



Pierre-Michel Menger et Simon Paye (dir.)

## Big data et traçabilité numérique Les sciences sociales face à la quantification massive des individus

Collège de France

---

# Introduction

Pierre-Michel Menger

---

Éditeur : Collège de France  
Lieu d'édition : Paris  
Année d'édition : 2017  
Date de mise en ligne : 23 octobre 2017  
Collection : Conférences  
ISBN électronique : 9782722604674

### Édition imprimée

Date de publication : 24 octobre 2017

Ce document vous est offert par Collège de France



<http://books.openedition.org>



### Référence électronique

MENGER, Pierre-Michel. *Introduction* In : *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus* [en ligne]. Paris : Collège de France, 2017 (généré le 24 octobre 2017). Disponible sur Internet : <<http://books.openedition.org/cdf/4990>>. ISBN : 9782722604674.

---

# Introduction

**Pierre-Michel Menger**

Professeur au Collège de France, titulaire de la chaire « Sociologie du travail créateur »

**L**A PRODUCTION EN MASSE de données numériques a rapidement ouvert la possibilité d'exploiter des informations en quantités inédites<sup>1</sup>. Un nombre croissant d'activités humaines, des plus privées aux plus publiques et relationnelles, peuvent désormais être analysées avec des moyens nouveaux dont la puissance d'exploration augmente très rapidement. Les traces numériques laissées par les usagers des moteurs de recherche, des réseaux sociaux ou des sites d'achat en ligne engendrent une masse d'informations sans précédent, non seulement sur les usages d'Internet en général, mais aussi sur les centres d'intérêt des individus, leurs pratiques de consommation et leurs orientations politiques ou religieuses. Au-delà d'Internet, une production comparable de données numériques de masse est assurée par un nombre croissant d'objets connectés qui peuvent être vus comme autant de capteurs des activités humaines (systèmes de géolocalisation, relevés de consommation d'électricité, systèmes de mobilité urbaine, cartes bancaires, passeports biométriques, etc.).

Depuis quelques années, l'expression « *big data* » (qu'on peut traduire par « données numériques massives ») sert à désigner ce phénomène. Elle tend à envelopper dans un même ensemble toutes les données numériques issues de l'activité des individus, qu'il s'agisse d'actes privés de consommation ou d'interventions dans les réseaux sociaux, ou de la production d'informations, de documents, d'archives dans les entreprises et les administrations, ou encore de la numérisation de tout le stock des données archivées par ces mêmes firmes, administrations, bibliothèques, et sources d'information.

Une nouvelle économie se structure autour d'un flux rapide d'innovations et d'applications destinées à stocker, à analyser, à visualiser et à échanger ces données, *via* des transactions marchandes ou hors marché. La recherche scientifique fondamentale et appliquée, l'innovation industrielle, le commerce, le marketing, l'expérimentation sociale, les politiques publiques sont invitées à se soucier de l'extraction et de l'exploitation de ce nouveau gisement informationnel. Les échelles d'analyse sont spectaculairement emboîtables depuis les cartographies planétaires d'échange et de diffusion d'informations jusqu'à la microsegmentation des comportements et des préférences des individus. Les unités d'analyse – la personne, l'organisation, le réseau, le groupe, la masse – sont aussi rendues beaucoup plus directement commensurables.

Les outils d'exploration de ces données bousculent non seulement l'analyse des comportements, mais encore les techniques d'influence et d'incitation. La détection à grande échelle de corrélations, la recherche automatisée de patterns, les algorithmes d'analyse sémantique prennent le pas sur les patients travaux d'enquête par échantillonnage et rivalisent avec

---

1. Je remercie chaleureusement Simon Paye pour ses remarques sur cette introduction et pour l'ensemble du travail accompli pour organiser le colloque qui s'est tenu au Collège de France et la production de cet ouvrage, qui en est issu.

l'expérimentation contrôlée. Profiler pour prédire les tendances et les usages, pour persuader les usagers et les consommateurs, et, ce faisant, pour obtenir que les anticipations issues des profilages se vérifient, voilà qui introduit de multiples boucles dans l'interaction avec les systèmes, boucles qui conduisent à ce dont le marketing a toujours rêvé : inventer une technologie silencieuse de la prophétie ou de la persuasion auto-réalisatrice, qui se substitue au bruit de la persuasion publicitaire traditionnelle.

Les entreprises numériques se multiplient et viennent nourrir les hypothèses de la recherche en sciences sociales. Lorsque les machines et les algorithmes apprennent progressivement à se passer d'instructions pour trouver, décider, exécuter, la création des métiers appariés à ces innovations obéit à la dynamique de concurrence des professions : de nouveaux savoirs et de nouvelles compétences, de nouveaux outils et de nouveaux services déplacent et redistribuent les aires d'expertise tout autant que la hiérarchie des problèmes, des méthodes et des chances de gain. Dans la production du savoir, une nouvelle génération de relations émerge entre la recherche fondamentale et la recherche appliquée, comme entre le monde académique et le monde industriel. Stanford et la Californie attirent un nombre croissant d'universitaires et de chercheurs. Cette agglomération de talents est multiculturelle, et ses comportements sont désinhibés : l'*academic entrepreneurialism* faisait parfois l'objet d'âpres controverses à la fin du xx<sup>e</sup> siècle (Slaughter et Leslie, 1997 ; McSherry, 2001 ; Bok, 2003 ; Popp Berman, 2012), mais les réticences à son égard sont aujourd'hui levées face à l'effervescence de l'innovation, à la densité des échanges disciplinaires et à l'affaiblissement des hiérarchies de légitimité qui plaçaient la science fondamentale très au-dessus de la science appliquée, et les disciplines nobles, productrices de théories et de modèles, très au-dessus des disciplines ancillaires, immergées dans l'exploration des données et dans l'amélioration incessante des technologies. Ces chercheurs veulent tirer parti de l'agglomération des entreprises de haute technologie, de la capitalisation boursière considérable des champions de l'économie numérique, mais aussi d'un éthos de la démesure, de l'audace sans limites, qui s'établit tranquillement comme une nouvelle norme. La finitude de la vie ? Parions de pouvoir dessiner sans trop tarder l'asymptote de la vie éternelle. Le défi climatique ? Inventons la ville verte, qui s'autocontrôle, et la fluidité du trafic routier, que promet la voiture intelligente et autonome. Le déplacement dans l'espace ? Un peuple d'automates deviennent agents livreurs (les drones) ou chauffeurs (encore les voitures intelligentes), et le monopole des agences spatiales publiques s'achève.

De nombreux travaux sont consacrés depuis près de vingt ans à la polarisation du marché du travail et à l'affaîssement du continent des emplois de qualification moyenne. Les relations de complémentarité ou de substitution entre les emplois et les technologies numériques expliquent pourquoi les tâches à forte intensité cognitive et créative (Autor, 2013 ; Autor et Handel, 2013 ; Brynjolfsson et McAfee, 2014) et à fort pouvoir relationnel (Deming, 2015) protègent les emplois supérieurs peu routiniers de la menace de robotisation à brève échéance, et pourquoi les tâches routinières, cognitives comme manuelles, sont exposées directement à la substitution du capital (robot, machines intelligentes, algorithmes) au travail. L'éthos de la créativité et de l'inventivité transgressives, qui suscite tant de fascination et de sidération, incarne en quelque sorte la nouvelle norme comportementale des humains non robotisables.

