

GRAZIA CECERE*Télécom École de management, Institut Telecom***FABRICE ROCHELANDET***Université Paris-Sud*

Modèle d'affaires numériques, données personnelles et sites web

Une évaluation empirique

La stratégie des firmes est un facteur clé de la régulation des données personnelles. Mais ont-elles intérêt à investir pour respecter la vie privée de leurs clients si cela réduit leurs performances ? Cette étude¹ analyse la relation entre modèles d'affaires numériques, modes d'exploitation des données personnelles et audience des sites web. Les auteurs mettent en évidence les catégories de sites dont le degré d'intrusion est significatif par rapport au modèle économique. Ils montrent alors que les sites avec une forte audience sont plus enclins à adopter des stratégies intrusives.

DOI:10.3166/RFG.224.111-124 © 2012 Lavoisier

1. Les auteurs tiennent à remercier les membres du laboratoire ADIS de l'Université Paris-Sud pour leurs conseils, en particulier Alain Rallet, Nicolas Soulié et Fabrice Le Guel. Ils remercient également les deux rapporteurs anonymes pour leurs remarques constructives. Ils restent évidemment seuls responsables des idées et des opinions émises dans ce papier.

La stratégie des firmes est un facteur clé pour évaluer la régulation des données personnelles (DP) dont l'enjeu est de contrôler les modes illicites de collecte et d'exploitation et d'inciter les exploitants à respecter la vie privée informationnelle des individus (Tabatoni, 2000). Les exploitations préjudiciables sont de fait nombreuses allant des messages non sollicités à des formes d'externalités « conséquentielles » comme la discrimination par les prix ou à l'embauche, ce qui justifierait un renforcement de la réglementation. Pour les défenseurs de l'autorégulation, la protection des DP s'inscrirait au contraire dans une logique vertueuse du « mieux-disant » selon laquelle chaque concurrent afficherait sa stratégie comme étant la plus respectueuse de la vie privée de ses clients (Swire, 1997). Or, dans ce débat, peu d'études empiriques se sont penchées sur la stratégie des firmes en matière d'exploitation des DP et de respect de la vie privée. La littérature repose davantage, côté offre, sur des modélisations et sur les effets de cette régulation sur les stratégies des firmes et, côté demande, sur des analyses centrées sur les comportements des consommateurs face à l'exploitation de leurs DP. Notre objectif est de mesurer la relation entre la performance des modèles d'affaires (MA) et la stratégie de gestion des DP. Les firmes ont-elles réellement intérêt à ne pas respecter la vie privée de leurs clients au risque de subir de moindres performances causées par des réactions de défiance ? Une première originalité de notre étude est de mettre en évidence une corrélation entre la nature des MA numériques et le niveau d'intrusion en matière de collecte et d'utili-

sation des DP. Une deuxième originalité est de mesurer la performance économique des MA identifiés selon leur « intrusivité ».

Le reste de l'article est composé comme suit. Une première partie propose une revue de la littérature. Une deuxième partie détaille les propositions testables empiriquement. Puis nous explicitons nos méthodes d'analyse et la construction de l'échantillon, avant, dans une dernière partie, de présenter nos principaux résultats. La conclusion formule des recommandations managériales et de politique publique.

I – LITTÉRATURE

La vie privée est une notion difficile à appréhender, car elle varie selon les époques, les cultures, les pays... et les individus eux-mêmes². Elle a donné lieu à une importante littérature en économie et en marketing (Hui et Png, 2006 ; Lanier et Saini, 2008). L'exploitation des DP par les entreprises y est présentée comme une source de bien-être, d'innovations et d'opportunités commerciales inédites, mais également de menaces sur la vie privée et d'externalités négatives pécuniaires (Rochelandet, 2010).

1. La gestion des données personnelles par les entreprises

La protection des DP représente des intérêts différents si l'on se place du point de vue des individus, de celui des entreprises ou de celui de la société en général. Dans certains cas, ces intérêts peuvent coïncider, par exemple lorsqu'une gestion respectueuse des DP s'avère un facteur de différenciation pour les exploitants (Swire, 1997). La pression du marché serait de la sorte une régu-

2. Que la vie privée renvoie à un comportement, un état et un droit de la personne, les économistes et les chercheurs en marketing retiennent essentiellement les deux dimensions du secret et de la quiétude.

lation efficace face aux moyens coûteux de la réglementation. Pour d'autres auteurs, les entreprises n'ont aucune incitation de nature concurrentielle à protéger la vie privée de leurs clients dès lors qu'elles ont intérêt à revendre les DP et que les protéger représente, au contraire, un coût net (Thomas et Maurer, 1997). La perspective de revenus supplémentaires obtenus avec les mêmes coûts fixes de collecte et de traitement crée ainsi une incitation excessive à la revente au détriment des individus, dès lors que les revendeurs n'intègrent pas les coûts des nuisances que les exploitants subséquents sont susceptibles de faire subir aux individus. La vie privée devrait donc être avant tout protégée sur un plan légal.

