Données personnelles et éthique : les enjeux économiques de la confiance

Par Patrick WAELBROECK [1]
Professeur d'économie, Télécom ParisTech

Nous laissons de nombreuses traces lorsque nous utilisons l'Internet ou d'autres outils numériques. Ces traces sont stockées et analysées par les moteurs de recherche et navigateurs du web. Si certains internautes laissent ces traces de manière plutôt involontaire, d'autres contribuent de manière active aux communautés socio-numériques (comme celles d'eBay, d'Amazon, de Wikipédia, de Twitter, de YouTube) en laissant des notes, des avis, des commentaires, des classements de produits et services, ou en publiant des fichiers. Par leurs traces et contributions, les internautes et les utilisateurs d'outils numériques sont des producteurs d'informations personnelles. L'économie numérique exploite ces traces et contributions pour construire ses modèles d'affaires. Les sites financés par la publicité ciblent, voire reciblent, les prospects afin de leur fournir des contenus et des prix personnalisés. De plus, les entreprises du Net telles qu'Amazon ou Netflix, pour développer leurs modèles commerciaux, utilisent ces données personnelles afin d'identifier les préférences des internautes et de leur faire des recommandations personnalisées. De même, eBay utilise les notes laissées par ses utilisateurs sur leurs transactions pour construire un système de réputation en ligne. YouTube et Facebook doivent leur existence au contenu généré par leurs utilisateurs.

Si l'économie numérique se nourrit des données personnelles fournies parfois volontairement par les internautes, il n'en demeure pas moins que de plus en plus d'utilisateurs de réseaux sociaux sont préoccupés par la manière dont leurs données sont exploitées. En effet, on compte désormais par millions les vols de données des clients de compagnies telles que Yahoo, Equifax, eBay, Sony, LinkedIn, ou encore plus récemment Cambridge Analytica. Les révélations de l'affaire Snowden sur la surveillance généralisée par les États ont également créé un sentiment de défiance envers les acteurs de l'économie numérique à tel point qu'en 2015, 21 % des internautes n'étaient prêts à partager aucune information ; ils n'étaient que 5 % en 2009 [2]. N'y a-t-il pas un coût sociétal au tout gratuit sur Internet ?

Le numérique bouleverse les conditions de l'échange à travers l'asymétrie d'information qu'il engendre, dans ce que F. Pasquale appelle la « black box society » : les utilisateurs d'outils numériques ne connaissent ni l'utilisation qui est faite de leurs données personnelles, ni le volume des données échangées par les entreprises qui les collectent. Pire encore, ces entreprises peuvent manipuler le contexte informationnel de la transaction pour mettre les individus dans un environnement qu'ils pensent être de confiance (Mantelero, 2013) afin de les inciter à divulguer plus d'informations personnelles.

Cet article cherche à apporter quelques éléments d'éclaircissement sur les enjeux économiques de la confiance, qui peut être appréhendée d'un point de vue économique par les risques liés à une transaction. Ces risques peuvent porter sur les termes d'une transaction individuelle ou sur l'environnement institutionnel dans lequel se déroule cette transaction.

(1) L'auteur remercie les membres de la chaire Valeurs et politiques des informations personnelles pour les discussions stimulantes ayant contribué à l'enrichissement de cet article.
(2) Baromètre de la confiance de l'ACSEL-CDC.

Comprendre la source économique des risques

Il s'agit de savoir tout d'abord quels types de données sont utilisés par les entreprises et quels sont les risques pour les consommateurs. La collecte de données personnelles crée des risques dans le cas de traces involontaires, laissées par un internaute dans son historique de navigation, car l'utilisateur n'a pas toujours conscience des conséquences des traitements effectués par les entreprises du numérique. Dans le cas de contributions volontaires, comme celle d'un consommateur commentant un blog ou évaluant la qualité d'un produit ou la réputation d'un vendeur, il existe toujours des risques liés au vol de données. Les données utilisées à mauvais escient peuvent conduire à des externalités négatives telles que le vol d'identité, le harcèlement en ligne et la divulgation d'informations intimes, l'utilisation frauduleuse de coordonnées bancaires, la discrimination par les prix, les publicités indésirables, ciblées ou non.

Les externalités négatives résultent d'une défaillance du marché lorsque les actions d'un agent économique exercent un effet négatif sur d'autres agents sans compensation liée à un mécanisme de marché. Ces externalités, négatives pour le consommateur, sont causées par des entreprises qui collectent trop de données par rapport à l'optimum social, et leur existence justifie économiquement la protection des données personnelles.

