of Computer Mediated Communication*, 4, 3

Sheehan B. et Hoy G. (1999), Flaming, complaining and abstaining: how online users respond to privacy concerns, *Journal of Advertising*, 28, 3, 37-51

Sheehan B. et Hoy G. (2000), Dimensions of privacy concern among online consumers, *Journal of Public Policy et Marketing*, 19, 1, 62-73

Sheehan K. et McMillan S. (1999), Response variation in e-mail surveys: an exploration, *Journal of Advertising Research*, 45-54

Sherman R. (1991), Rethinking privacy issues, *Direct Marketing*, April, 40-44

Sheth J. et Parvatiyar A. (1995), Relationship marketing in consumer markets: antecedents and consequences, *Journal of the Academy of Marketing Science*, 23, 4, 255-271

Sheth J. et Parvatiyar A. (1995), The evolution of relationship marketing, *International Business Review*, 4, 4, 397-418

Sheth J., Gardner D. et Garrett D. (1988), *Marketing theory: Evolution and evaluation*, Wiley & Sons, New-York

Shils E. (1966), Privacy: its constitution and vicissitudes, *Law and Contemporary Problems*, 31, 281

Shiman D. (1996), When e-mail becomes junk mail: the welfare implications of the advancement of communications technology, *Review of Industrial Organization*, 11, 35-48

Sieber J. (1982), Ethical dilemmas in social research, dans *The Ethics of Social Research: Surveys and Experiments*, J. Sieber ed., New York: Springer-Verlag, 1-29

Sigall H., Aronson E. et Van Hoose T. (1970), the cooperative subject: myth or reality?, *Journal of Experimental Social Psychology*, 6, 1-10

Simitis S. (1987), Reviewing privacy in an information society, *University of Pennsylvania Law Review*, 135, 707-746

Simmel A. (1971), Privacy is not an isolated freedom, dans *Privacy*, R. J. Pennock & J. W. Chapman Eds., New York: Atherton Press, 71-87

Simon H. (1976), From substantive to procedural rationality, dans *Method and Appraisal in Economics*, Cambridge: Cambridge University Press

Simon S. et Victor T. (1994), Customers'risk perceptions of electronic payment systems, *The International Journal of Bank Marketing*, 12, 8, 26-39

Singer E. (1978), The effect of informed consent procedures on respondents' reactions to surveys, *Journal of consumer Research*, 5, 49-57

Singer E. (1984), Public reactions to some ethical issues of social research: attitudes and behaviour, Journal of Consumer Research, 11, 501-509

Singer E., Groves R. et Corning A. (1999), Differential incentives: believes about practice, perceptions of equity and effects on survey participation, *The Public Opinion Quarterly*, 63, 2, 251-260

Singer E., Hoewyk J. et Neugebauer R. (2003), The impact of privacy and confidentiality concerns in the 2000 Census, *Public Opinion Quarterly*, 67, 368-384

Singer E., Mathiowetz N. et Couper M. (1993), the impact of privacy and confidentiality concerns on survey participation, *Public Opinion Quarterly*, 57, 465-482

Singer E. et Presser S. (1996), Public attitudes toward data sharing by federal agencies, *Annual Research Conference*, Bureau of the Census

Singh J. (1988), Consumer complaint intentions and behaviour: definitional and taxonomical issues, *Journal of Marketing*, 52, January, 93-107

Singh T. et Hill M. (2003), Consumer privacy and internet in Europe: a view from Germany, *Journal of Consumer Marketing*, 20, 7, 634-651

Siriex L. et Dubois P.-L. (1999), Vers un modèle qualité satisfaction intégrant la confiance, *Recherche et Applications en Marketing*, 14, 3, 1-22

Skinner S., Ferrell O. et Pride W. (1984), Personal and non personal incentives in mail surveys: immediate versus delayed inducements, *Journal of the Academy of Marketing Sciences*, 12, 1, 106-114

Smith H., Milberg S. et Burke S. (1996), Information privacy: measuring individuals' concerns about organizational practices, *MIS Quarterly*, 20, 2, 167-196

Smith C. (1997), Casting the net: surveying an internet population, *Journal of Computer Mediated Communication*, 3, 1

Smith T. (1983), The hidden 25% : an analysis of non response on the 1980 general social survey, *The Public Opinion Quarterly*, 47, 3, 386-404

Solove D. (2002), Conceptualizing privacy, *California Law Review*, 90, 1087-1155