De même que le « principe des 4 V » (volume, variété, vitesse, véridité) est la signature de la séduction exercée par les données numériques de masse, selon Serge Abiteboul (2012), la croissance est spectaculaire dans le volume des échanges, des expérimentations, des innovations, mais aussi des débats, interrogations et critiques qui touchent aux dimensions tout à la fois technologiques, industrielles, entrepreneuriales, sociales et éthiques des transformations en cours. Les marchés, la recherche-développement, les sociétés savantes, les législateurs, les organisations publiques internationales, les communautés de recherche, les associations de consommateurs, les médias ont, chacun depuis sa perspective, identifié un ensemble de questionnements communs, dont le pouvoir de rayonnement et de contagion illustre à lui seul la puissance des mécanismes de production et d'échange des contenus informationnels (Mayer-Schönberger et Cukier, 2013). Les débats ainsi lancés sont d'autant plus intenses que le maniement des données massives 1) s'applique à tous les aspects de la vie sociale et économique, 2) bouscule tous les partages habituels entre privé et public, entre intime et relationnel, entre secret et révélation, entre consentement passif et consentement explicite, entre usage et confiance, et 3) fonde son expansion sur la nature intrinsèquement relationnelle des données, qui décloisonne et fait communiquer toutes les dimensions et tous les domaines d'activité générateurs d'informations.

De toutes les propriétés mises en avant par les acteurs promoteurs des « *big data* », le volume et la croissance exponentielle de ces masses paraissent être les plus stupéfiantes. Pourtant, nous sommes devenus familiers des grandeurs extrêmes : pour connaître les bornes de notre monde, nous recourons, par exemple, à des ordres de grandeur dont l'échelle varie de  $5 \times 10^{-324}$  (en informatique, c'est la plus petite valeur positive différente de zéro qui peut être représentée par une valeur à virgule flottante à double précision) à  $10^{120}$  (quantité de toutes les différentes parties d'échecs qu'il est possible de jouer, dit « nombre de Shannon »). S'agissant des « *big data* », c'est en réalité la capacité de mise en relation ou de connectivité de ces données qui est le phénomène le plus remarquable. N'importe quelle parcelle de réalité peut être considérée comme une somme d'informations, informations dont la valeur vient de leur mise en relation. Et n'importe quelle relation observable et traçable entre un individu et un élément de son environnement peut être génératrice de données enregistrables, stockables, communicables et échangeables.

Le mouvement identifié sous le vocable « Big Data » est double. Il fait voir à la fois comment augmente le nombre d'éléments d'un environnement générateurs d'informations relationnelles traçables (le thème des objets connectés / capteurs), et comment se transforment les comportements des individus (humains et non humains) qui sont de plus en plus massivement émetteurs et récepteurs d'informations traçables. Les sciences sociales nous l'ont appris : quand on agit, on interagit. Et l'analyse économique nous a appris à graduer les interactions en fonction des informations dont disposent les acteurs pour agir, et en fonction de la symétrie ou de l'asymétrie informationnelles des relations d'interaction et d'échange. Or le puissant paradoxe qui est inhérent au développement des outils et des plates-formes numériques réside dans la construction et dans l'exploitation d'une asymétrie à laquelle les usagers et les consommateurs consentent : les outils sont relationnels, mais, sans l'asymétrie qu'ils exploitent, ils perdent leur efficacité. Cette asymétrie est invisible : les usagers ne savent que bien peu de choses des plates-formes, des algorithmes et de toute l'économie numérique

qui transforme en valeur les innombrables données personnelles recueillies sans cesse sur eux, au point que l'on a pu conceptualiser cette production de données captables comme un travail «bénévole à but lucratif» (Dujarier, 2016), fourni par les usagers des objets connectés. De multiples controverses et des demandes croissantes de régulation cherchent à rétablir la frontière entre ce qui appartient à l'individu et ce qu'il consent à laisser observer de lui. Mais le jeu de l'innovation déplace sans cesse ce pouvoir de l'asymétrie. L'économie numérique, incarnation d'une nouvelle révolution industrielle, représente aussi une concentration de l'expertise et de la créativité qui donne un pouvoir économique et technique intimidant à une jeune génération d'entrepreneurs, d'inventeurs, d'ingénieurs et d'informaticiens.

Au regard de cette émergence, quels gains peut-on attendre de cette nouvelle forme, graduable, de souveraineté des individus? L'accès à des informations auparavant introuvables ou inexistantes, les facilitations marchandes, la sociabilité démultipliée, les innovations attachées aux externalités de réseau? Et quels en sont les risques? Trop de publicité, trop de profilage, trop de prédiction segmentatrice enfermant les consommateurs dans un espace réduit de choix, trop d'avantages cumulatifs pour certains biens ou services, laissant une multitude d'autres biens et services dans un bruit informationnel indistinct? Quelles sont les résistances, les controverses et les révoltes? La réémergence des débats sur la protection de la vie privée, la définition de biens communs inappropriables, les nouveaux horizons politiques du piratage informatique? Les services secrets disposent de leviers extraordinaires avec l'exploitation des données de masse, et jouissent d'un monopole légal dans l'exploitation de l'asymétrie d'information, autrement dit d'une nouvelle forme de monopole de la violence légitime, pour reprendre la définition donnée par Max Weber de l'État et de son usage légal de la force. Admettra-t-on alors que les entreprises et les organisations qui peuvent prélever, gérer et revendre des données personnelles bénéficient d'une asymétrie d'information légitime, et contractualisable?

Finalement, cinq évolutions se superposent désormais :

- les innovations scientifiques et techniques;
- la multiplication des sources génératrices de données et leur connectivité;
- l'expansion du marché des outils et des usages d'exploitation des données de masse;
- la mise en controverse éthique et politique, et la mise en régulation juridique et législative, des déséquilibres informationnels exploités;
- la formation de nouveaux marchés de services transformant la *privacy*, la vie privée, en utilité ou «désutilité» graduable et négociable.

