Des machines et des hommes : quelles convergences ? Antoinette Rouvroy et Alain Ehrenberg.

ROUVROY, Antoinette; Ehrenberg, Alain

Published in:
Information psychiatrique

Publication date:
2021

Document Version
le PDF de l'éditeur

DES MACHINES ET DES HOMMES : QUELLES CONVERGENCES ?

Débat entre Antoinette Rouvroy et Alain Ehrenberg

Antoinette Rouvroy, Alain Ehrenberg

John Libbey Eurotext | « L'information psychiatrique »

2021/2 Volume 97 | pages 116 à 124

ISSN 0020-0204
DOI 10.1684/ipe.2021.2216

Article disponible en ligne à l'adresse :

https://www.cairn.info/revue-l-information-psychiatrique-2021-2-page-116.htm

L'Information psychiatrique 2021 ; 97 (2) : 116-24

Des machines et des hommes : quelles convergences ?

Débat entre Antoinette Rouvroy et Alain Ehrenberg

Antoinette Rouvroy [1]
Alain Ehrenberg [2]

[1] Docteur en sciences juridiques de l'Institut universitaire européen, chercheuse permanente du Fonds national de la recherche scientifique belge (FNRS) au Centre de recherche en information, droit et société de l'Université de Namur

[2] PhD, directeur de recherche émérite/Research director emeritus, CNRS Cermes 3, Centre de recherche médecine, sciences, santé, santé Mentale, société, Université de Paris, EHESS, CNRS, Inserm, 45 rue des Saints-Pères, 75270 Paris cedex 06

Débat entre Antoinette Rouvroy, juriste, philosophe du droit à Namur et Alain Ehrenberg, sociologue à Paris.

Co-présidents de séance Sonia Desmoulin-Canselier, juriste chargée de recherche CNRS-DataSanté à Nantes et Bernard Granger, professeur des universités, psychiatre à l'hôpital Cochin, Paris

Argument

La convergence des néotechnologies prépare des systèmes anthropotechniques qui concernent non seulement les dimensions techniques et économiques du monde commun, mais aussi les registres biologiques, psychologiques et sociaux. Cette transversalité des enjeux appelle à la réflexion éthique, juridique, psychologique de tous, citoyens et décideurs.

Comment les politiques publiques de santé mentale et de psychiatrie intègrent-elles ces impératifs ? Si les propositions des néotechnologies devaient apporter l'amélioration du service médical rendu attendue des neurosciences, comment influenceraient-elles les pratiques aujourd'hui soucieuses de prévention, d'accueil, de soin, d'inclusion en faveur de personnes vulnérables en souffrance ?

Correspondance : A. Rouvroy <antoinette.rouvroy@unamur.be>

Alain Ehrenberg : Ce petit exposé introductif à la conversation s'intitule « *Le cérébral et le digital. Quelle fabrique de l'autonomie* ? ». Je connais beaucoup mieux le cérébral que le digital dont je ne suis pas du tout un spécialiste. Ce que je ferai dans cet exposé, c'est quelques remarques sur la convergence ou la rencontre entre neurosciences cognitives et intelligence artificielle sous l'angle des imaginaires sociaux. « Neurosciences », « sciences cognitives », « sciences comportementales », ces appellations sont aujourd'hui partout. On a affaire à une sorte de raz-de-marée emportant dans un même mouvement psychologie et biologie, qui porte des promesses de progrès considérables dans des domaines aussi différents que la pathologie mentale, la gestion des ressources humaines ou les politiques éducatives, pour ne citer que ces quelques exemples. Dans un contexte global où la santé mentale et la souffrance psychique sont des préoccupations majeures de nos sociétés, ces disciplines suscitent des débats voire des polémiques. Aux espoirs, voire aux enthousiasmes des uns, sur le mode : « les neurosciences démontrent que ... », répondent les craintes de la réduction de l'homme à son cerveau, voire un biopouvoir au service du néolibéralisme. L'homme neuronal serait-il appelé à remplacer l'homme social ? Cette question n'est pas la bonne. Pourquoi ? Parce qu'elle nous enferme entre une mythologie du cerveau à laquelle répond une contre-mythologie du sujet et il me semble que l'on tourne en

Pour citer cet article : Rouvroy A, Ehrenberg A. Des machines et des hommes : quelles convergences ? *L'Information psychiatrique* 2021 ; 97 (2) : 116-24 doi:10.1684/ipe.2021.2216

rond. Le succès des neurosciences cognitives ne peut tenir aux seuls résultats scientifiques et médicaux quels qu'ils soient, pas plus qu'à d'obscures stratégies de pouvoir et de domination. L'engouement dont ces savoirs sont l'objet doit nécessairement reposer sur une autorité morale qu'ils permettent de mettre concrètement au travail.