Spence A. (1974), *Market signalling: informational transfer in hiring and related screening processes*, Cambridge, MA: Harvard University Press

Spiekermann S. (2005), The desire for privacy: insights into the views and nature of the early adopters of privacy services, *International Journal of Technology and Human Interaction*, 1, 1, 74-83

Spiekermann S., Grossklags J. et Berendt B. (2001), Stated privacy preferences versus actual behavior in EC environments: a reality check, dans *e-Finance*, Buhl H., Kreyer N. & Steck W. Eds., Berlin: Springer, 129-147, http://vasarely.wiwi.hu-berlin.de/Home/final_privacy.PDF

Spiekermann S., Grossklags J. et Berendt B. (2001), E-privacy in 2nd generation E-Commerce: privacy preferences versus actual behaviour, *ACM Conference on Electronic Commerce*, Tampa, 14-17 October

Sprecher S. (1992), How men and women expect to feel and behave in response to inequity in close relationships, *Social Psychology Quarterly*, 55, 1, 57-69

Sprott D., Hardesty D. et Miyazaki A. (1998), Disclosure of odds information: an empirical investigation of objective odds format and numeric complexity, *Journal of Public Policy and Marketing*, 17, 1, 11-23

Stanton J. (1998), An empirical assessment of data collection using the internet, *Personnel Psychology*, 51, 709-725

Stanton J. et Stam K. (2003), Information technology, privacy and power within organizations: a view from boundary theory and social exchange perspectives, *Surveillance and Society*, 1, 2, 152-190

Stewart K. et Segars A. (2002), An empirical examination of the concern for information privacy instrument, *Information Systems Research*, 13, 1, 36-49

Stone D. (1986), Relationship between introversion-extraversion, values regarding control over information and perceptions of invasion of privacy, *Perceptual and Motor Skills*, 62, 371-376

Stone E., Gueutal H., Gardener D. et McClure S. (1983), A field experiment comparing information privacy values, beliefs and attitudes across several types of organizations, *Journal of Applied Psychology*, 68, 3, 459-468

Stone E. et Stone D. (1990), Privacy in organizations: theoretical issues, research findings and protection mechanisms, *Research in Personnel and Human Resources Management*, 8, 349-411

Stone E., Stone D. et Hyatt D. (2003), Personnel selection procedures and invasion of privacy, *Journal of Social Issues*, 59, 2, 343-368

Storbacka K., Strandvik T. et Gronroos C. (1994), Managing customer relationships for profit: the dynamics of relationship quality, *International Journal of Service Industry Management*, 5, 5, 21-38

Strazzieri A. (1994), Mesurer l'implication durable vis-à-vis d'un produit indépendamment du risque perçu, *Recherche et Applications en Marketing*, 9, 1, 73-92.

Sweat J. (2000), Privacy paradox: customers want control and coupons, *InformationWeek*, April, 781, 52

Sykes C. (1999), *The End of Privacy*, New York: St. Martin's Press

Taylor D. (1979), Motivational bases, dans *Self-disclosure: origins, patterns and implications of openness in interpersonal relationships*, Chelune (ed.), San Francisco: Jossey-Bass, 110-151

Taylor D., Altman I. et Sorrentino R. (1969), Interpersonal exchange as a function of rewards and costs and situational factors : expectancy confirmation-disconfirmation, *Journal of Experimental Social Psychology*, 5, 324-339

Taylor D. et Altman I. (1975), Self-disclosure as a function of reward-cost outcomes, *Sociometry*, 38, 1, 18-31

Taylor D. (1968), The development of interpersonal relationships: social penetration processes, *The Journal of Social Psychology*, 75, 79-90

Taylor D. et Oberlander L. (1969), Person-perception and self-disclosure: motivational mechanisms in interpersonal processes, *Journal of Experimental Research in Personality*, 4, 14-28

Taylor S. (2003), How to build consumer trust while protecting privacy, *The Information Management Journal*, 73-74

Terrasse C. (2003), Proposition et validation d'une échelle de l'engagement à la marque, *Congrès de l'Association Française du Marketing*, Tunis

Tezinde T., Smith B. et Murphy J. (2002), Getting permission: exploring factors affecting permission marketing, *Journal of Interactive Marketing*, 16, 4, 28-36

Tih S. et Ennis S. (2004), Student subjects as internet services users surrogates, *EMAC Conference*, Milan, 24-27

Thibaut J. et Kelley H. (1959), *The Social Psychology of Groups*, John Wiley & Sons, New York

