

Dominique Cardon

À quoi rêvent les algorithmes

Nos vies à l'heure des *big data*



LA REPUBLIQUE DES IDEES

Seuil



Google, Facebook, Amazon, mais aussi les banques et les assureurs : la constitution d'énormes bases de données (les « *big data* ») confère une place de plus en plus centrale aux algorithmes. L'ambition de ce livre est de montrer comment ces nouvelles techniques de calcul bouleversent notre société. À travers le classement de l'information, la personnalisation publicitaire, la recommandation de produits, le ciblage des comportements ou l'orientation des déplacements, les mégacalculateurs sont en train de s'immiscer, de plus en plus intimement, dans la vie des individus. Or, loin d'être de simples outils techniques, les algorithmes véhiculent un projet politique. Comprendre leur logique, les valeurs et le type de société qu'ils promeuvent, c'est donner aux internautes les moyens de reprendre du pouvoir dans la société des calculs.

Dominique Cardon est sociologue au Laboratoire des usages d'Orange Labs et professeur associé à l'université de Marne-la-Vallée (LATTS). Avec *La Démocratie Internet* (Seuil/La République des Idées, 2010) et de nombreux articles, il s'est imposé comme l'un des meilleurs spécialistes du numérique et d'Internet.

www.seuil.com et www.repid.com



ISBN 978.2.02.127996.2/Imprimé en France 10.2015

11,80 €

Table

INTRODUCTION	
Comprendre la révolution des calculs	7
Chiffrer le monde	8
Ouvrir la boîte noire	12
CHAPITRE PREMIER	
Quatre familles de calcul numérique	17
À côté du web : l'imprécise popularité des clics ...	19
Au-dessus du web : l'autorité des méritants	24
À l'intérieur du web : la fabrique de la réputation .	29
Au-dessous du web : la prédiction par les traces ...	33
CHAPITRE 2	
La révolution dans les calculs	39
La manipulation du réel	40
Le débordement des catégories	44
Calculer au plus près	47
Corrélations sans causes	51
CHAPITRE 3	
Les signaux et les traces	55
Les nouveaux gisements de données	55

Des machines « statistiques »	58
Le signal et la trace.	62
Un comportementalisme radical.	66
Signaux sans traces et traces sans signaux	71
La quantification de soi	76
Les algorithmes sont-ils biaisés ?	79
L'« idiotie » des algorithmes	84
CHAPITRE 4	
La société des calculs	89
La « tyrannie du centre »	89
La coordination virale de l'attention.	91
La sécession des excellents	94
Digital labor	96
« Passer en manuel »	100
CONCLUSION	
La route et le paysage	105

INTRODUCTION

Comprendre la révolution des calculs

Un nouvel objet a fait son entrée dans nos vies : les algorithmes.

Ce terme d'informatique a une signification bien plus large qu'on ne le croit. Comme la recette de cuisine, un algorithme est une série d'instructions permettant d'obtenir un résultat. À très grande vitesse, il opère un ensemble de calculs à partir de gigantesques masses de données (les « *big data* »). Il hiérarchise l'information, devine ce qui nous intéresse, sélectionne les biens que nous préférons et s'efforce de nous suppléer dans de nombreuses tâches. Nous fabriquons ces calculateurs, mais en retour ils nous construisent.

Il n'est plus beaucoup de gestes quotidiens, d'achats, de déplacements, de décisions personnelles ou professionnelles qui ne soient orientés par une infrastructure de calculs. Quand elle vient soudainement à disparaître, comme lorsqu'une panne interrompt le trafic téléphonique, nous sommes désemparés. Pourtant, dès que nous pensons à la présence des calculateurs dans nos sociétés, nous maudissons la froide rationalité des machines et redoutons qu'elles ne prennent le pouvoir sur nous. Nous aimons leur opposer « notre » subtile sagacité.

Pourtant, les technologies trament notre monde depuis si longtemps qu'il est erroné de séparer les humains de leur environnement sociotechnique. Des premiers outils préhistoriques à l'invention de l'écriture, de la mécanisation de l'imprimerie à la numérisation de l'information, de la création des listes et des tableaux comptables au calcul scientifique, la longue histoire des technologies intellectuelles est au cœur de l'évolution de l'humanité. Il serait naïf de croire qu'elles n'ont pas transformé profondément ce que nous sommes, ce que nous savons, nos manières de penser et les représentations que nous avons de nous-mêmes. Nous vivons dans une telle proximité avec les technologies que ce couple ne peut plus être défait sans que nous amputions la meilleure part de nous-mêmes.

Comme l'invention du microscope a ouvert une nouvelle fenêtre sur la nature, les capteurs numériques sont en train de jeter leur filet sur le monde pour le rendre *mesurable en tout*. Le savoir et les connaissances, les photographies et les vidéos, nos mails et ce que nous racontons sur Internet, mais aussi nos clics, nos conversations, nos achats, notre corps, nos finances ou notre sommeil deviennent des données calculables.