Mon hypothèse c'est que ces disciplines cristallisent ou transfigurent dans des jeux de langages scientifiques certains idéaux de la modernité auxquels nous tenons fortement et qui sont les ressorts de leur autorité morale. Lesquels ? L'un des plus puissants idéaux de notre société de l'autonomie généralisée est l'individu capable, quels que soient ses handicaps, ses déviances, ses pathologies, de connaître des accomplissements en transformant ses handicaps en atouts. Je l'appelle « l'idéal du potentiel caché » : il est la forme spécifique par laquelle les populations diagnostiquées « malades mentales », « handicapées », « déviantes », et qui étaient il n'y a pas si longtemps enfermées, traitées dans des institutions closes sont devenues des individus capables non seulement de connaître des accomplissements malgré le mal qui les atteint, mais parfois plus encore grâce à lui. Alors cet idéal est me semble-t-il au cœur du message des neurosciences cognitives à savoir que l'individu, abordé sous l'angle de son cerveau, dispose toujours de ressources pour qu'il puisse se sortir des pires situations et entreprendre un itinéraire de transformation personnelle. Ce message est cristallisé par le concept biologique de « plasticité cérébrale », mais ce concept connaît une telle extension d'usage qu'il en est venu à incarner la possibilité de se changer soi-même sans limites et d'augmenter sa propre valeur. La définition de la plasticité cérébrale est beaucoup plus limitée : c'est l'idée que la nature n'a pas décidé totalement à l'avance de la répartition de nos fonctions mentales dans les différentes aires cérébrales, pour dire les choses très, très rapidement. Cet usage extensif de la plasticité cérébrale montre l'horizon d'attente d'où les neurosciences cognitives tirent une bonne part de leur autorité et qu'elles mettent au travail dans une société qui est imprégnée par les idées, les valeurs et les normes de l'autonomie, une société qui encourage le développement le plus large possible des capacités humaines. Le langage des neurosciences cognitives est d'abord un langage de l'action. Les pratiques qui s'en réclament – thérapies comportementales et cognitives, remédiation cognitive, *coaching*, etc. – sont généralement organisées comme des exercices, des entraînements dont le but est de modifier des habitudes de pensée ou de comportement, de les modifier jusqu'à ce qu'elles deviennent des réflexes inscrits dans la matière cérébrale grâce à la plasticité cérébrale, facilitant ainsi l'action de l'individu sur le modèle de l'entraînement du sportif ou du musicien. Donc, régularité, exercice, habitude qui augmentent la confiance en soi et l'empathie pour autrui, ces sciences ainsi ont dessiné les figures de l'« homme capable ». À travers la catégorie de l'exercice, ces disciplines cristallisent ou transfigurent dans des jeux de langage scientifique de types psychologique et biologique des idéaux sociaux traditionnels qui sont des idéaux de régularité et de fiabilité associés à des idéaux beaucoup plus récents de changement personnel et d'innovation. À travers cette association, les neurosciences cognitives constituent un des grands récits de l'individualisme contemporain.

Avec la rencontre entre les neurosciences cognitives et l'intelligence artificielle, entre le cérébral et le digital, les exercices ont trouvé de nouveaux supports dans la réalité virtuelle, les robots sociaux, la digitalisation des psychothérapies, etc., des nouveaux supports qui voient la machine non seulement démultiplier sa puissance de médiateur, mais encore passer du statut de médiateur à celui de partenaire, comme dans le cas de la robotique sociale. Les pratiques employant des technologies numériques visent à accroître, d'une part, les capacités de l'individu, qu'il s'agisse de prendre des habitudes ou d'accroître sa perspicacité en matière de reconnaissance des émotions (c'est l'« augmentation », pour reprendre l'un des deux mots-clés du monde numérique) et, d'autre part, les relations (ce sont les « connexions », pour reprendre l'autre mot-clé). Elles se déclinent des pathologies les plus lourdes (schizophrénie, autisme) jusqu'au développement personnel.