\*\*\*

Le 2 juin 2014 s'est tenu au Collège de France le colloque intitulé «*Big data*, entreprises et sciences sociales. Usages et partages des données numériques de masse», dont Simon Paye et moi-même avons pris l'initiative. On ne compte plus les publications sur le caractère «révolutionnaire» du phénomène des *big data*<sup>2</sup> dans le monde de la recherche. Avec un ensemble de collègues dont les contributions constituent la matière de ce livre, nous avons voulu explorer

---

2. Citons ainsi, parmi les travaux examinant l'impact de cette révolution des données sur la recherche en sciences sociales, King (2009, 2011), Lazer *et al.* (2009), Bail (2013), National Research Council (2013), Ruppert, Law et Savage (2013), Einav et Levin (2014), Kitchin (2014), Wyly (2014), Venturini *et al.* (2014).

les positions et les défis des sciences sociales face à la prolifération des traces numériques et de leur marchandisation. Il est par exemple annoncé que les données numériques de masse ouvriront la voie à une science sans théorie : plus besoin de s'embarrasser d'hypothèses et de références scientifiques si les données, par leur incroyable richesse, permettent de suivre en temps réel l'évolution des comportements, à l'instar d'une sorte de météorologie du social. De même, les acteurs du marketing, des assurances ou du conseil en organisation promettent l'avènement d'une nouvelle ère dans laquelle l'entreprise devenue intelligente peut prendre des décisions fondées sur une quasi-omniscience de ses dynamiques internes et de son environnement. Pour ceux qui tiennent ces discours, les traces numériques des individus connectés ne sont rien de moins qu'un nouvel Eldorado informationnel.

L'ouvrage examine, dans une première partie, les usages des données numériques et les logiques qui les sous-tendent, qu'elles soient marchandes, scientifiques, policières ou communautaires. Puis les pratiques concrètes et les opérations au quotidien sont analysées dans une deuxième partie, à partir d'enquêtes de terrain localisées. Les contributions de cette partie explorent en détail les comportements et les modèles émergents, tout en menant une réflexion parallèle sur les enjeux sociaux, éthiques, politiques auxquels ils donnent lieu. Une troisième partie est consacrée à l'utilisation des données numériques dans le cadre de la recherche en sciences humaines et sociales. L'enjeu est de clarifier, à partir de cas concrets, les perspectives que ces nouvelles sources de données ouvrent à la sociologie, à l'économie, aux sciences humaines, sans négliger les problèmes pratiques d'accès, de stockage ou d'interprétation qu'elles soulèvent. Enfin, une postface interroge la place du travail humain dans l'industrie du numérique.

La brève présentation des chapitres qui suit permettra de faire apparaître la triple ambition de cet ouvrage collectif : 1) donner à comprendre le rôle de l'asymétrie des relations utilisateurs / plates-formes dans la captation, l'exploitation et la marchandisation des traces numériques ; 2) rassembler des connaissances empiriques nouvelles sur la genèse et le développement de marchés liés à l'exploitation des données numériques, en faisant apparaître les différents intermédiaires qui s'insèrent dans les chaînes de valeur ; 3) offrir une synthèse de nos réflexions sur les enjeux cognitifs, épistémologiques et méthodologiques des usages de données numériques de masse en sciences sociales.

Frank Cochoy et Jean-Sébastien Vayre proposent une généalogie des relations d'échange qui sont au fondement de la production et de l'utilisation des données massives. Sur un marché opèrent des entreprises et des consommateurs. Ce marché est lui-même la somme des relations d'échange que les entreprises entretiennent avec d'autres entreprises, et des relations que les entreprises nouent avec les consommateurs individuels. Dans tous les cas, la vente de biens et de services fait appel à des dispositifs destinés à faire connaître les produits, à les valoriser, à persuader les acheteurs de les adopter sans en avoir fait l'expérience préalable, et à les convaincre de devenir fidèles, à l'aide d'arguments visant à soutenir une réputation de qualité, pour un rapport donné entre prix et caractéristiques du bien. Chaque concurrent agit de même. Le marketing et ses analyses des caractéristiques des consommateurs, de leurs besoins et de leurs préférences, la publicité, les techniques de fidélisation de la clientèle sont autant de dispositifs générateurs de données à prélever, à exploiter et à faire remonter vers le producteur. Les dispositifs numériques de prélèvement

massif de données personnelles, de traitement et d'exploitation marchande de celles-ci ne sont qu'une extension de ces dispositifs, selon Cochoy et Vayre. Les auteurs détaillent ce traçage généalogique en insistant notamment sur les savoirs mobilisés et sur les techniques mises en œuvre pour transformer l'individu consommateur ou usager (dans le cas des services publics) en grandeur statistique et pour identifier, grâce au calcul probabiliste, les comportements prévisibles, les préférences influençables, les demandes à satisfaire et les innovations profitables. Avec les techniques numériques, les individus ne sont plus simplement des représentants de classes statistiques issues de patients travaux économétriques de classification et de segmentation, ils sont désormais identifiés directement, par leurs caractéristiques personnelles, par leurs comportements et leurs trajectoires de consommation, par leurs préférences et par leurs relations. Ils sont traçables, visualisables et susceptibles d'être individuellement influencés. Qui donc traite toutes ces données pour alimenter tous les dispositifs d'action sur le consommateur ? Une armée d'ingénieurs et de statisticiens ? Non, des machines programmées pour se rendre autonomes et apprendre à travailler seules. Les discours sur les innovations sont incessants, mais l'efficacité de ces dernières est plus incertaine que ne le promettent les multiples sociétés de conseil et les fournisseurs de services numériques qui enjoignent aux entreprises de ne pas manquer le virage des *big data*. En replaçant les pratiques de marchandisation des données numériques dans le cadre d'un « marché des dispositifs marchands », les auteurs se donnent les moyens d'interroger à nouveaux frais la performativité de ces promesses.

Par le titre de sa contribution, Bernard Harcourt suggère d'emblée une parenté théorique avec l'œuvre de Michel Foucault. Qui détient un savoir détient un pouvoir. Le savoir produit à partir des données massives est traité, exploité et échangé dans des conditions, et pour des usages, dont chaque personne privée n'a au mieux qu'une infime idée. L'économie biface des plates-formes repose, comme on le sait, d'un côté, sur la mise à disposition d'un service gratuit ou très peu coûteux pour le consommateur, et, de l'autre, sur le consentement de ce consommateur à abandonner tout contrôle sur les informations que produit son utilisation du service. Mais cet échange n'est que la pointe d'un iceberg d'opérations et d'échanges dont le volume et la variété ont très rapidement et considérablement augmenté. Pour les connaître, le journalisme d'investigation, l'expertise citoyenne, le travail scientifique militant et l'enquête historique offrent leurs solutions et leurs lots de révélations, comme le montre Harcourt au début de sa contribution. Mais, objectera-t-on, les savoirs produits à partir des données individuelles ne sont pas intrinsèquement condamnables, puisqu'ils peuvent servir le bien commun. C'est l'argument qui soutient l'édifice entier des différentes sciences impliquées dans la connaissance des individus, de leur santé, de leurs comportements et de leurs préférences. La mise en place de normes procédurales et de garanties éthiques voudrait permettre de dépasser les difficultés liées à l'asymétrie entre l'expertise du scientifique et le défaut d'expertise du profane. Et la connaissance qui sera issue de l'exploitation de ces sources numériques ne peut devenir un bien public que si sa production est soumise au libre examen de ses méthodes, de ses résultats et de ses usages possibles, et si son utilisation est non rivale et non exclusive.