Aussi est-il essentiel de comprendre, de discuter et de critiquer la manière dont les algorithmes impriment leurs marques sur nos existences, jusqu'à devenir indiscutables et même invisibles. L'objet de ce livre est de comprendre ce que la révolution des calculs apportée par les *big data* est en train de faire à nos sociétés. Il décrit le monde auquel rêvent les algorithmes, avant que nous nous réveillions – trop tard.

Chiffrer le monde

Avant leur spectaculaire entrée dans nos vies quotidiennes, les calculs étaient surtout l'affaire d'États et d'entreprises.

Longtemps, la mesure statistique a été une question de spécialistes. Le grand public n'en percevait l'écho qu'à travers la publication d'indicateurs simplifiés venant justifier des choix de politiques publiques. Indispensable colonne vertébrale des États et des marchés, les grandes institutions statistiques ont très vite été gouvernées par des professionnels de la mesure, usant d'outils et de modèles de plus en plus complexes.

Instruments de connaissance, les statistiques étaient aussi conçues comme des instruments politiques aux mains des décideurs. En « photographiant » le monde, elles donnaient aux hommes de pouvoir des outils pour évaluer, choisir et faire agir¹. Depuis leur tour d'ivoire, statisticiens, sociologues et économètres veillaient à ce que l'existence des mesures n'influence pas le comportement des « mesurés ». À partir des politiques néolibérales des années 1980, on assiste à une généralisation de la calculabilité et à une systématisation de la politique des indicateurs. La présence des quantificateurs dans la vie sociale se fait partout sentir. Baromètres, indices et palmarès entreprennent de chiffrer des activités qui, jusqu'alors, n'étaient pas mesurées ou dont la quantification ne faisait pas l'objet d'une attention constante et inquiète.

Les instruments statistiques sont devenus une technique de gouvernement. L'évaluation des politiques publiques en fonction d'objectifs chiffrés s'est généralisée. Les palmarès d'écoles, d'hôpitaux ou de régions où il fait bon vivre font la une des magazines. Les outils de gestion s'introduisent dans les activités les plus quotidiennes des salariés. Les systèmes de notation financière branchent leurs résultats sur une interminable chaîne de mécanismes comptables². Sous prétexte d'efficacité, les indicateurs

1. Alain Desrosières, *La Politique des grands nombres. Histoire de la raison statistique*, Paris, La Découverte, 2000.

2. Pierre Lascoumes et Patrick Le Galès (dir.), *Gouverner par les instruments*, Paris, Les Presses de Sciences Po, 2004.

se sont répandus dans la société pour fournir à ceux qui étaient mesurés des chiffres destinés à orienter leurs comportements¹.

L'objectif de ces indicateurs est moins de connaître le réel que de « conduire les conduites² » des individus pour qu'ils le transforment. Les statisticiens traditionnels se sont trouvés désemparés devant ce déluge de chiffres peu fiables, mais, désormais, ils n'ont guère de prise sur la manière dont les entreprises et les administrations se nourrissent, jusqu'à l'asphyxie, de chiffres destinés à comparer et à évaluer, dans une logique de compétition et de performance. Le tournant de la « politique des indicateurs », qui a vu les statistiques descendre dans le monde social, continue d'étendre les dispositifs de commensuration à un nombre toujours plus important de secteurs d'activités³.

Aujourd'hui, une nouvelle vague d'extension de la calculabilité est en marche. Son ampleur est inédite et ses conséquences, bien qu'encore difficiles à évaluer, sont considérables. Sur la logique des indicateurs chiffrés se greffe désormais celle du calcul algorithmique embarqué à l'intérieur des interfaces numériques. En rencontrant l'informatique, les chiffres sont devenus des signaux numériques (listes, boutons, compteurs, recommandations, fils d'actualité, publicité personnalisée, trajet GPS, etc.) qui habillent toutes les interfaces que, d'un clic, nous ne cessons de caresser. Ils pénètrent si intimement notre vie quotidienne que nous percevons mal les longues chaînes qui conduisent des sympathiques écrans colorés aux grandes infrastructures statistiques que la révolution numérique installe dans de lointains serveurs de données.

1. Isabelle Bruno et Emmanuel Didier, *Benchmarking. L'État sous pression statistique*, Paris, Zones, 2013.

2. Michel Foucault, « Le sujet et le pouvoir », in *Dits et écrits*, Paris, Gallimard, coll. « Quarto », 1982, tome II, p. 1041-1062.

