

Stephens-Davidowitz, S. (2017). *Everybody lies. Big data, new data and what the Internet can tell us about who we really are*. New York, NY : HarperCollins

Léandre Bouffard

Volume 39, Number 1, 2018

URI: <https://id.erudit.org/iderudit/1044853ar>

DOI: <https://doi.org/10.7202/1044853ar>

[See table of contents](#)

Publisher(s)

Revue québécoise de psychologie

ISSN

2560-6530 (digital)

[Explore this journal](#)

Cite this review

Bouffard, L. (2018). Review of [Stephens-Davidowitz, S. (2017). *Everybody lies. Big data, new data and what the Internet can tell us about who we really are*. New York, NY : HarperCollins]. *Revue québécoise de psychologie*, 39(1), 255–259. <https://doi.org/10.7202/1044853ar>

Stephens-Davidowitz, S. (2017). *Everybody lies. Big data, new data and what the Internet can tell us about who we really are*. New York, NY : HarperCollins.

Par *Everybody lies*, je veux dire que les gens mentent – aux amis, aux sondages et à eux-mêmes – pour se donner contenance (p. 239).

Voici quelques exemples de « mensonges » dévoilés par les données accumulées chez Google (p. 157) :

- Les gens disent ne pas vouloir traquer leurs amis. Les données révèlent qu'ils jugent leurs amis constamment. Ceci a permis à M. Zuckerberg (fondateur de Facebook) de valoir plus de 55 milliards \$.
- Les gens disent ne pas acheter de produits provenant de manufactures où les travailleurs sont mal payés. Les données révèlent qu'ils achètent de beaux produits à « prix raisonnables ». Ceci a permis à P. Knight (cofondateur de Nike) de valoir plus de 25 milliards \$.
- Les gens disent aimer écouter les nouvelles du matin. Les données révèlent qu'ils écoutent les récits de nains ayant des relations sexuelles avec des vedettes de la porno. Ceci a permis à W. Stern (animateur de radio) de valoir 500 millions \$.
- Les gens disent avoir intérêt à *s'informer* sur le BDSM¹. Les données révèlent qu'ils aiment plutôt lire des livres pornos. Ceci a permis à E. L. James de vendre 125 millions de copies de *Fifty shades of gray* (plusieurs millions de personnes sont allées voir le film).
- Les gens disent vouloir que les politiciens présentent leur programme. Les données révèlent qu'ils veulent des politiciens qui leur épargnent les détails, mais paraissent fermes et affirmés. Ceci a permis à D. Trump d'être élu.

Au cours d'une journée ordinaire en ce début du XXI^e siècle, les humains génèrent 2,5 millions de trillions d'octets de données (p. 15) qui peuvent être très révélatrices. Comme le déclare Stephens-Davidowitz, les recherches d'information que les gens font sur Internet sont en elles-mêmes une source d'information sur, par exemple, le sexe, l'emploi, l'économie, le racisme et l'islamophobie, le terrorisme, l'anxiété et la dépression, les idées de suicide, l'élection d'un Président noir et celle de Trump. Le présent livre a pour objectif de fournir des exemples de ce qui peut être fait avec toutes ces données (*Big Data*).

Pour l'auteur, comme pour bien des scientifiques et des observateurs, il y a quelque chose de révolutionnaire dans l'information que l'on peut

1. BDSM : Bondage, Dominance, Sadism, Masochism.

obtenir sur la psychologie humaine² à partir de toutes ces données éparses. Cette information provenant de l'analyse des données permet de répondre à des questions telles que : Combien d'hommes sont gais? Est-ce que la publicité est efficace? Les médias sont-ils biaisés? Qui triche dans son rapport d'impôts? Est-ce important d'être admis dans une université prestigieuse? Pouvons-nous prévoir le cours des actions? Qu'est-ce qui prédit un cancer du pancréas ou la dépression? (p. 22)