Thiétart et al. (2003), *Méthodes de recherche en management*, Dunod, Paris

Thomson J. (1975), The right to privacy, *Philosophy and Public affairs*, 4, 4, 295-314

Thomson J. (1984), The right to privacy, dans *Philosophical Dimensions of Privacy: An Anthology*, Schoeman F. ed., New York: Cambridge University Press

Tolchinsky P., McCuddy M., Adams J., Ganster D., Woodman R. et Fromkin H. (1981), Employee perceptions of invasion of privacy: a field simulation experiment, *Journal of Applied Psychology*, 66, 3, June, 308-313

Tourangeau R. et Smith T. (1998), Asking sensitive questions : the impact of data collection mode, question format and question context, *Public Opinion Quarterly*, 60, 275-304

Triandis H. (1971), *Attitude and attitude change*, John Willey and Sons, Inc., New York

Tsalikis J. et Fritzsche D. (1989), Business ethics: a literature review with a focus on marketing ethics, *Journal of Business Ethics*, 8, 695-743

Tse A. (1998), Comparing the response rate, response speed and response quality of two methods of sending questionnaires: e-mail vs. mail, *Journal of the Market Research Society*, 40, 4, 333-351

Turn R. (1985), Privacy protection, *Annual Review of Information Science and Technology*, 20, 27-50

Tuten T., Bosnjak M. et Bandilla W. (2000), Banner-advertised Web-surveys, *Marketing Research*, 11, 4, 17-21

Tybout A., Sternthal B. et Calder B. J. (1983), Information availability as a determinant of multiple request effectiveness, *Journal of Marketing Research*, 20, August, 280-290

Valette-Florence P. (1989), Conceptualisation et mesure de l'implication, *Recherche et Applications en Marketing*, 4, 1, 57-78

Van Kenhove P., Wijnen K. et De Wulf K. (2002), The influence of topic involvement on mail survey response behaviour, *Psychology and Marketing*, 19, 3, 293-301

Varian H. (1996), Economic aspects of personal privacy, www.sims.berkeley.edu

Vavra T. (1993), The database marketing imperative, *Marketing Management*, 2, 1, 46-59

Vehovar V., Manfreda K. et Batagelj Z. (2000), Design issues in web surveys, Proceedings of the Survey Research Methods Section, American Statistical Association, 983-988

Vernette E. (1991), L'efficacité des instruments d'études : évaluation des échelles de mesure, *Recherche et Applications en Marketing*, 6, 2, 43-65

Vidmar N. et Flaherty D. (1985), Concern for personal privacy in an electronic age, *Journal of Communication*, 91-103

Volle P. (1996), Impact du marketing promotionnel des distributeurs sur le choix du point de vente et rôle modérateur de variables individuelles, Thèse de Doctorat en Sciences de Gestion, Université Paris Dauphine

Vondracek F. et Marshall M. (1971), Self-disclosure and interpersonal trust: an exploratory study, *Psychological Reports*, 28, 235-240

Vondracek F. (1969), Behavioral measurement of self-disclosure, *Psychological Reports*, 25, 914

Vroom V. (1964), *Work and Motivation*, New York: Wiley

Wacheux F. (1996), *Méthodes Qualitatives et Recherche en Gestion*, Economica

Walczuch R., Seelen J. et Lundgren H. (2001), Psychological determinants for consumer trust in e-retailing, *Proceedings of the 8th Research Symposium on Emerging Electronic Markets*, Maastricht

Whalen J., Pitts R. et Wong J. (1991), Exploring the structure of ethical attributions as a component of the consumer decision model: the vicarious versus personal perspective, *Journal of Business Ethics*, 10, 285-293

Wallet-Wodka D. (2003), Adoption d'un intermédiaire d'achat par le consommateur : une comparaison réel-virtuel, Thèse de Doctorat en Sciences de Gestion, Université Paris Dauphine

Wang H., Lee M. et Wang C. (1998), Consumer privacy concerns about Internet marketing, *Communications of the ACM*, 41, 3, 63-70

Wang P. et Petrison L. (1993), Direct marketing activities and personal privacy, *Journal of Direct Marketing*, 7, 1, 7-19

Ward S., Bridges K. et Chitty B. (2005), Do incentives matter? An examination of online privacy concerns and willingness to provide personal and financial information, *Journal of Marketing Communications*, 11, 1, 21-40

Warren S. et Brandeis L. (1890), The right to privacy, *Harvard Law Review*, 4, 5