3. Barbara Cassin (dir.), *Derrière les grilles. Sortons du tout-évaluation*, Paris, Mille et une nuits, 2014.

À très grande vitesse, un nombre croissant de domaines – la culture, le savoir et l’information, mais aussi la santé, la ville, les transports, le travail, la finance et même l’amour et le sexe – sont désormais outillés par des algorithmes. Ils organisent et structurent les informations, aident à prendre des décisions ou automatisent des processus que nous avons l’habitude de contrôler nous-mêmes. Deux dynamiques s’avancent pour nous faire entrer dans cette nouvelle *société des calculs*.

La première est l’accélération du processus de numérisation de nos sociétés, qui nourrit de gigantesques bases de données d’informations, lesquelles n’avaient jamais été enregistrées, rendues accessibles et facilement manipulables. Un torrent de données se déverse aujourd’hui sur Internet. Chaque jour, 3,3 milliards de requêtes sont effectuées sur les 30 000 milliards de pages indexées par Google ; plus de 350 millions de photos et 4,5 milliards de *likes* sont distribués sur Facebook ; 144 milliards d’e-mails sont échangés par 3 milliards d’internautes. Si l’on numérisait toutes les communications et les écrits depuis l’aube de l’humanité jusqu’en 2003, il faudrait 5 milliards de gigabits pour les mettre en mémoire. Aujourd’hui, nous générons ce volume d’informations numériques en deux jours !

À l’instar des grandes révolutions industrielles, toutes initiées par l’exploitation d’un nouveau type d’énergie, le « nouvel or » des données numériques constitue, pour les promoteurs des *big data*, un gisement de valeur susceptible de relancer l’innovation, la productivité et la croissance. Aussi invitent-ils les institutions et les entreprises à conduire des politiques d’ouverture des données (*open data*), les citoyens à s’emparer des données pour surveiller les puissants (*sousveillance*) et les médias à pratiquer le « journalisme de données » (*data journalism*).

Si le discours public se concentre aujourd’hui sur le volume extravagant des données numérisées et les menaces que leur extraction fait peser sur la vie privée des individus, le principal défi que doivent affronter les *big data* est de donner du sens à ce

magma de données brutes. Aussi la deuxième dynamique qui nourrit la société des calculs est-elle le développement de procédés, les algorithmes, donnant aux ordinateurs des instructions mathématiques pour trier, traiter, agréger et représenter les informations. Venues de mondes différents, ceux du marketing, des marchés financiers ou de l'actuariat, de puissantes techniques statistiques (notamment celles que l'on qualifie de « prédictives ») se déploient à grande échelle en profitant de l'exceptionnelle augmentation des capacités de calcul des ordinateurs.

Greffés à nos écrans, les classements, palmarès, compteurs, cartes, recommandations et notes de toutes sortes figurent les pointes émergées de la calculabilité des traces de nos activités. À partir de données toujours plus inattendues (déplacements, tickets de caisse, clics sur Internet, consommation électronique, temps de lecture d'un livre sur les tablettes électroniques, nombre de pas enregistrés par un podomètre), les algorithmes chiffrent le monde, le classent et prédisent notre avenir.

Ouvrir la boîte noire

Omniprésents, ces calculs restent pour nous mystérieux. Ils orientent des décisions, appareillent des processus automatiques et justifient des choix politiques, mais nous interrogeons rarement la manière dont ils ont été produits. Nous regardons leurs effets sans examiner leur fabrication. Quelles sont les données qui servent au calcul ? Comment l'information a-t-elle été quantifiée ? Quels sont les principes de représentation qui animent le modèle statistique mis en œuvre pour classer tel objet plutôt que tel autre ? Qui pilote le codage des calculs et quels sont ses objectifs ?

Habités par un sentiment d'incompétence, nous préférons ignorer les conditions de fonctionnement de la société des calculs, en laissant les clés aux statisticiens, aux informaticiens et aux

économètres. La complexification des modèles algorithmiques mis en œuvre dans les nouvelles infrastructures informationnelles contribue à imposer le silence à ceux qui sont soumis à leurs effets. Elle désarme aussi ceux qui entreprennent de critiquer l'avènement de la froide rationalité des calculs, sans chercher à en comprendre le fonctionnement. Par facilité autant que par ignorance, la critique du nouvel empire des calculs s'est réfugiée dans une pseudo-opposition entre les « humains » et les « machines ». Elle dénonce confortablement la rationalisation néolibérale du monde, la tyrannie de l'évaluation ou les accidents automatisés du *high-frequency trading*.