Le problème de l'utilisation abusive des données personnelles est renforcé par l'asymétrie d'information en ligne. D'une part, il est difficile pour un consommateur de vérifier comment ses données sont utilisées par les compagnies qui les collectent et les traitent, et encore plus de savoir si cette utilisation est conforme ou non à la législation. Ceci est encore plus vrai à l'ère du Big Data où des bases de données indépendantes contenant peu d'informations personnelles peuvent être combinées facilement pour identifier une personne. D'autre part, un individu est difficilement capable d'évaluer techniquement le niveau de sécurité informatique dont font l'objet ses données pendant leur transmission et leur stockage. Parfois, les entreprises elles-mêmes ne sont pas toujours en mesure d'évaluer totalement la sécurité de leur système d'information : elles ignorent parfois si elles ont subi une cyberattaque. Les enjeux économiques sont de taille. Dans les travaux qui lui ont valu le prix Nobel d'économie, Akerlof a montré que des situations d'asymétrie d'information pouvaient conduire à la disparition de marchés. L'économie numérique n'échappe pas à cette théorie, et la détérioration de la confiance crée des risques qui peuvent conduire certains internautes à se « débrancher » [(3)].

On peut distinguer deux types de risques. Le premier est lié à la transaction individuelle. Il s'agit essentiellement de risques idiosyncratiques de contrepartie. Le contrat va-t-il être respecté ? Tous les termes du contrat seront-ils appliqués ? Ces risques ne sont pas toujours assurables car il est difficile de formuler toutes les éventualités du contexte du contrat, une situation que les économistes qualifient de contrat incomplet. Le deuxième type de risques est lié à l'environnement dans lequel le contrat s'exécute. Ces risques sont de nature systémique. Existe-t-il un recours en cas de problème ? Il est évident que la transaction s'inscrit dans un cadre juridique. Néanmoins les bénéfices d'une procédure légale ne couvrent souvent pas les coûts de transaction pour de faibles montants. Illustrons ces différents points dans le contexte d'une transaction eBay. Un acheteur commande un produit. Sera-t-il livré à temps ? Le produit correspond-il bien à celui commandé ? Il existe bien un risque idiosyncratique lié à la transaction. Il est atténué par un environnement de confiance dans lequel se déroule la transaction. En effet, le risque systémique peut être réduit à travers deux mécanismes. Le premier est un mécanisme punitif : l'acheteur trompé et déçu peut donner une mauvaise évaluation au vendeur. Le deuxième mécanisme est lié au tiers de confiance Paypal qui permet d'entamer une procédure de litige qui peut aboutir à un remboursement.

(3) Étude sur les données personnelles, chaire Valeurs et politiques des informations personnelles, 2017, https://cvpip.wp.imt.fr/donnees-personnelles-et-confiance-quelles-strategies-pour-les-citoyens-consommateurs-en-2017/

Enjeux économiques des problèmes d'éthique soulevés par les stratégies numériques

En plus des externalités négatives évoquées précédemment, l'utilisation des données personnelles dans les stratégies commerciales des acteurs du numérique soulève des considérations éthiques qui ont des conséquences économiques de premier ordre. Nous en analysons trois : le libre arbitre, ou possibilité de choisir dans un environnement informationnel neutre ; l'autonomie, ou capacité à se conformer à ce qui est bon pour soi ; la discrimination, ou absence de biais systématique.

Libre arbitre et autonomie

Pour pouvoir prendre une décision économique optimale, les consommateurs doivent pouvoir choisir dans un environnement informationnel neutre. Deux phénomènes, exacerbés par les stratégies de l'économie numérique évoquées précédemment, posent problème.

Premièrement, les consommateurs reçoivent des informations filtrées par les plateformes telles que Google ou Facebook. Par exemple, le moteur de recherche Google filtre les résultats de recherche en fonction de la géolocalisation, de l'historique de navigation et du profil publicitaire. Facebook filtre les informations en fonction des « likes » et du profil de l'utilisateur. Ces filtres d'information peuvent influencer le comportement des internautes. Ils soulèvent des problèmes économiques importants liés principalement à la construction des préférences. En effet, les bulles informationnelles sont créées par des algorithmes qui génèrent un univers spécifique à un internaute, univers qui peut potentiellement influencer la manière dont il pense, se comporte et achète.