Webb E. (1981), *Nonreactive measures in the social sciences*, Boston: Houston Mifflin

Weber J. (1992), Scenarios in business ethics research: review, critical assessment and recommendations, *Business Ethics Quarterly*, 2, 2, 137-160

Weber S. et Cook T. (1972), Subject effects in laboratory research: an examination of subject roles, demand characteristics and valid inference, *Psychological Bulletin*, 77, 4, 273-295

Weible R. (1993), Privacy and data: an empirical study of the influence of types of data and situational context upon privacy perceptions, Doctoral Dissertation

Weible R. et Wallace J. (1998), Cyber research: the impact of the internet on data collection, *Marketing Research*, 10, 3, 19-31

Weisband S. et Reinig B. (1995), Managing user perceptions of email privacy, *Communications of the ACM*, 38, 12, 40-47

Weisband S. et Kiesler S. (1996), Self disclosure on computer forms: Meta-analysis and implications, Conference on Human Factors in Computing Systems, Vancouver, B.C

West L. (1971), A study of the validity of the self-disclosure inventory for adolescents, *Perceptual and Motor Skills*, 33, 91-100

West D. (1999), Modern day hunters: database and relationship marketers, *DTC Time*, 10-12

West S., Finch J. et Curran P. (1995), Structural equation models with nonnormal variables: problems and remedies, dans *Structural Equation Modelling: Concepts, Issues and Application*, Rick H. Hoyle ed., 56-75

Westin A. (1967), *Privacy and Freedom*, New York: Atheneum

Westin A. (1970), *Privacy and Freedom*, Bodley Head, London

Westin A. et Louis Harris & Associates (1991), *Equifax-Harris Consumer Privacy Survey 1991*, New York: Louis Harris & Associates

Westin A. (1992), Two key factors that belong in a macro ergonomic analysis of electronic monitoring: employee perceptions of fairness and the climate of organizational trust or distrust, *Applied Ergonomics*, 23, 1, 35-42

Westin A. (1992), The Equifax Canada Report on Consumers and Privacy in the Information Age, Equifax, Canada

Westin A. (1995), The Equifax Canada Report on Consumers and Privacy in the Information Age, Equifax, Canada

Westin A. (2003), Social and political dimensions of privacy, *Journal of Social Issues*, 59, 2, 431-453

Westin A. et Harris & Associates (1996), *1996 Equifax-Harris consumer privacy survey*, Atlanta, GA: Equifax Inc.

Westin A. et Harris & Associates (1998), *E-commerce and privacy: what net users want*, Hackensack, NJ: Privacy & American Business

Whalen J., Pitts R. et Wong J. (1991), Exploring the structure of ethical attributions as a component of the consumer decision model : the vicarious versus personal perspective, *Journal of Business Ethics*, 10, 285-293

White J. (1994), Individual characteristics and social knowledge in ethical reasoning, *Psychological Report*, 75, 627-649

White T. (2004), Consumer disclosure and disclosure avoidance: a motivational framework, *Journal of Consumer Psychology*, 14, 1-2, 41-51

Williams D. (2001), The relevance of legal awareness in e-commerce security, *Journal of Database Marketing*, 8, 3, 217-223

Wish M., Kaplan S. et Deutsch M. (1976), Perceived dimensions of interpersonal relations, *Journal of Personality and social Psychology*, 33, 4, 409-420

Wolfe M. et Golan M. (1976), Privacy and institutionalization, *Meeting of the Design Research Association*, Vancouver, B.C.

Woodruff R. (1997), Customer value : the next source for competitive advantage, *Journal of The Academy of Marketing Science*, 25, 2, 139-153

Woodman R. et al. (1982), A survey of employee perceptions of information privacy in organizations, *Academy of Management Journal*, 25, 3, 647-663

Worthy M., Gary A. et Kahn G. (1969), Self-disclosure as an exchange process, *Journal of Personality and Social Psychology*, 13, 59-63

Wortman C. (1975), Some determinants of perceived control, *Journal of Personality and Social Psychology*, 31, 2, 282-294

Wyner G. (1996), Database research, *Marketing Research*, 8, 1, 60-63

Yammarino F., Skinner S. et Childers T. (1991), Understanding mail survey response behavior : a meta analysis, *Public Opinion Quarterly*, 55, 4, 613-639

Yavas U. (1994), Research note: students as subjects in advertising and marketing research, *International Marketing Review*, 11, 4, 35-43

Yin R. (1989), *Case studies research: design and methods*, Sage Publications, Newbury Park, CA