Le but de la science des données, comme toute science, est de « comprendre le monde ». Cette science est à la fois intuitive et rigoureuse. Elle exige de se méfier de ses propres biais. Elle consiste moins à examiner d'énormes quantités de données qu'à analyser des données pertinentes et à se poser des questions nouvelles. Ces questions paraissent parfois insignifiantes, mais elles peuvent orienter néanmoins le chercheur vers quelque chose de nouveau sur le comportement humain. Comme dans toute science, il ne faut pas se limiter aux seules données de l'Internet, car les autres sources dites traditionnelles sont également utiles. À la suite de cette démystification de la science des données (chapitre 1), l'auteur consacre quatre chapitres – le cœur de l'ouvrage – aux « forces » des *Big Data*.

La première force de la science des données consiste à offrir de nouveaux types de données, ce que l'auteur appelle des « données réimaginées ». Par exemple, pendant une période de chômage massif, on pourrait s'attendre à ce que les recherches sur Google portent sur des mots comme « Bureau de l'emploi » ou « Nouvel emploi ». Or, une étude de Stephens-Davidowitz a constaté que les clics les plus fréquents se rapportent à des sites pornographiques (*slutload*, *spider solitaire*, *rawtube*). Il y a trouvé des informations qui n'avaient jamais été collectées (ou même imaginées). La révolution *Big Data* consiste donc à fureter dans des endroits nouveaux.

La deuxième force de la science des données consiste à fournir des données « honnêtes » et objectives, ce que l'auteur appelle un « sérum de vérité digitale ». En guise d'exemple, le lecteur peut se référer à la liste des « mensonges » présentée au début du présent texte. Ajoutons l'anecdote suivante. Ahmadinejad, président de l'Iran, a déclaré en 2007 que dans son pays il n'y avait pas d'homosexuels. Un rapide coup d'œil de Stephens-Davidowitz dans l'Internet lui a fait voir « un intérêt significatif pour la porno gay en Iran » (p. 119). Il apparaît donc que le sérum de vérité digitale nous fait voir le monde pire que nous l'avions pensé. Ce peut

2. Il ne s'agit donc pas d'informations sur des individus particuliers, mais de données anonymes qui permettent de dresser un portrait de catégories de personnes ou d'une société.

être déprimant. Par contre, comprenant mieux la réalité, il devient possible d'améliorer notre vie et celle de nos semblables.

La troisième force de la science des données permet d'examiner un petit nombre de personnes ou de détails. De même qu'il faut un grand nombre de pixels dans une photo pour bien voir un détail, de même avec l'Internet il faut assez de données pour focaliser un point précis. Par exemple, on a observé que les possibilités d'accéder à de meilleures conditions financières que ses parents sont faibles aux États-Unis : 7,5 % pour l'ensemble du pays (au Canada, c'est 13,5 %). Toutefois, on a noté des différences marquées selon les régions : Charlotteville, NC : 4,4 %; Washington, DC : 10,5 %; San José, CA : 12,9 %. Avec les données disponibles, il est également possible d'établir de semblables statistiques pour l'espérance de vie, la provenance des riches, le nombre de femmes enceintes, la criminalité au jour le jour, etc. Un outil spécial (dont l'auteur fait grand cas) consiste à trouver son « double » (*doppelgänger*) ou ceux qui nous ressemblent. C'est la meilleure façon de prédire ce que nous ferons dans un avenir prochain. Si, par exemple, un médecin trouve dans une banque de données un ensemble d'individus ayant déjà eu les mêmes symptômes que son client, son diagnostic sera beaucoup plus approprié que s'il se basait seulement sur sa propre expérience.