Hier, automatisme et quantification étaient synonymes de standardisation et s'opposaient à la créativité. Ils se trouvaient aux antipodes de tout ce qui pouvait se rapporter à la singularité individuelle. Aujourd'hui, à l'inverse, ils convoquent la créativité et l'intelligence relationnelle. Dans ce changement de statut, la machine se trouve en position de coach, elle accompagne le développement des potentiels de l'individu, elle le met en situation d'être l'agent de son propre changement. On peut distinguer trois types de programmes ou de machines : la réalité virtuelle, les robots sociaux et les thérapies automatisées.

Pour les personnes qui sont particulièrement dépourvues de compétences relationnelles et sociales, comme les personnes schizophrènes ou autistes, la réalité virtuelle offre des perspectives. Qu'apportent ces nouvelles technologies ? En gros, ces programmes possèdent le double avantage de permettre des entraînements dans des situations qui sont proches du monde réel, ils surmontent le « fossé du spectateur » dans lequel il est possible de déployer différentes modalités d'actions sensorielles, visuelles, etc., tout en conservant les avantages du dispositif expérimental pour mesurer l'activité psychologique ou cognitive et l'activité cérébrale. Ces programmes démultiplient les possibilités d'explorer la diversité des compétences, que ce soit avec *joystick* ou casque.

C'est aussi le cas avec les robots sociaux qui fonctionnent comme supports relationnels de nos compétences sociales, de nos capacités à établir et à maintenir des relations en étant le plus à l'aise possible

dans leurs contingences. Les robots possèdent de multiples fonctions mais l'une d'entre elles nous concerne plus particulièrement, c'est celle de la robotique sociale, donc affective qui consiste à travailler sur des émotions et les compétences sociales. Les robots sociaux sont conçus comme des interfaces qui emploient des clés de communication sociale selon plusieurs modalités. Ils peuvent être conçus pour toutes sortes de rôles : attirer l'attention et la maintenir, faciliter les limitations, ce qui est décisif pour l'apprentissage, stimuler l'attention pour une activité avec une autre personne, etc. Dans le cas de l'autisme, ils permettent aux enfants d'apprendre à la fois à exprimer leurs émotions et à reconnaître celles d'autrui, le robot a un grand avantage pour les enfants autistes, c'est qu'il les rassure grâce à des relations claires, prévisibles dans lesquelles il n'y a ni ambiguïté ni malentendu possible.

Citons l'exemple de l'atelier Rob'Autisme au CHU de Nantes qui travaille avec des adolescents atteints de troubles du spectre autistique. La rencontre entre la robotique, la création et la théorie comportementale a permis à ces adolescents de programmer un robot, le robot NAO, pour créer une histoire. Le robot fonctionne ici comme un personnage de fiction que les participants font s'exprimer et le but est d'améliorer les capacités de communication et de socialisation de ces adolescents. Ils travaillent en binôme, ils sont accompagnés d'un soignant pour programmer le robot, ce qui implique de s'ajuster l'un à l'autre et donc à toutes sortes d'expérimentations relationnelles. Il faut aussi demander l'aide technique de la roboticienne présente, tout cela déclenche donc des interactions multiples entre adolescents, entre adolescents et professionnels et avec le robot. Les adolescents sont mis là dans la situation d'être les acteurs d'une histoire qui est élaborée par eux-mêmes et jouée par le robot. Cet outil a été véritablement conçu comme un travail sur ce qui permet d'agir avec les autres. Reconnaissance des émotions et compétences sociales sont au centre de ce travail thérapeutique. De plus, comme l'écrivent les auteurs d'une étude sur Rob'Autisme, les familles ont pu découvrir chez leurs enfants « des potentialités, des capacités, des modes d'être qu'elles ne soupçonnaient pas ». Nous sommes en plein dans l'idéal du « potentiel caché ».