Si cette défiance constitue une sorte de contrepoison, elle reste plus gratifiante que véritablement efficace. La critique de la raison calculatoire ne peut opposer qu'une rêverie pastorale à la marche automatisée des grands systèmes technologiques mondiaux. Pour vraiment critiquer une dynamique qui possède de si puissants moteurs économiques et culturels, il est nécessaire d'*entrer dans les calculs*, d'explorer leurs rouages et d'identifier leurs visions du monde. Avant de réduire la logique calculatoire aux intérêts économiques de ceux qui la fabriquent, il faut commencer par allonger les algorithmes sur le divan et entendre la variété de leurs désirs. Cet examen est indispensable si l'on veut débattre publiquement des calculs que nous voulons et de ceux dont nous ne voulons pas, contrôler leurs agissements et leur opposer des calculs alternatifs. Une radiographie critique des algorithmes est un enjeu démocratique aussi essentiel qu'inaperçu.

L'objet de ce livre est d'éclairer les enjeux sociaux, éthiques et politiques qui accompagnent le développement du calcul algorithmique, en prêtant attention au principal foyer des bouleversements en cours : celui des données numériques et, plus spécifiquement, du classement de l'information sur le web. Ma conviction est que, face au déploiement de la société des calculs, il est nécessaire d'encourager la diffusion d'une culture statistique vers un public beaucoup plus large que celui des seuls spécialistes.

Mais le propos de ce livre n'est pas mathématique : il est pleinement politique. La manière dont nous fabriquons les outils de calculs, dont ils produisent des significations, dont nous utilisons leurs résultats, trame les mondes sociaux dans lesquels nous sommes amenés à vivre, à penser et à juger. Les calculs habitent nos sociétés bien plus centralement que ne l'imaginent ceux qui voudraient les réduire à des fonctions mathématiques et rejeter la technique hors de la société, comme un *alien* menaçant. Les calculateurs fabriquent notre réel, l'organisent et l'orientent. Ils produisent des conventions et des systèmes d'équivalence qui sélectionnent certains objets au détriment d'autres, imposent une hiérarchisation des valeurs qui en vient progressivement à *dessiner les cadres cognitifs et culturels* de nos sociétés.

Comme l'ont souligné beaucoup de travaux d'histoire et de sociologie, les objets techniques ne fonctionnent que parce qu'ils opèrent dans un « milieu associé » qui les rend efficaces et pertinents¹. Les calculs ne calculent vraiment que dans une société qui a pris des plis spécifiques pour se rendre calculable. Aussi faut-il comprendre comment nos sociétés secrètent certaines manières de se chiffrer plutôt que d'autres. Que valorisent-elles dans leur façon de compter et de classer ?

Il suffit d'ouvrir la boîte noire des calculateurs pour constater qu'ils servent des desseins très différents. Selon la nature des données enregistrées, la manière de les catégoriser, le choix des techniques statistiques ou les options de visualisation des résultats, le fait de modifier les paramètres du calcul conduit à valoriser des choses très différentes. Face aux visées productivistes de la mesure du PIB, des économistes hétérodoxes opposent d'autres « indicateurs de richesse », comme l'indice de développement humain (IDH) du Programme des Nations unies pour

1. Gilbert Simondon, *Du mode d'existence des objets techniques*, Paris, Aubier, 1989 ; et Bruno Latour, *La Science en action*, Paris, La Découverte, 1989.

le développement (PNUD) popularisé par Amartya Sen. Ils voudraient déplacer le centre de gravité sur lequel reposent les calculs macro-économiques mondiaux vers la prise en compte de nouvelles variables, comme l'espérance de vie à la naissance, le niveau d'éducation, la qualité de vie ou le bonheur¹.

En modifiant la traditionnelle mesure de répartition des revenus par décile pour la décomposer en centiles, Camille Landais et Thomas Piketty ont fait apparaître l'explosion récente des écarts de richesse en faveur du 1 % de la population, qui n'apparaissait pas avec un filtre moins fin. La perspective nouvelle offerte par ce changement de lunette statistique a inspiré le slogan « Nous sommes les 99 % » au sein du mouvement des Indignés et Occupy, au début des années 2010². Il ne faut pas grand-chose – choisir d'autres variables, changer l'échelle, calculer autrement – pour faire du chiffrage la meilleure arme contre d'autres chiffrages³. Il est crucial de lutter contre cette sorte de fatalisme qui nous conduit à imputer aux mesures ce que, en réalité, nous leur avons demandé de faire.

On croit volontiers qu'un unique moteur anime la guerre de conquête entreprise par les calculs : la performance économique. Dans cet ouvrage, on n'abordera pas directement les enjeux économiques de la domination des grandes plateformes du web, les fameux GAFAs (Google, Apple, Facebook, Amazon). Leurs ambitions, leurs intérêts, leur culture californienne sont à la une des magazines et sont désormais bien connus. Ce livre ne propose pas de critiquer les algorithmes de l'extérieur, en en

1. Éloi Laurent et Jacques Le Cacheux, *Un nouveau monde économique. Mesurer le bien-être et la soutenabilité au XXI^e siècle*, Paris, Odile Jacob, 2015.