Zaichkowsky J. (1985), Measuring the involvement construct, *Journal of Consumer Research*, 12, December, 341-352

Zaichkowsky J. (1994), The personal involvement inventory: reduction, revision and application to advertising, *Journal of Advertising*, 23, 4, 61-69

Zaltman G., Pinson C. et Angelmar R. (1973), *Metatheory and consumer research*, New York: Holt Rinehart and Winston

Zeithaml V. (1988), Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, *Journal of Marketing*, 52, 2, July, 2-22

Zeithalm V., Berry L. et Parasuraman L. (1988), Communication and control processes in the delivery of services processes, *Journal of Marketing*, 52, April, 36-58

Zeithalm V., Berry L. et Parasuraman L. (1996), The behavioral consequences of service quality, *Journal of Marketing*, April, 60, 31-46

Zerbe W. et Paulhus D. (1987), Socially desirable responding in organizational behavior : a reconception, *Academy of Management Review*, 12, 2, 250-264

Zhang Y., Wang C. et Chen J. (2001), Chinese online consumers' responses to web-based data collection efforts : a comparison with American online consumers, *Journal of Database Marketing*, 8, 4, 360-369

Vu : le président

_____

Vu : les suffragants

_____

_____

_____

_____

_____

Vu et permis d'imprimer :

Le Vice-Président du Conseil Scientifique chargé de la Recherche de l'Université Paris-Dauphine

Dévoilement de soi et réponses du consommateur face à une sollicitation de ses données personnelles : une application aux formulaires sur Internet

RÉSUMÉ (FRANCAIS)

Les entreprises doivent récolter des informations nominatives sur leurs clients, afin de leur faire des propositions commerciales attractives et ainsi les fidéliser. Toutefois, cette collecte de données personnelles et l'utilisation qui en est faite se heurtent au droit au respect de la vie privée des personnes. Ce phénomène est d'autant plus prégnant sur Internet, qui constitue de ce fait un excellent terrain d'application.
L'objectif de notre recherche est alors de mesurer et d'expliquer les réponses des consommateurs français face aux sollicitations d'informations personnelles dont ils font l'objet lors de leurs contacts avec les entreprises.
Notre travail s'appuie d'abord sur une revue de la littérature et sur deux études exploratoires et propose d'identifier l'ensemble des facteurs susceptibles d'influencer l'attitude et le comportement des consommateurs. Cela nous conduit à élaborer un modèle conceptuel visant à décrire l'ensemble du processus de réponse de l'individu, depuis son exposition à la sollicitation jusqu'à sa réponse. Ce modèle est ensuite testé par le biais d'une méthodologie expérimentale. Trois études successives sont ainsi menées, auprès de publics variés.
Les résultats permettent de valider l'ensemble des étapes du processus de réponse. Ils montrent aussi que l'individu se décide davantage en fonction de la situation dans laquelle il se trouve que d'après ses convictions personnelles ; les facteurs situationnels exercent donc une influence majeure. Si l'effet de la politique de respect de la vie privée et de la quantité de données demandées est avéré, l'influence du contexte dans lequel les informations sont sollicitées et de la familiarité avec l'entreprise sollicitante est plus délicate à appréhender.

Mots clés : Dévoilement de soi, Données personnelles, Respect de la Vie privée, Internet, Expérimentation

Self-disclosure and consumer's response behavior towards a request of its personal data: the case of Internet application forms

RÉSUMÉ (ANGLAIS)

The companies must collect personal data on their customers, in order to make drawing proposals and thus create their loyalty. However, this acquisition and the use which is made of, run up against the right to privacy. This phenomenon is even more important on the Internet, which actually constitutes an excellent ground of application.
The objective of our research is to measure and explain the answers of the French consumers towards the requests of personal information. Our work is first based on a review of the literature and on two exploratory studies and proposes to identify the factors having an influence on the consumers' attitudes and behaviors. This approach results in a conceptual model which aims at exposing the decision process of the individual, from its exposure to the request, to its answer. Then, we test this model by the means of an original experimental methodology. Three successive studies are thus undertaken, with a varied public. The results make possible to validate all the stages of the process of answer. They also show that the individual decides more on the basis of situation than according to his personal convictions; so the situational factors have a major impact. If the effect of the privacy policy and the quantity of required data is proven, that corresponding to the context in which information is requested and to the level of familiarity with the company is more delicate to apprehend.

Key words : Self-disclosure, Personal data, Privacy, Internet, Experiment