On voit parfaitement l'intérêt d'un non-humain pour inscrire dans la société des personnes qui ont des difficultés à trouver leur place. De ce point de vue, les robots fonctionnent comme des substituts ou comme des interlocuteurs et non comme des agents exécutifs que seraient des drones, par exemple. Ils sont donc potentiellement des acteurs sociaux, ce qui veut dire qu'on les retrouvera dans toutes sortes de domaines où il s'agit de susciter des réactions affectives, de l'empathie ou de la théorie de l'esprit.

Thérapies assistées par ordinateurs, programmes de modification des biais cognitifs, assistants personnels de communication (Siri), diagnostics effectués par des ordinateurs, surveillance en continu des variations d'états du patient, etc., la digitalisation des relations est aujourd'hui un domaine en expansion, qu'il s'agisse de traiter des pathologies comme l'anxiété et la dépression ou de s'exercer à mieux contrôler ses émotions avant un entretien d'embauche. Ces programmes, ces technologies sont essentiellement utilisables dans les approches cognitives et comportementales parce qu'elles consistent en des entraînements et des exercices. Ce n'est pas applicable pour des cures analytiques, très clairement. On a là affaire à un ensemble de technologies partenaires qui concrétisent un vieux rêve de la psychologie cognitive qui est de distribuer la psychologie, c'est-à-dire de faire en sorte que chacun puisse disposer en ses propres mains de l'instrument thérapeutique, dans une sorte d'auto-thérapie ou d'auto-accompagnement. Désormais, on peut dire que la thérapie et l'accompagnement sont non seulement en train de devenir accessibles partout, tout le temps et sur n'importe quel support, mais encore qu'elles peuvent avoir lieu sans aucune référence à un soutien humain. En passant du statut d'intermédiaire à celui de partenaire, la machine se trouve en position de *coach* et ses techniques offrent d'ailleurs une alternative pour ceux qui répondent « non » à la question : « Voulez-vous parler de vos problèmes ? ». Il y aurait là un avantage pour le traitement de certaines populations comme les militaires qui parce qu'ils doivent se montrer « durs » évitent d'aller « chez le psy ».

C'est tout un imaginaire capacitaire qui s'est développé dans une floraison d'accompagnements en ligne, organisés autour de l'augmentation par les connexions, assistés par la plateforme, l'objet connecté et la donnée. Nous avons affaire à des programmes et des machines fondées sur l'idée que les individus peuvent, voire doivent s'aider eux-mêmes, augmenter leur pouvoir d'agir.

Antoinette Rouvroy : Je parlerai pour ma part de la montée en puissance des « *data sciences* », ou du « tournant algorithmique » qui se fait ressentir dans une part croissante des secteurs d'activité et de gouvernement. L'engouement contemporain pour les nouveaux dispositifs algorithmiques de modélisation du social et de profilage des individus n'est pas tant la cause que le symptôme d'une nouvelle manière de (nous) gouverner ou plutôt d'une nouvelle manière de nous dispenser de nous gouverner nous-mêmes. Rendues possibles et nécessaires par la disponibilité de fait de quantités massives de données numériques proliférant continuellement et à grande vitesse des comportements et interactions des individus, les nouvelles techniques algorithmiques permettent d'automatiser et d'accélérer la transformation des données « brutes » en informations opérationnelles. Visant à caractériser anticipativement les comportements possibles, le profilage algorithmique

vise à optimiser en temps réel les interactions avec les individus, promettant à la fois aux individus et aux bureaucraties publiques et privées de les dispenser à la fois d'avoir à se confronter à la part incompressible d'incertitude radicale (l'excès du possible sur le probable) dont est grosse la matérialité du monde, et d'avoir à assumer la tâche – traumatisante à bien des égards – d'interpréter, décider, trancher – bref, de (nous) gouverner sur fond d'incertitude.