2. Camille Landais, « Les hauts revenus en France (1998-2006). Une explosion des inégalités ? », École d'économie de Paris [disponible sur [http://ur1.cal/ge0sw](http://ur1.cal.ge0sw)].

3. Isabelle Bruno, Emmanuel Didier et Julien Prévieux (dir.), *Stat-Activisme, Comment lutter avec des nombres ?*, Paris, Zones, 2014.

faisant les reflets des intérêts de leurs concepteurs, mais de comprendre de l'intérieur la manière dont ils produisent des effets (plus ou moins critiquables) sur nos sociétés.

C'est, la plupart du temps, au nom de l'efficacité que les techniques calculatoires colonisent des univers toujours plus nombreux. Les recherches sur Google sont de plus en plus personnalisées, afin de répondre au mieux à nos attentes et d'anticiper des désirs que nous ne connaissons pas encore. Amazon voudrait nous envoyer des livres avant même que nous ne les ayons commandés tellement, forte de ses calculs, l'entreprise pense savoir ce que nous voudrions lire. Depuis quelques années, le marché de la publicité numérique nourrit l'espoir que, devenue « personnelle », la publicité ciblée perdra son caractère intempestif pour devenir, aux yeux de ceux qu'elle vise, une information comme une autre.

Ces manières de chiffrer l'information font souvent l'objet de critiques. Elles enferment les individus dans la bulle de leurs propres choix, plient leur destin dans l'entonnoir du probable et nourrissent la précision du ciblage d'une capture disproportionnée d'informations personnelles. Mais elles n'adviennent que parce qu'elles font écho à des transformations des modes de vie et des aspirations que suscitent les processus d'individualisation de nos sociétés. La thèse de ce livre est que, si les logiques de personnalisation s'installent aujourd'hui dans nos vies, c'est parce qu'elles calculent une forme nouvelle du social, la société des comportements, où se recompose la relation entre le centre de la société et des individus de plus en plus autonomes.

CHAPITRE PREMIER

Quatre familles de calcul numérique

Afin d'approcher de façon simple les enjeux qui président aux classements de l'information numérique, je proposerai un récit faisant se succéder quatre manières de produire de la visibilité avec des calculs. De façon métaphorique, on identifiera ces quatre familles en fonction de la place qu'occupe le calculateur par rapport au monde qu'il entend décrire. Les mesures peuvent se trouver *à côté, au-dessus, dans* ou *en dessous* des données numériques. La réalité des calculateurs du web n'a rien d'aussi géométrique. Mais le défi que voudrait relever cet ouvrage est de proposer une clé de lecture, afin que chacun se sente autorisé à ouvrir la boîte noire des algorithmes du web.

Résumons tout de suite, avant de les explorer, ces quatre familles. Les mesures d'audience, en premier lieu, se placent à côté du web pour dénombrer les clics des internautes et ordonner la *popularité* des sites. La famille de mesures issue du Page-Rank, l'algorithme de classement de l'information du moteur de recherche de Google, voudrait se situer au-dessus du web, afin de hiérarchiser l'*autorité* des sites au moyen des liens hypertextes qu'ils s'échangent. Les mesures de *réputation*, qui se sont développées avec les réseaux sociaux d'Internet et les sites de notation, se positionnent à l'intérieur du web, afin de donner aux internautes

des compteurs qui valorisent la *réputation* des personnes et des produits. Enfin, les mesures *prédictives* destinées à personnaliser les informations présentées à l'utilisateur déploient, sous le web, des méthodes statistiques d'apprentissage pour calculer les traces de navigation des internautes et leur prédire leur comportement à partir de celui des autres.

Le parcours que nous allons entreprendre à travers ces quatre manières de classer l'information numérique permettra de dégager les différentes valeurs qui nourrissent les choix que font les algorithmes : la popularité, l'autorité, la réputation et la prédiction. On montrera notamment comment ceux-ci prélèvent sur le web des données différentes (clics, liens, *likes* et traces) pour les soumettre à des opérations répondant à différentes conventions statistiques que résume le tableau ci-dessous.

	À côté 	Au-dessus 	Dans 	Au-dessous 
Exemples	Médiamétrie, Google Analytics, affichage publicitaire	PageRank de Google, Digg, Wikipédia	Nombre d'amis Facebook, Retweet de Twitter, notes et avis	Recommandation Amazon, publicité comportementale
Données	Vues	Liens	Likes	Traces
Population	Échantillon représentatif	Vote censitaire, communautés	Réseau social, affinitaire, déclaratif	Comportements individuels implicites
Forme du calcul	Vote	Classements méritocratiques	<i>Benchmark</i>	<i>Machine learning</i>
Principe	<i>Popularité</i>	<i>Autorité</i>	<i>Réputation</i>	<i>Prédiction</i>

En distinguant ces quatre familles, souvent confondues et mal comprises, on voudrait faire apparaître à quel point les techniques de calcul justifient, selon des principes différents, la manière dont elles produisent un ordre de tel type plutôt que de tel autre. Entrer dans les calculs, c'est comprendre que les enjeux techniques et industriels qui opposent les acteurs du web nourrissent aussi une compétition sur la bonne manière d'organiser la visibilité des informations.