L'enquête, nous montre Harcourt, doit se déployer sur les deux versants : qui agit et pour quel usage ? Une fois fixées les équations de base de la science et de l'industrie des données (quantités produites, quantités à stocker, catégories d'opérateurs, capacités tech-

nologiques de traitement), une économie politique des données doit différencier les types d'informations recueillies et traitées, en fonction des types d'entités émettrices de données – les humains, les machines, les entreprises, les objets, les administrations – et de la nature des relations qu'entretiennent ces entités. Au plus loin de ce que peut être, à l'ère numérique, la production de biens publics de connaissance, et au plus loin de l'alignement que celle-ci suppose entre les rationalités économique, politique et démocratique, Harcourt montre comment la connaissance extraite des données est traitée comme un bien marchand et une ressource de surveillance sécuritaire. Le marché développe ses chaînes d'intermédiation, au long desquelles les données personnelles deviennent des biens de transaction ordinaires, et les caractéristiques des individus des variables ordinaires de tarification différentielle, alors qu'elles peuvent concerner les aspects les plus intimes et les plus vulnérables de la personne. Comme le montrent aussi Cochoy et Vayre, le marketing dispose ainsi d'instruments sans précédent d'analyse des comportements de consommation. Mais où s'arrête le marketing et où commence le contrôle exercé sur le consommateur ? Le rapprochement avec la surveillance d'État pratiquée par la *National Security Agency* (l'Agence nationale de sécurité américaine) et son programme PRISM souligne que les mêmes sources de données et les mêmes techniques d'analyse sont au service de pratiques de profilage et de probabilités prédictives, dont les buts sont différents, mais les rationalités semblables. Le défi démocratique que représente la régulation d'une économie politique des données aussi puissante et aussi aisément nourrie d'innovations incessantes est réellement considérable.

Guillaume Tiffon se demande comment qualifier l'activité des internautes dès lors que les traces informationnelles des comportements de ceux-ci ont pour les entreprises et leurs plates-formes une valeur directement capitalisable et exploitable. L'internaute n'est-il que le nouvel avatar du client mis à contribution de multiples manières pour interagir avec des machines qui se sont substituées à des personnels ? Avant de répondre, il faut d'abord identifier les diverses modalités de ce que peut être la « mise au travail du client ». Se fondant sur un ensemble d'enquêtes de terrain qu'il a réalisées, Tiffon détaille les diverses contributions qui sont apportées par les clients, sans être contractualisées ou monétarisées, pour accroître l'efficacité productive, dans la grande distribution, dans les centres d'appel, dans la restauration rapide ou dans les cabinets de kinésithérapie. Les multiples formes de *self-service* et d'interaction avec des automates (caisses, distributeurs, bornes d'enregistrement, etc.) en sont les exemples les plus familiers, mais ce ne sont pas les seuls. Ces contributions sont bien un travail qui demande au consommateur du temps et des compétences pour se rendre complémentaire des machines ou pour gérer des interactions dans des relations de service. Tiffon conclut que, parmi les définitions usuelles du travail, y compris celle qu'avait Marx de la création de valeur par extorsion de la plus-value, aucune ne semble s'appliquer à l'activité de l'internaute. Ce n'est pas le Marx de la théorie de la valeur-travail qui aide à penser la position de l'internaute devant un écran dont l'envers est un prédateur d'informations, mais plutôt le Marx contempteur de la consommation, et de cet autre type d'aliénation qu'est la manipulation des besoins individuels par le marché – et plus encore Foucault, mis en avant par Harcourt, pour ses analyses sur la mise sous surveillance des individus par des dispositifs de contrôle et de pouvoir auxquels ils sont, au mieux, invités à consentir en grande, ou même en complète, méconnaissance de cause.



Éric Dagiral et Sylvain Parasia examinent l'émergence de la science des données, son épistémologie si particulière de l'induction, et les comportements des praticiens actuels. En enquêtant auprès de *data scientists*, en revisitant la littérature académique et en analysant la conception et les évolutions d'un des logiciels favoris des praticiens, les deux auteurs montrent que les *data sciences* sont au point de convergence d'une série de transformations. En statistique, dès les années 1960, l'intérêt se tourne vers une démarche exploratoire. L'idéal de précision et de rigueur qui fait de la statistique un simple champ d'application des mathématiques conduit à sacrifier d'autres heuristiques, comme l'exploration et l'induction. Or les outils informatiques ont précisément pour vertu de permettre de manipuler rapidement de grandes quantités de données, pour y détecter des relations et des patterns, dont la signification n'est pas tirée d'un modèle hypothético-déductif, mais approchée et interprétée par tâtonnements. Le développement de l'informatique met de multiples logiciels statistiques entre les mains de non-statisticiens. La collecte, la gestion et le traitement de données, d'abord considérés comme des opérations techniques d'application, deviennent innovants : pratiqués par des informaticiens et des firmes informatiques, ils font naître des outils d'analyse qui ne dérivent pas simplement de recherches académiques. Enfin, quand, au début des années 2000, la science des données se constitue, c'est par une levée des frontières entre statistique et informatique, et entre chercheurs universitaires et chercheurs des entreprises. Stanford attire par exemple des économistes qui veulent accéder à des données d'une variété et d'une ampleur inédites, en collaborant avec des ingénieurs et des statisticiens. Dans la recherche génétique et biologique, les collaborations entre l'université et les entreprises se sont considérablement développées et sont un levier décisif pour l'innovation. La science des données s'émancipe quand une diversité grandissante d'acteurs contribuent à ses avancées. C'est ce que montre l'étude des différents apports à l'évolution du logiciel très populaire R, ou encore l'accueil réservé à la *data science* dans des milieux très divers, mais tous directement intéressés par l'exploitation de grandes quantités d'informations – journalistes, *marketers*, chercheurs en biomédecine, etc.

Quand les parties prenantes se multiplient, elles déploient des savoirs et des technologies qui n'obéissent plus à un modèle dominant de prélèvement et d'exploitation des informations individuelles. Un monde de transactions émerge. Les données peuvent être prélevées sur les individus à leur insu. Mais les données émises par chaque individu constituent aussi un bien ou un service qui pourra lui être vendu pour lui permettre de connaître et de contrôler son environnement, sa santé, sa sécurité, ses déplacements, ses relations, ses transactions, ses interactions de sociabilité : c'est le *quantified self*. Or la qualité de ce service marchand de quantification individualisée dépend de l'exploitation à grande échelle des données recueillies, qu'il s'agisse de coordonner les comportements, d'analyser les risques collectifs, d'identifier les mécanismes de contagion ou encore d'augmenter le pouvoir prédictif des connaissances accumulées. Un ensemble de techniques et d'algorithmes qui nous renseignent sur notre état de santé, sur les qualités de notre nourriture, sur les variations de notre état physique au travail et sur toute autre dimension de notre comportement augmentent la puissance des services marchands qui nous sont vendus, mais ils peuvent alimenter aussi les bases de données qui sont exploitées par la recherche publique à des fins d'intérêt général. Comment établir une distinction stable et efficace entre ce qui sera extrait de nos données

personnelles pour améliorer le bien public fondé sur la connaissance, et ce qui, à travers les multiples canaux de diffusion possibles des données, alimentera des usages commerciaux et servira des intérêts particuliers? L'individu émetteur de données personnelles est face à une asymétrie sans solution : comment peut-il concevoir tous les usages possibles de la quantité et de la diversité considérables de données qui le concernent ou l'incluent, afin de décider d'être non seulement un consommateur souverain, mais aussi un producteur souverain d'informations exploitables?