Deuxièmement, Amazon, Netflix et Spotify, parmi beaucoup d'autres entreprises du numérique, exécutent des algorithmes pour fournir des recommandations de produits personnalisés en fonction de l'historique de navigation et des achats d'une personne.

Si l'environnement informationnel peut être manipulé, les agents économiques ne prennent plus les décisions optimales. Les questions du libre arbitre et de l'autonomie se traduisent par des choix économiques qui peuvent être manipulés par des filtres informationnels et des recommandations ciblées.

Discrimination et biais

La collecte de données personnelles permet également de pratiquer des stratégies de discrimination par les prix, à savoir vendre le même produit ou service à différents prix nets à des consommateurs différents. Le prix net comprend les frais de livraison et de production. Les entreprises cherchent le meilleur prix auquel elles peuvent vendre leurs produits et services en fonction de la disponibilité à payer des clients potentiels. Pour les produits numériques, la forme de discrimination la plus répandue consiste à développer des stratégies pour identifier plusieurs groupes de consommateurs et proposer différentes versions d'un même produit ou service à ces groupes. Par exemple, un fabricant de logiciels propose un même produit avec différentes fonctionnalités : une version professionnelle complète et une version basique, ou étudiante, pour laquelle certaines fonctions ne sont pas disponibles. Les informations personnelles des consommateurs peuvent donc être utilisées pour personnaliser les offres à des clients ciblés, souvent à un coût très faible. Certains consommateurs bénéficient de prix bas, mais d'autres se voient proposer le produit à des prix plus élevés et peuvent décider de protéger leurs données personnelles pour éviter d'être discriminés.

Conséquences économiques

La manipulation de l'environnement informationnel et le ciblage posent des questions éthiques qui ont de réelles conséquences économiques, en particulier du point de vue de la concurrence et de l'innovation sur les marchés. Nous considérons par la suite cinq points préoccupants.

Premièrement, les algorithmes qui influencent les choix des consommateurs par l'imposition de valeurs externes modifient la manière dont les consommateurs construisent leur fonction d'utilité. On peut penser à Facebook qui censure des œuvres d'art jugées choquantes (*L'Origine du monde* par exemple) ou qui nourrit des algorithmes avec des données collectées sur des citoyens américains pour les appliquer à leurs utilisateurs à travers le monde. Les problématiques de la confiance deviennent alors des enjeux culturels. Deuxièmement, en l'absence de concurrence forte, les plateformes qui contrôlent l'accès des utilisateurs aux données peuvent développer des stratégies de forclusion ou refuser de faire commerce avec des entreprises tierces. Il existe de nombreux exemples de cette pratique. En juin 2012, Facebook imposait comme adresse de messagerie par défaut l'adresse du domaine @facebook.com. Toujours en juin 2012, Apple annonçait qu'il installait par défaut son logiciel de cartographie Plans à la place du logiciel Google Maps dans son nouveau système d'exploitation iOS. En avril 2013, Apple retirait de son App Store l'application AppGratis qui faisait chaque jour la promotion d'une application devenue temporairement gratuite sur le Store. En 2001, les États-Unis demandaient à Microsoft de supprimer l'offre groupée du système d'exploitation Windows et du navigateur Internet Explorer pour laisser au consommateur la possibilité d'installer le navigateur de son choix, et d'ouvrir l'interface de programmation à des compagnies tierces. En juin 2017, la Commission européenne inflige une amende de 2,4 milliards d'euros à Google pour avoir filtré des résultats de recherche en mettant en avant Google Shopping tout en défavorisant les concurrents. Troisièmement, qui garantira que les outils de protection de la vie privée seront disponibles pour les consommateurs ? Ces solutions techniques sont coûteuses à développer et vont à l'encontre des stratégies du numérique et de surveillance des États. De manière générale, la sécurité informatique est un bien public qui profite à tout le monde. Il existe donc un risque de sous-investissement des entreprises en protection des données personnelles qui peut conduire à davantage de fuites de données et de perte de confiance [(4).] Quatrièmement, les plateformes qui contrôlent l'accès aux données des internautes augmentent les coûts d'entrée sur le marché et peuvent freiner l'innovation qui nécessite des données personnelles. Cinquièmement, les algorithmes qui personnalisent les prix en fonction des consommateurs peuvent être détournés pour faciliter la collusion entre entreprises [(5)]. À titre d'illustration, on peut citer l'exemple du livre qui coûtait 23 millions de dollars, à la suite d'une surenchère algorithmique entre deux vendeurs désireux d'optimiser leur profit sur la plateforme Amazon Marketplace [(6)].