Cependant, pour réguler de manière optimale, il faudrait être en mesure de quantifier l'ensemble de ces externalités. La littérature repose essentiellement, côté offre, sur des modélisations du comportement des firmes recourant notamment à la théorie des jeux (Lee et Ahn, 2011 ; Acquisti et Varian, 2005) et sur les effets de la *privacy* sur l'innovation (Goldfarb et Tucker, 2011a) et, côté demande, sur des analyses empiriques centrées sur les attitudes et les comportements individuels des consommateurs (Cecere *et al.*, 2012 ; Culnan et Armstrong, 1999). En particulier, un certain nombre d'analyses en marketing étudient les relations entre les notions de confiance, de réputation et de préoccupation pour la vie privée, le plus souvent en se focalisant sur la vente à distance et en menant des questionnaires auprès des seuls consommateurs (Lancelot Miltgen, 2011 ; Eastlick *et al.*, 2006 ; Pan et Zinkhan, 2006). Mais peu d'études empiriques se sont penchées spécifiquement sur le comportement stratégique des firmes en matière de gestion des DP.

Les premières analyses dans ce domaine ont été réalisées dans les années 1990 pour la *Federal Trade Commission* (FTC, 1998) et, en réaction, une étude baptisée *Georgetown Study* menée en 1998 pour le secteur privé. La première montre que la majorité des 600 sites internet étudiés collectait des données, tandis qu'une minorité informait sur leurs stratégies d'exploitation. Afin de montrer la supériorité de l'autorégulation, la seconde étude montrait toutefois que l'affichage des pratiques de gestion des DP sur les sites commerciaux était croissant. Critiquant la *Georgetown Study*, Culnan (2000, p. 25) souligne qu'« [i]l serait utile d'évaluer si les sites internet avec les meilleures pratiques en matière de stratégie affichée de protection de la vie privée ont quelque chose en commun ». Notre étude s'insère précisément dans cette voie en comparant les caractéristiques des sites par rapport à leurs stratégies de gestion des DP. À la suite de ces études, Miyazaki et Fernandez (2000) étudient les pratiques des sites commerciaux en détaillant 17 catégories de produits, mais se contentent d'opérer un tri croisé sans chercher à établir de corrélation plus poussée. Milne et Culnan (2002) étudient l'évolution de la gestion des DP par les firmes et leur conformité à la loi : les sites les plus populaires sont plus enclins à afficher des chartes conformes aux préconisations de la FTC. Mais rien n'est dit sur leur niveau d'intrusion. Par la suite, ces auteurs et d'autres s'orientent sur le problème de la lisibilité de ces chartes (Pollach, 2005 ; Gurau *et al.*, 2003). Milne *et al.* (2006) montrent ainsi sur un échantillon de 312 chartes que leur qualité (longueur et complexité) se détériore entre 2001 et 2003. Autre facette de la gestion des DP, l'usage des cookies n'est pas en soi une violation de

la vie privée, si elle respecte l'anonymat de l'internaute ou si ce dernier consent à des usages plus intrusifs. En revanche, non consentis, leur collecte, leur recoupement avec d'autres informations, leur conservation sur une période non justifiée peuvent entraîner des résistances lorsque ces pratiques ne sont pas déclarées ouvertement (Chellappa et Sin, 2005 ; Miyazaki, 2008).

2. Gestion des données personnelles et modèle économique

L'exploitation des données est quant à elle difficile à quantifier sur un échantillon large d'entreprises. Certaines affaires comme le cas médiatisé d'Amazon.com ayant opéré en 2000 une discrimination par les prix entre ses clients ont été largement utilisées par la littérature (Acquisti et Varian, 2005). D'autres auteurs comme Sarathy et Robertson (2003) explicitent les déterminants de la protection des DP par les firmes (environnement institutionnel, caractéristiques) mais ne proposent aucune analyse empirique.

À notre connaissance, une seule étude établit une corrélation entre l'activité des sites et leur stratégie de gestion des DP (Park, 2011). Les firmes auraient d'autant plus tendance à afficher des mesures protectrices que leur site est ancien et leur audience internet élevée, tandis que leur modèle d'affaires n'aurait pas d'effet sur le degré de *privacy* affichée. Pour autant, la caractérisation binaire des modèles d'affaires, selon qu'il s'agisse d'« e-commerce » ou non, limite la portée des résultats obtenus. Or, les services sont très hétérogènes sur internet : informations, réseaux social, etc., tous justifiant des stratégies de gestion des DP plus ou moins intrusives. De plus, Park (2011) cantonne

son échantillon aux États-Unis, soit un régime juridique plutôt minimaliste pour les obligations légales pesant sur les firmes. L'intérêt de notre analyse est de porter sur des sites accessibles en France, donc soumis à une législation plus contraignante. L'analyse des chartes met ainsi en évidence une partie des stratégies de gestion des DP.