Afin de présenter ces quatre familles du calcul numérique, je suivrai un fil chronologique. Même si, au sein des actuelles plateformes du web, les quatre familles de calcul cohabitent et se mélangent les unes aux autres, elles sont apparues successivement dans l'histoire d'Internet. Leur chronologie fait apparaître les points de bascule qui nourrissent les tensions actuelles entre les différents écosystèmes du web.

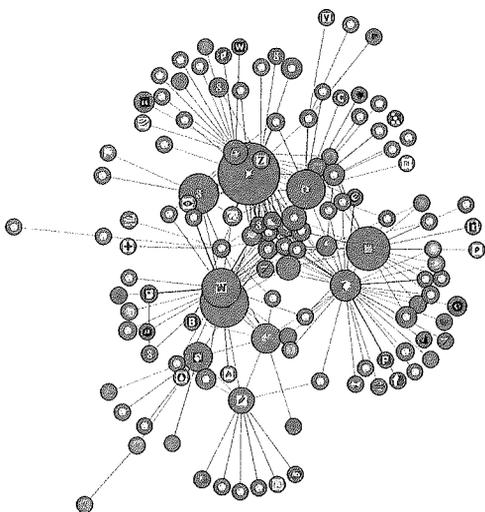
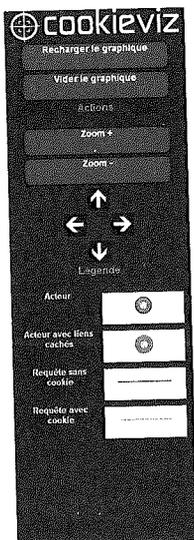
À côté du web : l'imprécise popularité des clics

La première technique de calcul se place à côté du web pour mesurer l'audience des sites et leur popularité. Alors que les pionniers d'Internet n'y prêtaient guère attention, quand le réseau a pris son essor, au début des années 1990, certains web-mestres ont cherché à connaître le nombre de visiteurs sur leur site. Ils ont fabriqué un indicateur dont le principe est directement inspiré des techniques de mesure d'audience des médias de masse : compter les clics des visiteurs.

Afin d'éviter de dénombrer plusieurs fois le même internaute, la notion de « visiteur unique » – vérifiée à travers l'adresse IP de l'ordinateur – est la principale unité de compte de la popularité des médias en ligne et, par simple équivalence, du tarif des publicités qu'ils accueillent. La mesure d'audience mime le vote démocratique : chaque internaute qui clique dispose d'une voix

pour que le cookie envoie à l'entreprise de Mark Zuckerberg une information sur son passage sur ce site.

CookieViz¹, une extension du navigateur développée par la CNIL, permet de visualiser les multiples trajectoires que prennent les données de l'utilisateur lorsque, croyant se connecter au site de son quotidien préféré, il est en réalité en train d'envoyer des informations à une dizaine de serveurs obscurs dispersés à travers le monde.



Lecture: En naviguant sur le site représenté par le rond au centre du schéma, l'internaute envoie des informations de navigation aux autres sites et serveurs de données représentés sur ce graphe.

Source: CNIL.

Aux formes traditionnelles de caractérisation de l'internaute dans les catégories du marketing (âge, niveau de vie, revenu)

1. Disponible au téléchargement: <http://www.cnil.fr/vos-droits/vos-traces/les-cookies/telechargez-cookieviz/>

CHAPITRE 2

La révolution dans les calculs

La révolution des *big data* se trouve moins dans l'accumulation des données que dans la manière de les calculer. Chaque famille de calculateurs suppose des conceptions et des usages différents de la statistique. L'arrivée des *data-analystes*, armés de leurs techniques de traitement massif des données, a mis le monde des statisticiens en ébullition. Trois secousses sont venues déplacer la manière dont nos sociétés se représentent à travers leurs chiffres : les mesurés sont eux-mêmes devenus plus facilement calculateurs ; les catégories parviennent moins bien à représenter des individus qui se singularisent de plus en plus ; les corrélations statistiques ne vont plus de la cause vers la conséquence, mais remontent des conséquences vers une estimation des causes probables.

Ces trois secousses ont fait vaciller la longue tradition statisticienne qui s'était constituée, en même temps que les États, pour proposer une cartographie de la nation autour de conventions stables et de catégories de description du monde social¹. Celles-ci garantissaient, d'une part, un certain degré

1. Emmanuel Didier, *En quoi consiste l'Amérique ? Les statistiques, le New Deal et la démocratie*, Paris, La Découverte, 2009.