Il nous est impossible de nous représenter la « couche » proliférante de données numériques. Imaginez : d'après certains calculs produits par IBM, au cours d'une vie, un individu « normalement connecté » devrait produire plus d'un million de gigabits de données. Chaque donnée, isolément, ne représente rien : sa valeur, son utilité, sa signification ne dépend aucunement de sa densité en information (qui peut être très faible) mais seulement de la possibilité qu'elle contribue, étant mise en corrélation avec d'autres données, à l'émergence de « patterns » ou « modèles » statistiquement significatifs, donc prédictifs. Les données numériques sont, en quelque sorte, arrachées à leur fonction de signifiants : ce sont de purs signaux asignifiants mais calculables qui contribuent à former ensemble une sorte de carte absolument plastique en expansion, laquelle produirait son propre territoire spéculatif. Je dis « territoire spéculatif » car les modélisations et « prédictions » produites par les algorithmes nourris de ces données massives ne sont aucunement descriptives d'un état de fait (les modélisations algorithmiques ne sont ni vraies ni fausses mais seulement suffisamment fiables pour justifier des stratégies d'action), elles ne disent rien de ce qui est, elles ne visent que le « possible » – la somme des opportunités, des propensions, des risques, statistiquement détectables à même les données numériques. Le « rêve » qu'incarnent les projets de gouvernementalité algorithmique, c'est que le monde se gouverne lui-même à travers les données qui en émanent et s'en détachent comme une empreinte déterritorialisée dispensant de toute représentation. À la crise générale de la représentation, l'idéologie technique des données massives apporte une réponse radicale : il n'y a plus rien à représenter, tout est toujours déjà dans les données, accessible en très haute définition, en temps réel.

La disponibilité de ces quantités massives de données numériques rend non seulement possible mais nécessaires de nouvelles techniques d'analyse, de nouvelles logiques de traitement, capables d'automatiser et d'accélérer la transformation de ces données « brutes » en information opérationnelle. Alors que dans les pratiques statistiques « classiques », on convient d'abord de catégories statistiques, qui, ensuite, orientent et bornent la collecte de données, dans le contexte des données massives le processus est exactement inverse : les catégories, hypothèses ou attributs ne sont pas décidés d'avance, ils ne président ni ne bornent la récolte des données, mais résultent ou surgissent de leur traitement. Donc, dans cette logique purement inductive, on ne fait plus tellement d'hypothèses à propos des causes des phénomènes pour ne s'intéresser qu'à la détection anticipative – spéculative – des effets. Ce désintérêt pour les causes accélère la production de modélisations opérationnelles. Il y aurait même une sorte de "*trade-off*" ou de compromis à faire entre l'explicabilité de la modélisation et sa fiabilité, sa finesse : rendre le résultat du calcul algorithmique explicable présupposerait d'opérer a priori une sélection dans les données : c'est-à-dire de revenir vers des formes de pratiques statistiques plus « classiques », ce qui ne pourrait se faire qu'au détriment de la « fiabilité » ou de la « finesse » des modélisations algorithmiques, capables de tenir compte, précisément, de données qui auraient été exclues des traitements statistiques classiques. L'effet « haute définition » de cette approche par les données massives, s'il permet de détecter des régularités du monde qui ne sont observables que sur les très grands nombres, s'il permet dès lors de faire droit, de façon inédite, à la complexité des phénomènes, se paie cependant d'une série de risques épistémiques qui n'ont rien de trivial. L'exemple le plus évident – à côté des biais préexistants dans les données (qui ne reflètent pas tant les « faits » que les « effets » de causes dont on se désintéresse), des biais de programmation (la détermination de la fonction objective, c'est-à-dire de ce que chaque algorithme doit optimiser, la fixation initiale des « métriques », etc.), des biais d'apprentissage (privilégiant l'obtention d'une solution non équivoque quitte à négliger les risques de « faux positifs » ou de « faux négatifs ») – est celui des *spurious correlations*. Chacun sait que « corrélation n'est pas causalité ». Une corrélation entre un signal A et un signal B peut signifier que A cause B, ou que B cause A, ou que C, que l'on n'a pas vu, cause A et B, ou encore que la corrélation détectée entre A et B est le pur effet du hasard. Il est donc impossible d'interpréter ces résultats. Plus les quantités de données augmentent, plus grandes sont, statistiquement, les « chances » de découvrir des corrélations qui ne correspondent strictement à rien dans le monde physique, n'étant là que par l'effet du hasard. Fonder des décisions, dans quelque secteur d'activité ou de gouvernement que ce soit, sur des recommandations algorithmiques découlant de ces corrélations « aléatoires » équivaudrait à s'en remettre purement et simplement au hasard tout en ayant l'impression d'avoir rationalisé le processus décisionnel en l'automatisant au moins en partie. Il convient donc de problématiser l'engouement contemporain pour l'« intelligence des données ».