3. Les modèles d'affaires numériques

Il existe de très nombreuses définitions à propos de la notion de modèle d'affaires (Zott *et al.*, 2011), laquelle renvoie à la manière dont une firme génère des revenus à partir d'une proposition de valeur originale, de schémas d'investissement et de revenus et d'une combinaison spécifique des ressources (compétences propres, acquisition de ressources externes).

Un MA numérique se décline partiellement ou intégralement en ligne. Certaines firmes se servent ainsi d'internet uniquement comme vitrine publicitaire. À l'opposé, d'autres firmes génèrent leurs revenus en réalisant intégralement en ligne leurs transactions (Google, eBay). Comprendre un modèle d'affaires numérique ne se résume pas donc aux modes de rémunération des sites internet, que les revenus correspondent à une marge sur des ventes de produits en ligne ou à des recettes publicitaires. Ainsi un certain nombre de travaux proposent des méthodes plus étendues pour opérer des typologies de modèles d'affaires. Dubosson-Torbay *et al.* (2002) avancent des critères tels que le rôle de l'utilisateur ou la nature de l'offre.

Cependant, les typologies proposées sont rarement satisfaisantes en raison de l'évolution des modèles et du chevauchement des catégories proposées. Un même service en ligne peut reposer sur différents types de

revenus ou proposer des valeurs différentes à ses utilisateurs : agréger des contenus, mettre en relation ses utilisateurs en créant une communauté, vendre des services, etc. Nous proposons donc une méthodologie originale pour tenter de pallier ces limites.

4. L'audience comme indicateur de performance ?

L'audience internet constitue un indicateur des performances des firmes opérant un modèle d'affaires numérique. Il repose sur une notion de performance vraisemblablement plus adaptée aux pratiques de l'économie numérique que les mesures financières traditionnelles³. De plus, l'audience web est un indicateur d'autant plus pertinent qu'elle est un déterminant essentiel des tarifs proposés aux annonceurs, la publicité s'avérant une source de financement prédominante des sites internet. On retrouve ici la logique des marchés bifaces.

Dans les écosystèmes numériques, les revenus sont de plus en plus déconnectés des investissements et dépendent davantage de la valeur que leur attribuent les utilisateurs des services et que s'approprient les opérateurs dans des revenus récurrents comme les abonnements télécoms (Rallet et Rochelandet, 2011). L'audience et le nombre d'utilisateurs déterminent ainsi une partie non négligeable de ces revenus indirects récurrents. Enfin, l'audience est un élément étroitement lié à la notoriété et à la réputation des opérateurs et des sites. Cette relation a du reste été étudiée en marketing à propos du respect de la vie

privée et des risques perçus en matière d'achat sur internet (Eastlick *et al.*, 2006).

II – HYPOTHÈSES ET PRÉDICTIONS

La littérature souligne l'arbitrage entre les bénéfices liés aux innovations numériques et les coûts sociaux engendrés par l'exploitation des DP. Ces bénéfices sont toutefois difficiles à quantifier, car ils entraînent des effets divergents : des effets de réseaux qui renforcent les positions dominantes mais également permettent aux nouveaux entrants de les contester en provoquant une chute de l'audience des acteurs en place (Facebook *vs.* Myspace) ; un potentiel de valeur gigantesque mais une fragilité des modèles de revenus. Il est néanmoins possible d'identifier des MA numériques fondés sur des propositions de valeur et des schémas de financement originaux et d'évaluer leur niveau d'« intrusivité », ainsi que leur audience sur internet. Sur cette base, nous proposons d'approximer l'arbitrage innovation/coût social en déterminant, d'une part, quels modèles sont associés à des stratégies d'intrusion élevées et, d'autre part, si leur intrusivité affecte leurs performances. Ce qui donne les deux prédictions suivantes :

P1. *Le degré d'intrusion d'un service est corrélé avec la nature de son modèle économique.*

Les besoins en DP plus ou moins sensibles et leurs modes d'exploitation plus ou moins intrusifs varient selon les MA. Par exemple, un service de réseau social nécessite des DP

3. Facebook génère le plus de consultations et d'utilisateurs réguliers, ce qui reflète indéniablement une valeur sociale importante. Or, son chiffre d'affaires est relativement faible par rapport à d'autres services en ligne comme Google, lequel génère une audience comparable. Ainsi la capitalisation financière de Facebook et sa valeur sociale s'expliquent davantage par son audience web que par son chiffre d'affaires.

pour appairer au mieux ses utilisateurs (et générer l'utilité du service) et se financer (revente de profils à des partenaires commerciaux). Les sites de vente à distance utilisent des données de navigation (*clickstream data*) moins pour personnaliser leurs messages publicitaires que pour optimiser la navigation des clients (Goldfarb et Tucker, 2011b).