Finalement, le contraste est grand entre l'histoire longue des hybridations mêlant la statistique et l'informatique, telle qu'elle est relatée par Dagiral et Parasie, d'une part, et la mise à l'agenda récente et spectaculaire de la question des données numériques de masse dans les organisations de recherche, les entreprises, les administrations et l'arène publique, d'autre part. Les attentes sont considérables, les injonctions à la modernisation sont lancées quotidiennement, les espoirs d'avancées scientifiques sont devenus omniprésents, mais les controverses et les critiques sont aussi très vives contre la transformation des individus en matrices d'informations exploitables sans limitations ni contrôles efficaces.

David Pontille et Didier Torny procèdent à une analyse généalogique et à un état des lieux de l'outil bibliométrique, devenu l'un des socles de l'évaluation de la qualité scientifique des chercheurs et des institutions, et l'un des instruments de pilotage de la recherche et de la gestion des recrutements et des carrières. Ils distinguent ainsi trois configurations qui ont situé la science elle-même à l'avant-garde de l'exploitation de données massives, et ils montrent que l'inventivité algorithmique et métrologique est affaire de jeux d'acteurs, de variété des sources de données et de stratégies concurrentielles. On connaît l'une des fonctions traditionnellement remplies par les revues scientifiques : celles-ci se substituent, pour une large part, aux universités et aux instituts de recherche pour évaluer, sélectionner et valider la production de recherche. La scientométrie fut mise au point par Eugene Garfield, qui s'appuyait sur les travaux théoriques de Derek J. de Solla Price. Elle a consisté à mesurer la valeur des revues et la valeur des articles à leur audience (à leurs citations), et à introduire ces informations dans les outils d'évaluation des individus, de hiérarchisation des revues et de classement des universités. L'analyse de Pontille et de Torny retrace les initiatives des firmes qui ont concurrencé l'entreprise pionnière de Garfield, qu'il s'agisse d'éditeurs comme Elsevier, ou du moteur de recherche Google Scholar et des indices bibliométriques qu'il produit. Parce que le but d'une recherche est d'être aussi accessible et visible que possible, et que la traçabilité de son appropriation est un progrès irrécusable, la production scientifique engendre un monde de données idéalement renseignées, idéalement individualisables, et idéalement exploitables. L'ingénierie bibliométrique a développé, à partir de ce matériau de qualité exceptionnelle, une cascade d'innovations pour étendre son emprise de mesure et d'évaluation à la plus large production scientifique possible, et pour proposer une grande variété d'indices et de systèmes de cotation de la valeur des supports de publication. Ici aussi, la situation apparaît de part en part asymétrique. Les revues n'existeraient pas sans la matière première des travaux scientifiques dont elles publient les résultats. Mais l'accès à ces revues est souvent payant, et les bouquets d'abonnement sont coûteux. L'utilisation systématique des données bibliométriques pour mesurer la productivité ou pour cerner le potentiel des individus peut se substituer au grain fin des procédures collégiales d'évaluation

qui font appel à une pluralité de sources d'appréciation. Les revues se multiplient, et sollicitent autant qu'elles absorbent la production d'une population mondialisée de chercheurs. Les algorithmes bibliométriques se diversifient pour rapporter la valeur d'une production scientifique à d'autres critères d'audience et d'impact que sa citation dans le seul périmètre des revues. Progressivement, le jeu de la science fait face à des dilemmes radicaux : faut-il consolider le pouvoir marchand des éditeurs jusqu'à transformer l'évaluation scientifique en une somme de technologies métrologiques et indicielles qui alimentent un marché de services destinés aux chercheurs et à leurs institutions pour augmenter leur productivité et leur impact ? Ou faut-il libérer complètement l'accès aux productions scientifiques à travers des archives ouvertes, qui suspendent le pouvoir sélectif et le travail éditorial des pairs, quitte à pousser chaque scientifique à devenir le gestionnaire de sa propre visibilité et à activer à cette fin de multiples sources d'impact, d'audience, d'attention, d'usage ? Dans ce dernier cas, rechercher et mesurer l'attention des publics les plus variés revient-il à leur reconnaître une expertise profane légitime, qui complète l'expertise professionnelle des pairs, voire s'y substitue ?

Dans leur contribution, Jérôme Denis et Samuel Goëta étudient le phénomène de l'ouverture des données publiques (*open data*), contemporain de celui des données massives, et précisent les modalités de son expansion. L'ouverture des données est le foyer d'une mobilisation qui oppose directement l'accès public à l'exploitation industrielle et commerciale des informations, données et connaissances en jeu. En théorie, un individu ou une organisation producteurs ou émetteurs de données qui ont une valeur pour autrui peut soit s'en réserver l'utilisation, soit autoriser celle-ci en s'assurant du contrôle des usages qui seront faits des données, soit en ouvrant complètement l'utilisation si ces données sont considérées comme des biens publics, ni rivaux ni exclusifs. Dans tous les autres cas, quand l'émetteur des données ne peut exercer qu'un contrôle partiel, sporadique, ou même aucun contrôle du tout, sur l'usage et sur la transformation des informations qu'il a émises, le partenaire de l'échange (plate-forme, site web, moteur de recherche) tirera un avantage de la collecte des informations auprès de chaque individu. Cet avantage est certes infinitésimal, si chaque cas est considéré isolément, mais il devient potentiellement considérable s'il agrège ces informations avec celles obtenues auprès de tous les autres utilisateurs du service. Ce levier essentiel de l'exploitation à grande échelle d'une asymétrie informationnelle est un cas classique de fonctionnement imparfait des marchés : un acteur en sait davantage que son partenaire dans l'échange sur les profits qu'il peut tirer de cet avantage, alors même que les deux parties sont en relation d'échange et que l'échange devrait être transparent pour être équilibré et efficient. Le phénomène de l'Open Data ne se présente pas comme une solution à de telles défaillances, puisqu'il concerne d'abord toutes les données qui ont vocation à devenir des biens publics d'information et de connaissance, que ces données soient produites par des acteurs publics (administrations publiques ou autres), par des acteurs dont les activités sont financées par des fonds publics (chercheurs scientifiques notamment) ou par des producteurs de connaissances et d'outils techniques désireux d'en rendre l'utilisation totalement libre (créateurs du Web 2.0, militants des logiciels

libres). Pourtant, le motif de l'asymétrie informationnelle réapparaît quand il s'agit de définir les caractéristiques désirables de ces données ouvertes.