L'hypothèse que je voudrais mettre en discussion est que l'engouement contemporain pour des formes de gouvernementalité algorithmique du monde social – c'est-à-dire pour une nouvelle manière de gouverner qui n'en passe plus par aucune catégorie statistique ou aucune catégorie socialement éprouvée,

mais par des profilages de plus en plus nombreux, évolutifs en temps réel, s'ajustant aux comportements de chacun sans plus les rapporter à aucune norme commune, mais en les évaluant à l'aune de métriques hyper-mobiles articulées aux comportements de tous les autres –, correspondrait assez exactement à un certain idéal de l'individu occidental contemporain « émancipé » de toute assignation identitaire fixe. La gouvernementalité algorithmique est le mode de gouvernement qui répondrait parfaitement à la « haine de la moyenne » que partagent les hyper-(in)dividus occidentaux. L'hyper-(in)dividu contemporain ne veut plus que les bureaucraties privées ou publiques s'adressent à lui en tant qu'éléments d'une catégorie statistique : « Je refuse que l'on s'adresse à moi comme la ménagère de moins de 50 ans, je suis exceptionnelle et j'attends que l'on s'adresse à moi en tant qu'être exceptionnel et que l'on s'adresse à moi dans ma singularité ». Or, c'est précisément ce que promet la gouvernementalité algorithmique. Dans la mesure où contrairement à ce qui se passe dans les pratiques statistiques classiques, on tient compte de toutes les données disponibles, y compris celles qui sont les plus éloignées de la moyenne, et dans la mesure où les « profils » ne précèdent ni ne bornent pas la récolte des données (comme les anciennes catégories statistiques), mais résultent de leur analyse, les nouveaux traitements algorithmiques permettent une hyper-personnalisation des interactions qui n'en passe plus par la rencontre des personnes individuelles mais par des traitements de données massives à l'échelle industrielle. À cet engouement pour la « personnalisation » ou pour ce gouvernement en « haute définition » qui implique aussi une forme inédite de surveillance de masse, s'articule, sur le plan politique, l'obsession de faire payer à chacun son « coût réel ». Donc, c'est là vraiment une nouvelle stratégie de gestion de l'incertitude qui permettrait de ne plus en passer par aucune solidarité sociale, par aucune mutualisation des risques... à la limite on pourrait évacuer même les assureurs puisque l'assureur a précisément pour métier de socialiser les risques, c'est-à-dire de fabriquer des populations, qui sont des catégories statistiques. Ici, on semble glisser dans une société post-actuarielle dans laquelle on n'aurait plus besoin d'assurance, chacun paierait pour son risque réel qui ne serait même plus un risque puisque le risque ne se calcule par définition que sur des populations. Dans cette société post-actuarielle, les primes d'assurance pourraient évoluer, il y aurait une sorte de fluctuations en temps réel, en fonction du comportement journalier de chacun. Évidemment, tout cela produit un certain type de subjectivité : des effets d'incitation qui peuvent, suivant les contextes, être favorables (l'évaluation de la prime d'assurance automobile en fonction du comportement de l'automobiliste enregistré et analysé en temps réel peut inciter à une conduite prudente et à des comportements vertueux), mais cette « société de notation » – ou l'évaluation du « crédit » de chacun en temps réel sans passer par aucune catégorie actuarielle antécédente – peut aussi produire des effets paradoxaux.

Premièrement, cette appréhension des comportements et propensions humaines en « haute définition » et en temps réel, présuppose une certaine tolérance pour ce qui relève à l'évidence de la surveillance de masse. Voilà qui est assez contradictoire avec les idéaux de liberté, d'autonomie, d'autodétermination qui, prétendument, motivent l'engouement pour la « personnalisation » des interactions et la mise à charge de chacun de son « coût réel ».