CHAPITRE 2

La révolution dans les calculs

La révolution des *big data* se trouve moins dans l'accumulation des données que dans la manière de les calculer. Chaque famille de calculateurs suppose des conceptions et des usages différents de la statistique. L'arrivée des *data-analystes*, armés de leurs techniques de traitement massif des données, a mis le monde des statisticiens en ébullition. Trois secousses sont venues déplacer la manière dont nos sociétés se représentent à travers leurs chiffres : les mesurés sont eux-mêmes devenus plus facilement calculateurs ; les catégories parviennent moins bien à représenter des individus qui se singularisent de plus en plus ; les corrélations statistiques ne vont plus de la cause vers la conséquence, mais remontent des conséquences vers une estimation des causes probables.

Ces trois secousses ont fait vaciller la longue tradition statisticienne qui s'était constituée, en même temps que les États, pour proposer une cartographie de la nation autour de conventions stables et de catégories de description du monde social¹. Celles-ci garantissaient, d'une part, un certain degré

1. Emmanuel Didier, *En quoi consiste l'Amérique ? Les statistiques, le New Deal et la démocratie*, Paris, La Découverte, 2009.

CHAPITRE 3

Les signaux et les traces

Les promoteurs des *big data* font preuve d'un optimisme statistique à toute épreuve. Si Internet a libéré les individus du filtre des médias qui les empêchait de s'exprimer, il faudrait désormais libérer les données des fichiers et des modèles qui les cadennassent. Non sans naïveté, ils soutiennent qu'une fois les données brutes « libérées », il suffira de les calculer pour que les vérités mathématiques sous-jacentes au monde social apparaissent et permettent de réduire les erreurs des gouvernants, les approximations de la médecine ou le gaspillage des marchés.

Accessibles, croisées et livrées aux algorithmes, les données pourraient alors, elles aussi, exprimer des choses qui leur étaient interdites ou qui restaient jusqu'alors inconnues en l'absence de mesures objectives. Si notre monde est imparfait, c'est que nous manquons de données pour le corriger.

Les nouveaux gisements de données

Il est vrai que beaucoup d'entreprises et d'institutions disposent de riches bases de données et les exploitent mal. En

CHAPITRE 4

La société des calculs

Il est important de porter un regard critique sur le fonctionnement des calculateurs plutôt que de les laisser agir silencieusement. Mais cet ouvrage voudrait aussi proposer une lecture plus politique du type de société qui rend aujourd'hui possible le déploiement des algorithmes.

La « tyrannie du centre »

La généralisation des calculs offre un point de vue original pour saisir, à l'état chiffré, la manière dont nos existences se transforment à l'heure des *big data*. Les algorithmes ont entrepris de calculer la société par le bas, depuis le comportement des internautes. Leurs concepteurs partagent l'idée que les informations ne doivent pas être choisies par les journalistes, que les publicités ne peuvent pas être les mêmes pour tous, que les catégories d'appartenance traditionnelles représentent mal les individus et que chacun doit pouvoir choisir librement ses « contenus » sans subir le paternalisme des prescripteurs. Les zélotes californiens des *big data* ont pour projet de refabriquer nos sociétés à partir d'un réel

L'« idiotie » des algorithmes

Les effets des algorithmes de la quatrième famille, ceux prédisent les comportements des internautes à partir des tr prélevées sous le web, sont beaucoup plus difficiles à critiquer. Leurs agissements se réalisent souterrainement, dans les bases données des *data brokers*, et sans qu'il soit possible de comparer leurs résultats (puisqu'ils ont été personnalisés pour chaque). Parce qu'ils fonctionnent comme de purs automatismes procéduraux, les algorithmes donnent souvent des résultats statistiques imparfaits, stupides ou choquants.

Si, lorsque l'on tape le nom de certaines personnes dans Google, le moteur de recherche suggère parfois d'y ajouter « juif », c'est parce que beaucoup d'internautes l'ont déjà fait. Lorsqu'au 1^{er} janvier, Facebook propose aux utilisateurs un résumé illustré de leur année, l'algorithme sélectionne les publications qui ont suscité le plus d'interactions avec les amis, quitte à mettre en valeur la mort d'un proche. Les algorithmes suivent leurs procédures bêtement et ils manquent d'autant plus de tact et de sens moral que, ne calculant que des traces de comportements, ils font disparaître les catégories qui pourraient les empêcher de prendre en considération tel ou tel résultat. Les nouveaux calculateurs aspirent à être le reflet idiot d'une régularité statistique.