P2. *Le degré d'intrusion des sites internet influence ses performances. En particulier, une stratégie d'intrusion croissante est susceptible d'être associée à une audience plus faible.*

Dans un monde fondé sur une rationalité et une information parfaites, le degré d'intrusion d'un MA numérique devrait réduire ses performances. En effet, les utilisateurs d'un service en ligne trop intrusif peuvent subir des coûts liés à l'exploitation abusive

de leurs DP : *spamming*, discriminations sur le marché du travail ou du crédit, etc. Par conséquent, plus le service est intrusif, plus le bénéfice net espéré de son utilisation diminue et donc moins les individus devraient avoir tendance à l'utiliser. En particulier, un effet de réputation négatif prévaudrait et, l'anticipant, les firmes adopteraient des stratégies moins intrusives.

III – DONNÉES ET MÉTHODOLOGIE

Nous avons sélectionné en juin 2011 les 100 sites internet les plus visités en France selon le classement effectué par la société Alexa (<http://www.alexa.com>)⁴. La nature de ces sites dans l'échantillon est très hétérogène. Toutefois, les données non observables réduisent l'échantillon à 93 sites. Un relevé d'audience a été effectué afin d'opé-

MÉTHODOLOGIE : TROIS PHASES

1) Une classification ascendante hiérarchique (CAH) a été réalisée avec la méthode Ward de façon à déterminer différentes catégories de MA en fonction de variables discriminantes. Le CAH (ou classification « par agrégation ») procède par fusions successives de clusters déjà existants. À chaque étape, les deux clusters qui vont fusionner sont ceux dont la « distance » est la plus faible entre deux groupes de points. Il existe de nombreuses définitions d'une telle distance, la plus utilisée étant la distance de Ward. Une analyse discriminante (test du chi-deux) a été ensuite effectuée entre les catégories issues de la CAH et les variables dépendantes pour vérifier la robustesse de nos résultats.

2) Un indice synthétique du degré d'intrusion a été ensuite élaboré et corrélé avec les MA identifiés selon le test de Pearson.

3) Nous avons mesuré la relation entre l'audience web et les MA avec l'analyse de covariance Ancova permettant de comparer une variable quantitative dépendante avec plusieurs variables discrètes et une variable continue.

4. Alexa réalise un classement de tous les sites du web par pays en fonction du nombre de visiteurs uniques et du nombre de pages vues par chaque visiteur. Nous avons exclu les sites pornographiques ainsi que les sites purement B2B.

rer un classement des MA entre eux globalement et par catégorie.

1. Une classification des MA

Pour établir notre classification des MA numériques, nous avons adopté une approche de type « bottom-up » en nous basant sur les éléments classiques permettant d'identifier un modèle d'affaires : une proposition de valeur, un modèle de revenus et d'investissement et une combinaison de ressources (compétences propres, réseaux de partenariat). Nous avons codifié les MA des sites internet sous la forme d'une base de données originale, construite à partir d'une série de critères (tableau 1) qui émergent clairement dans la revue de la littérature. Une CAH appliquée à notre base de données a permis de mettre en évidence différentes catégories de MA numériques.

2. MA et degré d'intrusion des sites

L'intrusion est le fait pour une personne physique ou morale de s'introduire inopportunistement et sans y être conviée dans un espace non public. La littérature en marketing suggère qu'il est difficile d'avoir une mesure exhaustive et exacte du niveau d'intrusion réelle des sites internet dans la vie privée des individus. Pour notre part, nous nous focalisons sur l'intrusion *potentielle* à partir d'une analyse des stratégies de gestion des DP des sites internet de notre échantillon.

Nous n'avons donc pas intégré de mesures de l'intrusion *perçue* par les individus, par exemple à travers l'impact des modes de collecte des DP sur leurs préférences ou leurs comportements (Pan et Zinkhan, 2006). D'une part, parce qu'associer à notre étude une enquête auprès d'internautes dépasserait son cadre exploratoire. D'autre part, les travaux en économie comporte-

mentaliste soulignent que cette intrusion n'a qu'une faible incidence sur les comportements des individus.