C'est assurément toute une nouvelle organisation de l'économie de la connaissance et de l'innovation qui veut imposer ses exigences : exigences de démocratisation, de transparence des opérations de production et de traitement des informations, mais aussi exigences de libre exploitation d'une matière aussi proche que possible de son état brut, afin que son traitement, qui sera générateur de valeur ajoutée, puisse préserver un lien traçable avec la source. Mais, comme le montrent Denis et Goëta, les données présentées comme brutes sont en réalité des données « brutifiées ». L'opération de « brutification » consiste à transformer des données de travail originales, construites et formatées selon des règles et des besoins internes, en données génériques susceptibles de traitements statistiques. Mais ne revient-elle pas à rétablir un filtre permettant aux administrations de conserver une marge de discrétion ? Si « brutification » équivaut à standardisation, la démocratisation de l'accès aux données conduit-elle si heureusement qu'on le dit à la diversification des sources et à la démocratisation des expertises qui peuvent s'en saisir ? Si, selon l'argument bien connu, une concurrence plus parfaite procure des gains d'efficience, le rétablissement d'un équilibre informationnel peut permettre d'améliorer le fonctionnement des marchés de biens et de services, de multiplier la production de connaissances, et de décentraliser le pouvoir d'innovation. Il doit aussi rendre les acteurs publics plus directement et plus continûment responsables de leurs actions, ce qui rapproche alors l'ouverture des données d'un dispositif de contrôle propre au principe de l'audit, dont l'extension est pourtant controversée. Enfin, il participe de l'émergence d'une *crowd science* ou science citoyenne, qui, selon les cas, sera complémentaire ou rivale de la science professionnelle. Cette dernière peut au demeurant trouver elle-même de nouveaux leviers d'innovation dans une circulation ouverte des données publiques, tout en ayant à chercher son propre équilibre entre la culture de l'accès et la protection du monopole temporaire de l'innovateur.

La sociologie est, dès sa fondation à la fin du XIX<sup>e</sup> siècle, une science des relations et des interactions autant qu'une science de la détermination des actions et des comportements des individus et des groupes. La sociologie des réseaux a connu un essor considérable à partir des années 1970 aux États-Unis. Elle pouvait fonder son développement sur l'une des inspirations originaires de la sociologie américaine, celle qui, comme l'ont souligné successivement John Dewey, George Herbert Mead, Herbert Blumer, Erving Goffman, Howard Becker, Harrison White et Andrew Abbott, fait de « l'acte social », ou de l'action immergée dans un réseau sans cesse changeant d'interactions, le fondement de ce qu'est une société (Menger, 2013). La sociologie des réseaux a formalisé et mathématisé cette intuition originare. L'invention du Web 2.0 pouvait livrer à cette sociologie relationnelle une importante matière première et des situations de quasi-expérimentation, fondées sur les ressources illimitées d'interactivité, mais aussi sur une immersion des individus dans un monde de relations entre des entités de types multiples, dont les humains ne sont qu'une catégorie. Une technologie conforme à une ontologie whiteheadienne du monde social (Menger, 2016) semble être apparue.

À partir des travaux du laboratoire Orange Labs dont il est membre, Jean-Samuel Beuscart procède à un bilan des usages scientifiques des données du Web, et détaille les avancées et les limites du déploiement de cette sociologie dans les recherches sur les usages des nouveaux

médias. Le Web paraît abolir toutes les contraintes sur la formation et le développement des relations entre toutes les entités qui peuplent la Toile, et favoriser un monde horizontal, sans hiérarchies établies *a priori*. Mais les inégalités d'attention, de visibilité et de réputation que reçoit chacun existent bel et bien et nous rappellent que le Web obéit à des principes de gravitation relationnelle : ces principes sont bien différents de ceux d'une démocratie horizontale de liberté d'accès et d'échange. Les comportements des internautes participatifs sont en effet distribués très asymétriquement : la connectivité, tout comme la réputation au sein de la communauté des participants, obéissent à une loi de puissance bien connue qui veut que beaucoup contribuent peu, et peu contribuent énormément. Beuscart rappelle aussi que la matière première des liens, des traces, des usages des sites participatifs, qui est disponible pour l'analyse sociologique, n'est pas un or brut, mais un ensemble d'actes et d'informations déjà modelés en fonction de l'administration des contacts par les sites du Web 2.0. Ni objectivité ni exhaustivité des données à extraire : la mise en garde est salutaire, et elle permet aux sociologues de se maintenir en état de vigilance épistémologique tout en innovant dans le traitement de cette matière sociale considérable des actes relationnels. C'est l'équilibre entre contrôle empirique, expérimentation méthodologique et innovation théorique qui peut permettre d'avancer dans l'étude de multiples questions classiques de la science sociale, telles que les pratiques culturelles, le jugement de goût, la constitution de l'expertise profane, la construction des réputations ou encore l'influence des consommateurs sur le comportement des marchés de biens et de services infiniment différenciés. La recherche peut ainsi montrer comment les individus échangent des jugements et des évaluations, en contournant les hiérarchies anciennes qui réservaient l'émission d'opinions expertes à des professionnels du jugement et de la critique, mais elle montrera aussi comment ces nouvelles formes d'échanges plus horizontaux peuvent enclencher des dynamiques d'affiliation des jugements dont les marchés cherchent à influencer le cours.

Que valent ces nouvelles données ? Complètent-elles l'information recueillie par les procédés ordinaires d'enquête directe sur les comportements relationnels et sur les déterminants des préférences, des jugements et des choix, ou peuvent-elles prétendre s'y substituer en réorientant l'agenda des recherches vers de nouvelles classes de faits sociaux ? À la lecture de la contribution de Beuscart et de celles des autres auteurs du volume, il me semble que quatre réponses émergent. Premièrement, les données relationnelles proposent surtout une géométrie du social, puisqu'elles renseignent le plus souvent sur des comportements et sur des interactions sans permettre de caractériser finement les acteurs au-delà des traces et des informations qu'ils produisent. En ce sens, la carte des réseaux et des dynamiques de liens entre les individus ne saurait prétendre fournir une représentation directe de la société et de ses transformations. Deuxièmement, des phénomènes sociaux dont on avait l'intuition sont brusquement révélés et documentés avec une ampleur empirique sans précédent, notamment pour ce qui touche à la tension entre la puissance d'échange horizontal entre les individus et les hiérarchies qui se constituent de manière endogène à partir des inégalités d'usage, d'influence, de visibilité, de contribution, de réputation. Troisièmement, autant que la morphologie et la dynamique des réseaux et de leurs effets sur les comportements individuels, c'est une nouvelle matière de l'action sociale, la matière informationnelle, qui s'offre à l'analyse sociologique, avec une densité et une complexité inédites. Enfin, la sociologie, comme d'autres sciences

sociales, peut construire de multiples liens de collaboration avec les sciences informatiques pour innover, et pour s'appliquer à elle-même le principe fondateur de la sociologie relationnelle : le changement et la transformation sont la donnée primitive du système d'action, et c'est l'inertie qui constitue le problème à explorer et à résoudre.