La quatrième force de la science des données consiste à exécuter des expérimentations qui établissent la causalité entre deux variables. Par exemple, le 27 février 2000, les ingénieurs de Google ont réalisé une expérimentation qui a révolutionné l'Internet. Les individus du groupe expérimental ont reçu 20 liens concernant leur recherche, tandis que ceux du groupe contrôle ont reçu les 10 liens fournis habituellement. Les ingénieurs ont ensuite mesuré la satisfaction de leurs clients sur la base de la fréquence de l'utilisation ultérieure de Google³. Donc, une expérimentation simple, rapide et peu coûteuse qu'on appelle dans le milieu « un test A/B ». En 2011, ce test avait été effectué 7000 fois chez Google et la méthode s'est répandue et fut adoptée par Facebook. Il est facile d'imaginer l'intérêt de cette méthode pour les grandes entreprises. Elles l'utilisent pour vérifier, par exemple, l'efficacité de leur publicité. On peut alors comprendre pourquoi se développe l'assuétude par rapport au monde virtuel : « Il y a mille spécialistes derrière l'écran dont le travail consiste à briser votre autorégulation », d'expliquer l'éthicien Tristan Harris (rapporté à la page 220). Enfin, l'auteur considère que ces « expérimentations naturelles » sont puissantes et seront utilisées bientôt par de nombreux scientifiques dans diverses disciplines.

3. L'auteur ne donne pas le résultat de cette expérimentation parce que son propos est plutôt de présenter la nouveauté de cette approche qui a été reprise chez Google et empruntée par d'autres entreprises.

Dans la dernière section, Stephens-Davidowitz nous prévient que les *Big Data* doivent être utilisées avec précautions. Aux prises avec un très grand nombre de variables et avec les règles de l'éthique, la science des données, comme les autres sciences, est limitée. Cependant, avec toutes les informations dont disposent les grandes corporations et les gouvernements, les possibilités de dérapage sont nombreuses : exploitation, discrimination, viol de la vie privée, etc. Le citoyen doit être vigilant et utiliser, lui aussi, le pouvoir que donnent les *Big Data*.

En conclusion, l'auteur déplore le fait que la majorité des universitaires ignorent l'explosion des données provoquée par l'âge digital. Il suggère : « Dans le champ de la psychologie, les chercheurs pourraient utiliser les outils de la Silicon Valley pour améliorer considérablement leur recherche » (p. 277).

APPRÉCIATION

Everybody lies est l'œuvre d'un jeune chercheur, brillant diplômé de Harvard en économie et féru de la science des données. Après avoir travaillé chez Google comme expert en science des données, il collabore maintenant avec le *New York Times* et est conférencier invité à la Wartom School of Management de l'Université de Pennsylvanie.

Everybody lies se lit comme un roman, pour reprendre un poncif. Il est bien documenté et bien structuré. À l'intérieur des chapitres, les résultats de recherches, les anecdotes et les statistiques s'accumulent comme une liste d'épicerie en un joyeux mélange présenté avec passion et humour. À certains moments, nous sommes en face de nos mensonges et nos prétentions, derrière les apparences, car l'analyse des données révèle le côté sombre de la nature humaine, mais cette connaissance nous donne également la possibilité de nous connaître et d'agir. Il s'agit d'une façon novatrice – révolutionnaire même – d'approcher les questions et phénomènes que les chercheurs en sciences sociales devraient considérer avec attention et ouverture.

Le titre – *Everybody lies* – est sans doute un titre accrocheur, mais plutôt... mensonger. C'est le sous-titre qui traduit véritablement le contenu de l'ouvrage.

Ce livre intéressera tous les humains que nous sommes, car nous pouvons en tirer profit selon notre niveau de culture : les étudiants, les professeurs, les politiciens, les entrepreneurs et, évidemment, les mordus de l'Internet.

Ce livre témoigne d'une nouvelle façon d'étudier l'esprit [...] et offre une vue sans précédent du psychisme humain. [...] S. Stephens-Davidowitz

trace une nouvelle route aux sciences sociales pour le vingt et unième siècle (Steven Pinker, préface, p. x et xi).

Léandre Bouffard⁴
Université de Sherbrooke

4. Courriel de correspondance : Leandrebouffard1939@yahoo.ca