Plus généralement, nous nous intéressons à la manière dont les exploitants articulent leurs modèles économiques et leurs stratégies de gestion des DP. Le fait qu'un site collecte des données sensibles, pratique l'*opt-in* ou utilise des *cookies flash* peut être corrélé avec une stratégie intrusive aboutissant à des exploitations non respectueuses de la vie privée. De même, une lisibilité difficile des chartes de *privacy* peut être une mesure de l'intrusion des sites (Milne et Culnan, 2002). Sur cette base, nous avons donc calculé un indice synthétique sur la base de 11 items présentés dans le tableau 2 et qui indiquent dans quelles mesures les sites internet collectent et exploitent de manière intrusive des données sur leurs utilisateurs. Nous avons ensuite testé la corrélation entre cet indice et les différents MA identifiés.

3. Audience web et MA

Afin de tester la relation entre l'audience web et les MA identifiés, nous avons effectué une analyse de covariance Ancova en prenant l'audience comme variable quantitative dépendante, ainsi que les MA identifiés (variables discrètes) et le degré d'intrusion (variable continue) comme variables indépendantes. L'audience est mesurée par l'indicateur *Traffic rank* d'Alexa, moyenne des pages les plus consultées sur les trois derniers mois.

IV – RÉSULTATS

1. Identification des modèles d'affaires

Les résultats de la CAH sont présentés dans le tableau 3. Le dendrogramme met en

Tableau 1 – Variables binaires utilisées pour la CAH

Variable	Description
Vente en ligne	Vente uniquement en ligne
Vente multican	Vente des produits en ligne et en magasin
Contenu	Distribution du contenu édité par la firme
Websearch	Moteur de recherche, un annuaire ou un portail sur d'autres sites ou services en ligne
Affiliation	Proposition de liens sponsorisés ou des espaces pour la communication professionnelle
Agrégateur	Agrégation de contenus produits ou édités par d'autres firmes
Infomed	Mise en relation des offreurs et des demandeurs
SNS	Mise en réseau de ses utilisateurs
Communauté	Le site génère explicitement des communautés
Contrib_contenu	Contributions en contenus mis en ligne par ses utilisateurs
Contrib_conseils	Contributions informationnelles produites et mises en ligne par ses utilisateurs
Abonnement	Demande le paiement d'un forfait pour l'utilisation du service sur une durée déterminée
Commissions	Le site prélève un pourcentage sur les transactions
Dons	Possibilité de faire des dons
Marché biface	Financement assuré par acteurs professionnels à partir d'une audience générée par la mise à disposition gratuite de contenus et de services*
Freemium	Les services de base offerts par le site sont gratuits et les services améliorés sont payants
Publicité « classique »	Publicité indépendamment du comportement de l'utilisateur
Publicité contextuelle	Publicité en fonction du contenu affiché à la demande de l'utilisateur
Publicité comportementale	Publicité en fonction du comportement de l'utilisateur ou de son entourage
Affiliation	Affiliation
Revente de données	Revente de DP directement ou indirectement sous la forme de profils clients (cf. Facebook)
Autres financements	Autres modes de financement
Autres retours	Fidélisation de nouveaux clients par rapport à un autre service

*Un marché biface (*two-sided market*) correspond à une plateforme technique ou de services reliant deux groupes d'utilisateurs distincts qui se procurent mutuellement des effets de réseaux positifs (par exemple, des consommateurs et des commerçants à travers le système des cartes bancaires). Il s'agit d'une notion très utilisée en économie de l'information pour expliquer nombre de modèles d'affaires numériques où des firmes peuvent tarifer avec profit des biens informationnels à un prix nul.

Tableau 2 – Mesures du degré d'intrusion

Inscription obligatoire	0/1
Données obligatoires	0/1
Données facultatives	0/1
Données sensibles	0/1
Données de contact autres que le mail	0/1
Données relationnelles	0/1
<i>Cookies</i>	0/1
<i>Cookies flash</i> (LSO)	0/1
Charte de <i>privacy</i>	0/1
Contrôle effectif de l'utilisateur	0/1
Lisibilité de la charte de <i>privacy</i>	0/1

Tableau 3 – Résultats de la CAH : identification des MA

	Services en ligne
Plateforme	01.net, Appartager, Boursorama, Clubic, eBay, EverestPoker, Leboncoin.fr, L'internaute, Megavideo, Onvasortir.com, Pap, Partypoker.fr, PayPal, Seloger.com, venteprivée, Vivastreet, Zynga.com
Téléchargement	Deezer, Leserve, ImageShack, Megaupload, Skyrock, Streamiz.com
Institutionnel	Adobe.com, Banquepopulaire, Bouygues Telecom, Caisse-épargne, Crédit-agricole, Ikea, Laposte.net, LCL, Mappy, Meteofrance, Orange, RATP, SFR, Société Générale
Information	Canalplus, Lemonde.fr, Lefigaro.fr, Leparisien.fr, Lequipe.fr, Lexpress.fr, Programme-tv
e-Commerce	3Suisses, Allociné, Amazon.fr, Apple, Brandalley, Cdiscount, Fnac, Groupon, Jeuxvidéo.com, Laredoute, Leguide.com, Microsoft, Pixmania, Priceminister, PRIXMOINSCHER, Rueducommerce, Voyagesncf.com
Contributifs	20minutes, Aufeminin, Viedemerde.fr, Bing, Comment ça marche, Conduit.com, Dailymotion, Facebook, Flickr, France2.fr, Google, Pagesjaunes, Tfl.fr, Tripadvisor, Vimeo, Voilà.fr, Wikipedia, Worldreference, Yahoo, YouTube
Blog	Blogger, Canalblog, CenterblogLinkedIn, Meetic, Msn/Windows live, MySpace, Overblog, Twitter, Viadeo, Wordpress