Il ne faut sans doute pas négliger le phénomène de mode derrière l'engouement pour ces nouveaux gisements de connaissances promises. Mais le diagnostic suppose de ne pas basculer pour autant dans la mode de la critique négatrice (Cardon, 2015). Qu'en est-il, au juste, des innovations pour les sciences sociales ? Les protocoles de recherche se modifient quand les matériaux s'offrent en quantités illimitées. L'heuristique de l'analyse relationnelle et de la détection de formes gagne ses lettres de noblesse : de nouvelles corrélations, des configurations de relations, des tendances et des régularités peuvent être mises au jour, visualisées et interprétées. Mais la promesse d'une innovation de grande portée pour la science sociale est-elle tenue ? Une nouvelle physique du monde social est-elle en train de naître ? Pour l'instant, on n'en discerne pas les traits, du moins tant qu'on se contente d'assimiler la révolution des données à l'analyse des traces numériques des individus. En réalité, il faut, selon Ollion et Boelaert (2015), distinguer trois sources bien différentes de données : celles produites par les usages d'Internet ; celles, certainement plus décisives pour les sciences sociales, qui proviennent des actes de travail ainsi que de la production et de l'échange d'informations dans les entreprises, dans les organisations, dans les administrations ; et celles qui proviennent de la numérisation de sources documentaires prénumériques. Comment le travail du chercheur en sciences sociales est-il affecté par l'abondance et la variété de ces sources ? Les nouveaux impératifs sont cognitifs (l'acquisition de nouvelles compétences), heuristiques (la détection de patterns, et la tolérance au flottement inductif dans l'exploration des données) et méthodologiques (l'importation de méthodes familières à d'autres disciplines). Ils sont aussi éthiques et juridiques, puisqu'il s'agit bien de protéger les données personnelles exploitées par la recherche tout en favorisant l'accès à ces ressources nouvelles. Enfin, comme le montre le foisonnement des recherches exploitant les traces numériques, la matière sociale des comportements et des relations est accessible à une variété croissante d'analystes, physiciens, informaticiens, mathématiciens, mais aussi experts en maniement des données, professionnels de l'information et journalistes. Comme le souligne Jean-Samuel Beuscart, si les individus laissent une masse de traces, celles-ci sont de faible épaisseur sociale. L'ivresse « quantophrénique » ne doit alors ni masquer l'insuffisance de nouvelles propositions théoriques, ni faire oublier l'importance cruciale des données produites dans le contexte d'une recherche, quelle que soit la méthode retenue : observation, entretiens, questionnaires, etc. (Venturini *et al.*, 2014).

Dominique Boullier s'inscrit différemment dans ce débat puisqu'il se donne pour ambition d'identifier trois générations de sciences sociales, en se fondant sur un ensemble de variables analytiques dont le tableau est présenté à la fin de son chapitre. Si la science sociale est une science probabiliste, il est tentant de dessiner des configurations dans lesquelles le principe probabiliste dépend, en premier lieu, de l'état des techniques de collecte d'information employées pour caractériser la position des individus dans une totalité sociale, en deuxième lieu, des objectifs qu'ont les acteurs intéressés à la production du savoir social, et, en troisième lieu, des méthodes et des outils de traitement et de

représentation des informations caractérisant le comportement des entités élémentaires dont on cherche à décrire et à expliquer le comportement. L'argument est ici que de la science sociale d'Émile Durkheim à celle de l'acteur-réseau de Michel Callon et Bruno Latour, une réduction d'échelle intervient progressivement. La relation de l'individu à la totalité sociale, les conditions d'équilibrage de la différenciation individualisante et de la solidarité interindividuelle, et les perturbations de celle-ci en présence de crises d'anomie sociale, étaient au cœur du système durkheimien. L'analyse des comportements cherche ensuite à devenir plus prédictive, avec la technologie des sondages et des mesures d'opinion, et explore des réalités plus mésosociologiques comme les phénomènes de diffusion et de contagion informationnelle, dans lesquels les caractéristiques des individus, mais aussi leur position dans des réseaux de relations interindividuelles et leur exposition variable aux sources d'information, sont prises en compte par les modèles. La focale se resserre encore quand une topologie relationnelle caractérise les entités agissantes à l'aide des flux d'informations qu'elles émettent, reçoivent, associent, transforment et propagent sans cesse. Ici, c'est Gabriel Tarde, le rival de Durkheim, qui, selon l'argument de Boullier, triomphe, avec sa théorie d'inspiration leibnizienne. Mais comment, dans une telle ontologie sociale de troisième ordre, loger tout ce qui est révélé par les autres contributions au volume : l'exploitation stratégique des asymétries informationnelles, les amplifications de visibilité et de réputation, les techniques de contrôle, de persuasion et d'influence, et la forte concentration du marché des technologies numériques et des algorithmes, avec la domination des GAFA (Google, Apple, Facebook, Amazon) ? Boullier se demande lui-même, à la fin de sa contribution, comment éviter de conclure trop banalement qu'un modèle du social se substitue simplement à un autre.

L'ouvrage se clôt sur l'entretien qu'a réalisé Simon Paye avec Émilien, un étudiant qui finance ses études en travaillant pour un intermédiaire à qui des firmes comme Google ou YouTube sous-traitent des lots de tâches. C'est un « travailleur de la donnée » (Bastard *et al.*, 2013). Les GAFA mettent volontiers en avant les vertus novatrices, créatives et anticonventionnelles de l'organisation du travail de leurs salariés<sup>3</sup>. Au-delà des slogans sur l'« uberisation » de l'emploi, qui ne sont que des commodités publicitaires ou demi-savantes, il faut décrire et évaluer les pratiques d'emploi et de travail qui, comme les formules « auto-entrepreneuriales » de prestation, déplacent plusieurs des bornes de la flexibilité contractuelle et statutaire. Ces formes de *self-employment* s'affranchissent de l'appareillage juridique construit pour distinguer le travail salarié du travail indépendant, et mettent en avant les souplesses de la rémunération à la tâche. Cette allocation du travail est courante dans les pays où opèrent les entreprises pour qui travaille l'interviewé. L'entretien est riche d'enseignements et permet de prolonger les résultats de recherches récentes (Scholz, 2013 ; Cardon et Casilli, 2015). Les plates-formes telle que Google et YouTube, qui donnent accès à une infinité de sites et de contenus, et qui en structurent la consultation par les classements hiérarchiques des pages, doivent en permanence

---

3. Les témoignages sur les pressions à la productivité et les techniques managériales mises en œuvre n'ont pas moins de relief. Voir l'enquête de Jodi Kantor et David Streitfeld sur les pratiques managériales d'Amazon, parue dans le *New York Times*, « Inside Amazon: Wrestling big ideas in a bruising workplace », 15 août 2015.



contrôler et évaluer les sites référencés, la qualité des contenus proposés, la conformité de ceux-ci avec des dispositions législatives et éthiques (le contrôle des sites pornographiques est une des tâches fastidieuses qu'évoque Émilien), l'efficacité des requêtes qu'il est possible d'y lancer, ou encore la qualité des services proposés par la plate-forme, comme la valeur des traductions réalisées par Google Translate. Des « modérateurs » sont embauchés pour réaliser ce travail d'évaluation, de contrôle, de surveillance, de nettoyage. Ils forment une armée de prestataires dont Émilien, l'un d'entre eux, décrit le travail de façon très suggestive.