Deuxièmement, alors que les pratiques algorithmiques se parent d'une aura sinon d'objectivité, au moins d'impartialité, leurs effets ne sont rien moins qu'incompatibles avec les principes de non-discrimination et d'égalité d'opportunités. Si l'on se fonde sur le profil de consommateur, sur le type de grands magasins que vous fréquentez pour détecter votre risque, etc., à ce moment-là les personnes les moins fortunées, celles qui fréquentent les supermarchés dans lesquels il n'y a pas de rayon « bio », les supermarchés qui sont fréquentés aussi par des personnes qui ont des difficultés à rembourser leurs dettes, vont être moins bien « notées » par les algorithmes d'évaluation des risques de crédit, par exemple. Une telle gouvernementalité algorithmique aura plus probablement pour effet d'intensifier les disparités de moyens, de ressources d'opportunités, de bien-être au détriment de certains groupes qui cumuleront les désavantages « en cascade » que d'égaliser les conditions.

Troisièmement, comme les autres types de gouvernementalité, la gouvernementalité algorithmique « produit » les sujets qui lui sont adéquats. Le fait d'attribuer à des éléments infimes du quotidien, en eux-mêmes asignifiants, une valeur spéculative, d'en faire des signaux calculables, nourrit une nouvelle forme de capitalisme numérique (dans lequel l'objet d'accumulation, c'est la donnée) en même temps qu'une nouvelle forme de normativité. Dans l'exemple du « *quantified self* », c'est-à-dire l'auto-mesure de soi, l'auto-quantification de soi, donc l'élaboration d'un soi commensurable car chiffré, on participe nous-mêmes très volontiers à cette mise en nombre. Quantifier, expliquait Alain Desrosières, c'est exprimer ou faire exprimer sous une forme numérique, sous une forme de nombre, ce qui n'était exprimé antérieurement que par des mots. Cette traduction n'est possible qu'à condition qu'existent antérieurement des conventions d'équivalence, donc qu'il y ait eu des discussions, discussions qui sont évidemment court-circuitées par les processus algorithmiques. Le « *quantified self* » qu'est-ce que c'est ? C'est de la quantification continue sans convention de quantification préalablement établie, le temps réel et l'évolutivité dispensant de la convention antérieure tout en contribuant à la production sociale de normes de comportement, de performances, de santé éminemment

évolutives : les personnes ne vont plus être évaluées à l'aune de métriques stables comme une idée générale de normalité, un seuil de fonctionnement normal dans la société par exemple mais à l'aune de métriques qui deviennent hyper mobiles et qui dépendent du comportement de tous les autres puisque le « *quantified self* », mis en relation/compétition via Internet avec toutes les autres personnes qui se quantifient, peuvent constamment à la fois mesurer leurs progrès et se mesurer aux autres qui, eux aussi, essayent de progresser. L'on a affaire à une norme ou à une normalité qui s'échappe : plus on en approche, plus elle s'éloigne. Voilà qui rejoint un peu l'idée du perfectionnisme, « on n'est jamais assez normaux », d'où la « normopathie » dont parle Guillaume Le Blanc, dont il dit que « c'est une situation dans laquelle le langage, la pensée, les comportements normés en vue de la performance et de l'efficacité perdent tout pouvoir de contestation dès lors que la vie elle-même deviendrait un programme lui-même intégré à celui d'une immense machinerie acéphale ». La construction dynamique, interactive, en réseau, de cette normativité perfectionniste correspond parfaitement aux idéaux les plus « à la mode » de démocratisation, d'horizontalisation, de dispositifs « centrés sur la personne ». Hyper-individualisation et immanentisation de la norme correspondent tellement bien à l'esprit de l'époque qu'il est mal aisé de trouver des arguments pour s'y opposer. Quelles critiques peut-on adresser à ce genre de dispositifs dans la mesure où ce sont les pratiques des destinataires de la norme qui contribuent à redéfinir continuellement les objectifs de performance et de jouissance les inscrivant dans des processus de perfectionnement infini dont l'objectif recule au fur et à mesure que les individus progressent ? Cette microgestion algorithmique de la santé, du bien-être, de la productivité rend chaque individu entrepreneur de lui-même, maximisateur de son capital humain numérisé, ce qui ne veut pas dire nécessairement en bonne santé. Par exemple, afin de gagner des points au regard de son assureur sans faire l'effort d'aller courir soi-même, on peut imaginer que certains assurés sous-traitent leur course à pied quotidienne à des tâcherons équipés de leur bracelet connecté, payés à la performance, entretenant ainsi le capital humain numérisé de l'assuré, et améliorant son « crédit » assuranciel sans pour autant améliorer son état de santé. Les individus deviennent donc entrepreneurs non pas tant d'eux-mêmes que de leur capital humain numérisé, ils deviennent responsables non seulement de leur destin psychologique et psychique mais encore d'envoyer des signaux numériques qui leur valent de bonnes cotes de crédit, quitte à ruser, en misant sur la distinction entre identité physique et identité numérique. Ce faisant cette focalisation sur la micro gestion individuelle de la santé distrait la tension des causes environnementales et socio-économiques des problèmes de santé publique. Or, le design technologique pourrait, ou devrait à mon avis, en tout cas aussi faciliter, plutôt que le perfectionnisme sanitaire individuel, la délibération collective sur les déterminants non seulement comportementaux mais aussi environnementaux et socio-économiques de la santé et du bien-être. On est dans un paradoxe dans lequel finalement on a cette idée que les individus deviennent éminemment plastiques, que leur destin biologique et psychique est entre leurs mains comme l'a dit Alain Ehrenberg alors même que les structures socio-économiques sont vécues comme des faits de nature et donc inchangeables et indiscutables. Enfin, ce qui est paradoxal aussi c'est que l'on insiste beaucoup sur l'autonomie individuelle d'un côté mais en même temps Richard Thaler a gagné le prix Nobel d'économie pour sa « théorie du *nudge* ». Dans une posture de paternalisme libertarien, sans en passer par le prisme de la conscience des individus, il s'agit de façonner leurs environnements physiques et informationnels (les « architectures du choix ») de manière à orienter, de façon subliminale, sur le mode du réflex, leurs comportements. N'est-il pas paradoxal qu'à la fois on valorise le « *nudge* », c'est-à-dire l'orientation des comportements à un stade préconscient, la manipulation « pour le bien » et qu'en même temps on ne cesse de proclamer comme valeur suprême l'autonomie, l'auto-détermination d'un individu qui n'est même plus saisi ni pensé en tant que sujet d'énonciation de ses motivations, ni en tant qu'il s'inscrit dans des contextes collectifs, puisque l'on ne parle plus ni de sujet, ni de catégorie socialement éprouvée, mais d'individus, voire de dividus, en fait réduits à des amas de pulsions ?