Note : Pour simplifier l'analyse, nous avons renommé les différents MA selon leur type d'activités prédominantes.

évidence les associations entre les firmes et les différents groupes reliés aux différents nœuds. Il présente un détail de 20 branches et nous avons retenu 7 classes de modèle d'affaires. Par ailleurs, le faible nombre d'itérations suggère la pertinence de la méthode. Les statistiques descriptives présentées dans le tableau 4 montrent une grande hétérogénéité entre les firmes de notre échantillon. Le niveau de l'indice d'intrusion est relativement élevé (5,6/10). Il apparaît également que le modèle d'affaires « Contributifs » est le plus représenté avec 21 % de l'échantillon. Il est suivi par les classes « Plateforme » et « eCommerce » (18 % chacune), puis les classes « Institutionnel »

et « Blog » avec 15 % et 13 %. Les autres modèles sont plus marginaux.

2. Degré d'intrusion et nature des MA numériques

Il existe une corrélation statistiquement significative entre l'intrusion dans la vie de privé et certains MA (tableau 5). La statistique de Pearson rejette l'hypothèse nulle pour les MA numériques « Information » et « Contributif » qui s'avèrent les plus intrusifs. Ces résultats corroborent partiellement la prédiction P1 selon laquelle le degré d'intrusion d'un service est corrélé avec la nature de son modèle économique. Les deux modèles les plus intrusifs reposent sur l'accès aux DP des individus par l'utilisa-

Tableau 4 – Statistiques descriptives

Variable	Description	Obs	Mean	Std. dev.	Min	Max
Audience	Audience du site internet en France dans les trois derniers mois	93	2,736	7,907	0,001	49,673
Indice d'intrusion	Indice synthétique qui mesure le degré d'intrusion potentielle.	93	5,634	1,852	1	10
Plateforme	Activité de mise en relation entre offreur et demandeur	93	0,183	0,389	0	1
Téléchargement	Télécharger du contenu	93	0,064	0,247	0	1
Institutionnel	Entreprises qui utilisent internet comme un complément de leurs activités de leur business	93	0,150	0,359	0	1
Information	Sites d'information	93	0,075	0,265	0	1
e-Commerce	Sites orientés essentiellement vers la vente à distance	93	0,183	0,388	0	1
Contributifs	La valeur repose sur la contribution volontaire des usagers	93	0,215	0,413	0	1
Blog	Sites permettant de créer des blogs pour l'échange d'information sur des thèmes identifiés par les usagers	93	0,129	0,337	0	1

tion de *cookies* et à travers leurs données relationnelles à la fois pour leur fonctionnement et leurs modes de financement.

3. Audience, intrusion et MA

Les résultats de l'analyse Ancova (tableau 6) montrent que le modèle est globalement

significatif⁵. L'audience est influencée positivement et significativement par le degré d'intrusion, ce qui conduit au rejet de la prédiction P2. En d'autres termes, plus la firme adopte une stratégie potentiellement intrusive en matière de gestion des DP, plus ses performances sont élevées.

Tableau 5 – Degré d'intrusion et MA

	Pearson chi2(9)	Pr
Plateforme	12,4884	0,187
Téléchargement	3,1018	0,960
Institutionnel	7,2936	0,607
Information	15,0459	0,090*
e-Commerce	12,6729	0,178
Contributifs	19,8753	0,019**
Blog	6,5416	0,685

Tableau 6 – Analyse de covariance Ancova

Root MSE	6,867		R ²	0,37	0,245
Source	Partial SS	df	MS	F	Prob>F
Model	2120,002	15	141,333	3,00	0,0008***
Indice d'intrusion	1089,836	9	121,093	2,57	0,0122**
Plateforme	191,685	1	191,685	4,06	0,0473**
Téléchargement	53,673	1	53,672	1,14	0,2894
Institutionnel	79,321	1	79,321	1,68	0,1986
Information	87,103	1	87,103	1,85	0,1781
e-Commerce	82,725	1	82,725	1,75	0,1893
Contributifs	175,736	1	175,736	3,73	0,0572*

Modèle Blog référence. Nombre d'observations : 93. Niveau de significativité * p<0,10 ; ** p<0,05 ; *** p<0,01.