Conclusions et pistes de réflexion

La confiance repose sur l'évaluation des risques encourus lors d'une transaction effectuée dans l'économie numérique. Nous avons classé ces risques en deux catégories : idiosyncratiques et systémiques. Il est donc nécessaire pour développer la confiance dans l'économie numérique d'agir sur ces deux leviers. Premièrement, il faut renforcer la connaissance des internautes sur la manière dont les données sont utilisées et sur les externalités négatives. Deuxièmement, il s'agit de rétablir un sentiment de réciprocité et d'équité en garantissant un échange de valeur équitable (l'échange d'un service gratuit contre des données personnelles ne semble plus satisfaisant) et un mécanisme punitif crédible en cas d'utilisation abusive des données personnelles. La sanction renforcée du Règlement général sur la Protection des Données (RGPD) entrant en vigueur en mai 2018 va dans ce sens.

(4) Lire Dubus et Waelbroeck (2018) à ce sujet.
(5) Lire le rapport de l'OCDE (2017).
(6) Lire http://edition.cnn.com/2011/TECH/web/04/25/amazon.price.algorithm/index.html
Il n'y avait que deux vendeurs, des robots, dont l'un possédait le livre à vendre et voulait se situer juste au-dessous du prix demandé par l'autre, qui, lui, ne possédait pas le livre et demandait un prix sensiblement supérieur à celui du premier dans l'intention de lui acheter le livre et de réaliser un beau profit à la revente.

L'appropriation par les consommateurs des nouveaux outils de protection forme un autre moyen de construire la confiance. Des logiciels permettant de masquer les adresses IP, les extensions de navigateur Internet bloquant les scripts et les publicités rendent plus difficile l'identification des utilisateurs et la discrimination par les prix. Nous avons montré dans une enquête réalisée en 2017 sur un échantillon représentatif de la population française que ceux qui se protègent le plus sont ceux également qui achètent le plus sur Internet [7]. Ce résultat peut sembler paradoxal de premier abord, mais devient logique lorsqu'on comprend que ces outils permettent aux consommateurs d'acheter en toute confiance. On ne peut plus considérer les consommateurs comme passifs face aux stratégies des acteurs du numérique.

La question de la confiance dans l'utilisation des données porte également sur la valeur économique de l'anonymat. Une théorie simpliste postule qu'il existe un arbitrage entre valeur économique d'une part et protection de la vie privée et anonymat d'autre part. Il y aurait ainsi deux situations extrêmes : une situation où une personne est parfaitement identifiée et susceptible de recevoir des offres ciblées et une autre situation où la personne serait anonyme. Dans le premier cas, la valeur économique serait maximale dans le second ; les données n'auraient pas de valeur. Si l'on déplace le curseur vers le ciblage, on augmente la valeur économique au détriment de la protection de la vie privée. Inversement, si l'on protège la vie privée, on réduit la valeur économique des données. Cette théorie ignore les enjeux de confiance dans l'utilisation des données. Pour développer une relation client sur le long terme, on doit raisonner en termes de risques et d'externalités pour le client. Il existe donc une valeur économique à la protection de la vie privée, qui tourne autour de la relation de long terme et de la notion de confiance, et de la garantie du libre arbitre, de l'autonomie et de l'absence de discrimination. Les principes de pseudonymisation et de consentement explicite du RGPD vont dans ce sens.

Bibliographie

DUBUS A. et WAELBROECK P. (2018), «La notion de confiance en économie,» in : *Signes de confiance – L'impact des labels sur la gestion des données personnelles*, Levallois-Barth (Ed.), https://cvpip.wp.imt.fr/2018/01/30/8-03-2018-signes-de-confiance-limpact-des-labels-sur-la-gestion-des-donnees-personnelles/

MANTELERO A. (2013). "Competitive value of data protection: the impact of data protection regulation on online behavior", *International Data Privacy Law* 3(4): 229-238. DOI : 10.1093/idpl/ipt016

OCDE (2017), *Algorithms and Collusion: Competition Policy in the Digital Age*, www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm

(7) Étude sur les données personnelles, chaire Valeurs et politiques des informations personnelles, 2017, https://cvpip.wp.imt.fr/donnees-personnelles-et-confiance-quelles-strategies-pour-les-citoyens-consommateurs-en-2017/