Latanya Sweeney, une chercheuse en informatique américaine, a remarqué que, lorsqu'elle tapait son nom dans le moteur de recherche de Google, elle voyait apparaître la publicité « Latanya Sweeney, arrested ? ». Cette publicité propose un service de consultation en ligne, *instantcheckmate.com*, qui permet, entre autres choses, de savoir si les personnes ont un casier judiciaire. Or le nom de ses collègues blancs n'était pas associé au même type de publicité, mais plutôt à des propositions

divertissement et des vidéos sur YouTube¹. Il n'est pas besoin que Facebook biaise ou censure son algorithme ; il suffit qu'il le laisse calculer l'information que cliquent ses utilisateurs en reproduisant l'inégale répartition du capital culturel selon les réseaux relationnels².

En épousant les comportements des internautes, l'algorithme de Facebook reconduit les inégalités de nos sociétés en donnant aux mieux dotés les moyens d'enrichir leurs réseaux relationnels et d'accéder à plus de ressources et d'opportunités. Il est à craindre que, pour les autres, l'ouverture informationnelle ne produise pas les mêmes effets. À la différence des métriques d'audience ou d'autorité, les compteurs du web social ne sont pas faits pour aider l'Internet expressif à rencontrer l'Internet silencieux. Ils laissent chacun construire son espace informationnel en fonction de ses ressources sociales et culturelles.

« Passer en manuel »

Sans doute le rêve ultime des nouveaux calculs est-il d'installer un environnement technique invisible permettant partout et pour tout de nous orienter sans nous contraindre. Amatrice de science-fiction, la critique des algorithmes dramatise à l'envi les risques totalitaires d'une rationalisation des existences. La diversité des informations, l'exposition à des connaissances multiples, la variété des choix ont connu une explosion si massive avec Internet qu'il est toujours surprenant d'entendre certains soutenir que nous serions enfermés dans la prison des algorithmes et des plateformes. Non sans contradiction, ce sont souvent les mêmes

1. Les résultats du projet ANR « ALGOPOL » sont accessibles sur <http://algopol.huma-num.fr>

2. François Héran, « La sociabilité, une pratique culturelle », *Économie et Statistique*, n° 216, 1998.

CONCLUSION

La route et le paysage

Les médias ont longtemps été pour nous des gyrophares. Ils nous prenaient par la main pour nous mener tout en haut de la montagne. Là, depuis la table d'orientation, ils nous désignaient les éléments notoires du paysage : le sacro-saint « panorama ». Placée au centre de la société, la table d'orientation organise l'attention collective afin que tous, prêtant attention aux mêmes motifs, partagent le sentiment de faire société commune. Avec Internet, l'orientation de notre attention est libérée au motif que chacun, depuis son propre véhicule, peut se déplacer librement dans un univers proliférant d'informations. Contre le paternalisme de la table d'orientation et son panorama obligé, chacun peut organiser son voyage sans avoir à suivre les directions prescrites par un guide, fût-il de haute montagne.

La conquête de cette liberté est sans pareille. Elle déboussole, fait perdre du temps et prendre des risques. Elle ouvre à chacun le droit d'errer, de se tromper et de s'émerveiller. Pourtant, perdu parmi les mille et un choix possibles, il a fallu trouver d'autres manières de se repérer et de s'organiser. Les calculateurs ont apporté une solution originale et audacieuse à cette désorientation. Ils constituent un moyen efficace pour trier dans

l'abondance d'informations disponibles et pour guider l'utilisateur vers ses propres choix.

Comme les GPS dans les véhicules, les algorithmes se sont glissés silencieusement dans nos vies. Ils ne nous imposent pas la destination. Ils ne choisissent pas ce qui nous intéresse. Nous leur donnons la destination et ils nous demandent de suivre « leur » route. La conduite sous GPS s'est si fortement inscrite dans les pratiques des conducteurs que ceux-ci ont parfois perdu toute idée de la carte, des manières de la lire, de la diversité de ses chemins de traverse et des joies de l'égarément.

Les algorithmes nous ont libérés des voyages de groupe, des points de vue obligés et des arrêts obligatoires devant des panoramas à souvenirs. Ils procèdent d'un désir d'autonomie et de liberté. Mais ils contribuent aussi à assujettir l'internaute à cette route calculée, efficace, automatique, qui s'adapte à nos désirs en se réglant secrètement sur le trafic des autres. Avec la carte, nous avons perdu le paysage. Le chemin que nous suivons est le « meilleur » pour nous.

Mais nous ne savons plus bien identifier ce qu'il représente par rapport aux autres trajets possibles, aux routes alternatives et peu empruntées, à la manière dont la carte compose un ensemble. Nous n'allons pas en revenir aux voyages de groupe et à leur guide omniscient. En revanche, nous devons nous méfier du guidage automatique. Nous pouvons le comprendre et soumettre ceux qui le conçoivent à une critique vigilante. Il faut demander aux algorithmes de nous montrer et la route, et le paysage.