5. Le modèle est globalement significatif, puisque la somme des carrés de Model SS= 2120,001 avec un degré de liberté de 15, soit une moyenne Model MS= 141,3, ce qui correspond à une statistique de Fisher de 3,00. L'Ancova donne un coefficient de détermination R² de 0,37 qui fournit une bonne identification pour un échantillon de données réelles.

Deux explications peuvent être avancées pour expliquer un tel résultat, en apparence paradoxal, basé sur la logique du « crime paie ». Premièrement, ce résultat tendrait à confirmer, côté offre, les études comportementalistes sur la tendance des individus, victimes de distorsions psychologiques, à sous-estimer la valeur de la protection de leur vie privée face aux bénéfices liés à l'usage des sites web (Acquisti et Grossklags, 2008). Il s'agit ici d'un résultat notable, car la plupart des études portant sur le fameux paradoxe de la vie privée (les préférences déclarées pour la vie privée des individus ne reflètent pas leur comportement effectif de surexposition) se basent sur des expérimentations menées exclusivement du côté de la « demande ».

Deuxièmement, nos résultats tendraient à prouver empiriquement l'arbitrage entre l'utilité du service en ligne et le coût lié à la perte de *privacy*. On peut supposer que plus la stratégie de l'exploitant de DP est intrusive et donc plus il collecte des informations sur ses utilisateurs, plus il est en position d'améliorer son offre par rapport à leurs attentes. De même, plus il collecte des données de navigation (quels sites les internautes ont-ils consultés précédemment ?) ou sur les préférences de son audience, plus il sera en mesure de poster des liens sur ces autres sites ou des sites correspondant aux préférences moyennes de son audience et donc d'orienter vers lui du flux internet. S'agissant de l'influence des modèles d'affaires, la catégorie « Contributif », associée à un degré important d'intrusion dans la vie privée, a un coefficient positif et significatif en termes d'audience, ce qui permet à nouveau de rejeter la proposition P2. En revanche, le modèle « Plateforme » n'est pas significativement corrélé avec le degré

d'intrusion (cf. *supra*) mais l'est positivement avec l'audience. Les MA basés sur la mise en relation entre les offreurs et les demandeurs ont une performance importante grâce à la proposition de contenus et ne font pas appel pour cela à une intrusion forte dans la vie privée.

CONCLUSION

Cet article présente de manière exploratoire les premiers résultats d'une étude inédite sur les stratégies de gestion des DP en prenant en compte la nature des MA numériques adoptés par les exploitants de ces ressources. Nous déterminons en premier lieu les catégories de sites web pour lesquels le degré d'intrusion est significatif par rapport au modèle économique, à savoir les réseaux contributifs et les sites d'information. En second lieu, nous montrons que les sites les mieux classés ont une probabilité plus forte d'adopter une stratégie de gestion des DP moins respectueuse de la vie privée.

Cela tendrait à prouver que la pression du marché et l'autorégulation s'avèrent inefficaces face à d'autres modes de régulation, dès lors qu'elles s'appuient sur les effets de réputation des exploitants (Rallet et Rochelandet, 2011). Cela donnerait raison aux partisans d'une régulation plus orientée vers une certification des pratiques et une labellisation des sites. La réglementation suppléerait la défaillance des consommateurs qui se protégeraient insuffisamment. Cependant, cette étude présente certaines limites du fait de sa nature exploratoire. En premier lieu, l'échantillon analysé est relativement limité en taille, ce qui peut affecter sa représentation par rapport à l'audience totale sur internet et la pertinence de

notre analyse de données. En second lieu, il manque les variables contrôle qui permettent d'établir des estimations économétriques pour mesurer le poids et la significativité de nos variables d'intérêt. Une troisième limite tient à notre mesure de l'in-

trusion qui ne demeure que partielle. Une autre extension serait d'associer à notre analyse une enquête auprès d'un échantillon large d'internautes pour évaluer leur perception à partir des catégories issues de notre classification.