Le travail semble pouvoir être exercé hors de toute subordination : Émilien travaille où il veut, segmente ses tâches comme il l'entend, pourvu que chaque unité de travail accompli puisse correspondre à un lot minimal de tâches déterminées, ajuste son niveau d'effort en fonction de ses autres contraintes et du temps à donner à ses études, choisit ses horaires et son rythme de travail, peut moduler son niveau de concentration. Son bilan est simple :

C'est un jeu un peu. Ça demande pas une concentration extrême. [...] J'ai peut-être moyen de faire le double horaire si j'arrive à mettre en place ce qu'il faut. Là ça sera juste parfait. [...] Déjà rien que là, je gagne plus que ma mère, alors que je fous rien et qu'elle se casse le cul, je me dis que c'est magnifique. [...] Elle est agent immobilière. Avec un deuxième compte je vais être payé 3200 balles pour 30 heures de taf par semaine, où je veux, quand je veux, comme je veux. C'est déjà mieux que 90 % des gens salariés dans le pays quoi.

Mais l'entretien révèle sous une lumière crue les conditions de cette apparente « insubordination ». La prestation de services réalisée par Émilien peut être interrompue à tout moment par l'entreprise cliente, en fonction des performances attendues et observées. Le travail d'Émilien est organisé et distribué par des algorithmes, et fait l'objet d'un contrôle précis : ses faits et gestes sont chronométrés, ce qu'il produit est régulièrement évalué. Les écarts par rapport à la performance attendue, s'ils se répètent, conduisent à la fin de la relation commerciale. Un jeu tactique s'installe, qui rétablit une partie de l'équilibre bilatéral de l'échange commercial : Émilien sait développer des routines pour moduler son effort, parce qu'il est certain de sa haute productivité dans ce travail.

C'est en réalité à une mise en abyme du travail numérique que conduit cet entretien : Émilien observe et corrige ce que font les algorithmes des plates-formes tout en travaillant pour une entreprise qui emploie des algorithmes pour distribuer, quantifier, gérer et évaluer le travail demandé à ses prestataires. Et ceux-ci, s'ils sont assez agiles, peuvent exploiter les marges de manœuvre du dispositif algorithmique. Le paradoxe qui émerge est éloquent : on fait intervenir l'humain pour corriger les imperfections des algorithmes, tout en mobilisant des techniques de captation et d'exploitation des traces numériques pour contrôler l'activité des travailleurs de la donnée.



## Références

- Abiteboul S. (2012), *Sciences des données: de la logique du premier ordre à la Toile*, Paris, Fayard / Collège de France; en ligne: Paris, Collège de France, <http://books.openedition.org/cdf/506>.
- Autor D. (2013), «The “task approach” to labor markets: An overview», *Journal for Labour Market Research*, vol. 46, n° 3, p. 185-99.
- Autor D. et Handel M. (2013), «Putting tasks to the test: Human capital, job tasks and wages», *Journal of Labor Economics*, vol. 31, n° 2, S59-S96.
- Bail C. (2014), «The cultural environment: Measuring culture with big data», *Theory and Society*, vol. 43, n° 3, p. 465-482.
- Bastard I., Cardon D., Fouetillou G., Prieur C. et Raux S. (2013), «Travail et travailleurs de la donnée», *Internetactu.net*, <http://www.internetactu.net/2013/12/13/travail-et-travailleurs-de-la-donnee/> (dernière consultation le 3 avril 2017).
- Bok D. (2003), *Universities in the Marketplace. The Commercialization of Higher Education*, Princeton, Princeton University Press.
- Brynjolfsson E. et McAfee A. (2014), *The Second Age Machine*, New York, Norton.
- Cardon D. et Casilli A. (2015), *Qu'est-ce que le Digital Labor*, Paris, INA Éditions.
- Cardon D. (2015), *À quoi rêvent les algorithmes: nos vies à l'heure des big data*, Paris, Seuil.
- Deming D. (2015), «The growing importance of social skills in the labor market», Chicago, *NBER Working Paper*, n° 21473.
- Einav L. et Levin J. (2014), «Economics in the age of big data», *Science*, vol. 346, n° 6210.
- Kitchin R. (2014), «Big Data, new epistemologies and paradigm shifts», *Big Data and Society*, avril-juin, p. 1-12.
- King G. (2009), «The changing evidence base of social science research», in King G., Schlozman K. et Nie N. (dir.), *The Future of Political Science: 100 Perspectives*, New York, Routledge, p. 91-93.
- King G. (2011), «Ensuring the data rich future of the social sciences», *Science*, vol. 331, n° 11, p. 719-21.
- Lazer D., Pentland A., Adamic L., Aral S., Barabasi A.-B., Brewer D., Christakis N., Contractor N., Fowler J., Gutmann M., Jebara T., King G., Macy M., Roy D. et Van Alstyne M. (2009), «Computational social science», *Science*, vol. 323, n° 5915, p. 721-723.
- Mayer-Schönberger V. et Cukier K. (2013), *Big data. A revolution that will transform how we live, work, and think*, New York, Houghton Mifflin.
- McSherry C. (2001), *Who Owns Academic Work? Battling for Control of Intellectual Property*, Cambridge (Mass.), Harvard University Press.
- Menger P.-M. (2013), «La dramaturgie sociale du travail. Une conception interactionniste de la stratification», in Perrenoud M. (dir.), *Les Mondes pluriels de Howard S. Becker*, Paris, La Découverte, p. 206-237.
- Menger P.-M. (2016), «Temporalité, action et interaction», in Demazière D. et Jouvenet M. (dir.), *Andrew Abbott et l'héritage de l'école de Chicago*, Paris, Éditions de l'EHESS, p. 145-170.
- National Research Council (2013), *Frontiers in Massive Data Analysis*, Washington, National Academies Press.

- Ollion E. et Boelaert J. (2015), «Au-delà des *big data*. Les sciences sociales et la multiplication des données numériques», *Sociologie*, vol. 6, n° 3, p. 295-310 (en ligne: <https://sociologie.revues.org/2613>).
- Popp Berman E. (2012), *Creating the Market University. How Academic Science Became an Economic Engine*, Princeton, Princeton University Press.
- Ruppert E., Law J. et Savage M. (dir.) (2013), «Digital devices: The social life of methods», *Theory, Culture and Society*, numéro spécial, vol. 30, n° 4.
- Scholz T. (dir.) (2013), *Digital Labor. The Internet as Playground and Factory*, New York, Routledge.
- Slaughter S. et Leslie L., *Academic Capitalism. Politics, Policies and the Entrepreneurial University*, Baltimore, The Johns Hopkins University Press, 1997.
- Venturini T., Cardon D. et Cointet J.-P. (dir.) (2014), *Méthodes digitales. Approches quali/quantitative des données numériques*, *Réseaux*, n° 188 (en ligne: <https://www.cairn.info/revue-reseaux-2014-6.htm>).
- Wily E. (2014), «The new quantitative revolution», *Dialogues in Human Geography*, vol. 4, n° 1, p. 26-38.