BIBLIOGRAPHIE

- Acquisti A., Varian H. (2005). "Conditioning prices on purchase history", *Marketing Science*, vol. 24, n° 3, p. 367-381
- Acquisti A., Grossklags J. (2008). "What can behavioral economics teach us about privacy?", (eds) *Digital privacy*, Acquisti *et al.*, Auerbach Publications, New York, p. 363-377.
- Cecere G., Le Guel F., Soulié N. (2012). "Perceived internet privacy concerns on social networks in Europe", Document de travail ADIS.
- Chellappa R.K., Sin R.G. (2005). "Personalization versus privacy: An empirical examination of the online consumer's dilemma", *Information Technology and Management*, vol. 6, p. 181-202.
- Culnan M.J. (2000). "Protecting privacy online: Is self-regulation working?", *Journal of Public Policy et Marketing*, vol. 19, n° 1, p. 20-26.
- Culnan M.J., Armstrong P.K. (1999). "Information privacy concern, procedural fairness, and impersonal trust: An empirical investigation", *Organization Science*, vol. 10, n° 1, p. 103-115.
- Dubosson-Torbay M., Osterwalder A., Pigneur Y. (2002). "E-Business model design, classification and measurements", *Thunderbird International Business Review*, vol. 44, n° 1, p. 5-23.
- Eastlick M.A., Lotz S.L., Warrington P. (2006). "Understanding online B-to-C relationships: An integrated model of privacy concerns, trust, and commitment", *Journal of Business Research*, vol. 59, n° 8, p. 877-886.
- Federal Trade Commission (1998). *Privacy online: A report to Congress*.
- Goldfarb A., Tucker C. (2011a). "Privacy and innovation", NBER Working Paper n° 17124.
- Goldfarb A., Tucker C. (2011b). "Privacy regulation and online advertising", *Management Science*, vol. 57, n° 1, p. 57-71.
- Gurau C., Ranchhod A., Gauzente C. (2003). "'To legislate or not to legislate': A comparative study of privacy/personalisation factors affecting French, UK and US Web sites", *Journal of Consumer Marketing*, vol. 20, n° 7, p. 652-664.
- Hui K.L., Png I.P.L. (2006). "The economics of privacy", *Economics and Information Systems*, Hendershott T. (ed.) Handbooks in Information Systems, vol. 1, chapitre 9, Elsevier.
- Lancelot Miltgen C. (2011). « Vie privée et marketing. Étude de la décision de fournir des données personnelles dans un cadre commercial », *Réseaux*, vol. 29, n° 167, p. 131-166.

- Lanier C.D., Saini A. (2008). "Understanding consumer privacy: A review and future directions", *Academy of Marketing Science Review*, vol. 12, n° 2, p. 1-48.
- Lee D.J., Ahn J.H., Bang Y. (2011). "Managing consumer privacy concerns in personalization: A strategic analysis of privacy protection", *MIS Quarterly*, vol. 35, n° 2, p. 423-444.
- Milne G.R., Culnan M.J., Greene H. (2006). "A Longitudinal assessment of online privacy notice readability", *Journal of Public Policy & Marketing*, vol. 25, n° 2, p. 238-249.
- Milne G.R., Culnan M.J. (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*, vol. 18, n° 5, p. 345-359.
- Miyazaki A.D. (2008). "Online privacy and the disclosure of cookie use: Effects on consumer trust and anticipated patronage", *Journal of Public Policy & Marketing*, vol. 27, n° 1, p. 19-33
- Miyazaki A.D., Fernandez A. (2000). "Internet privacy and security: An examination of online retailer Disclosures", *Journal of Public Policy & Marketing*, vol. 19, n° 1, p. 54-61.
- Pan Y., Zinkhan G.M. (2006). "Exploring the impact of online privacy disclosures on consumer trust", *Journal of Retailing*, vol. 82, n° 4, p. 331-338.
- Park Y.J. (2011). "Provision of internet privacy and market conditions: An empirical analysis", *Telecommunications Policy*, vol. 35, p. 650-662.
- Pollach I. (2005). "A typology of communicative strategies in online privacy policies: Ethics, power and informed consent", *Journal of Business Ethics*, vol. 62, n° 3, p. 221-235.
- Rallet A., Rochelandet F. (2011a). « La régulation des données personnelles face au web relationnel : Une voie sans issue ? », *Réseaux*, vol. 29, n° 167, p. 19-47.
- Rallet A., Rochelandet F. (2011b). « Modèle d'affaires partenariaux et écosystèmes numériques », Document de travail ADIS.
- Rochelandet F. (2010). *L'économie des données personnelles et de la vie privée*, La Découverte.
- Sarathy R., Robertson C.J. (2003). "Strategic and ethical considerations in managing digital privacy", *Journal of Business Ethics*, vol. 46, n° 2, p. 111-126.
- Swire P.P. (1997). "Markets, self-regulation, and government enforcement in the protection of personal information", *Privacy and Self-Regulation in the Information Age*, U.S. Department of Commerce, Washington, D.C.
- Tabatoni P. (2000). « Vie privée et management de l'information », *Revue française de gestion*, n° 129, p. 108-116.
- Thomas R.E., Maurer V.G. (1997). "Database marketing practice: Protecting consumer privacy", *Journal of Public Policy & Marketing*, vol. 16, n° 1, p. 147-155.
- Zott C., Amit C., Massa L. (2011). "The business model: Recent developments and future research", *IESE Research Papers*.