Co-présidents : Merci pour ces deux exposés qui nous font un peu frémir Il faut peut-être faire la différence entre l'intelligence artificielle et la digitalisation de la vie fantasmées et l'intelligence artificielle et la digitalisation réelles. Ce que vous dîtes parfois de l'intelligence artificielle, et qui est aussi développé par certains autres auteurs, est loin de ses performances réelles. D'ailleurs, en médecine, en ce qui concerne l'analyse d'images, il est clair qu'une machine est bien plus performante tout simplement parce qu'elle analyse pixel par pixel, ce que ne peut pas faire l'œil humain. Mais, pour d'autres domaines, on verra si l'intelligence artificielle est supérieure à l'être humain. Les patients auront besoin de la validation par l'être humain parce que l'IA peut être facilement trompée. Si vous lui apprenez à reconnaître des chiens et à les différencier des chats, il suffit de deux ou trois petites manipulations pour l'induire en erreur. Tout ce que vous dîtes sur la différence entre les corrélations et les causes, paraît très important. Vous connaissez ce fameux vers de Virgile dans l'Énéide : « Bienheureux celui qui connaît la cause des choses. » La recherche des causes en médecine, c'est la quête du Graal, et en psychiatrie on a encore beaucoup de travail à faire. L'attention attirée sur les corrélations doit se compléter par un travail de physiopathologie et de recherche des

causes, ce n'est pas forcément antinomique. Confondre cause et corrélation n'est pas le seul fait de l'intelligence artificielle. L'intelligence humaine dans ce domaine est très douée.

Antoinette Rouvroy : Je n'ai aucune prétention à dire que l'intelligence artificielle ne fait pas des choses magnifiques et je n'ai aucune prétention à la comparer à l'intelligence humaine. Elles font des choses très différentes, qui n'ont à la limite rien à voir, et qui, pour cette raison, peuvent être complémentaires. Donc dire que l'intelligence artificielle va devenir supérieure à l'intelligence humaine me semble une aberration. L'intelligence artificielle peut synthétiser, à très vive allure, de la diversité ; elle peut détecter des régularités du monde qui ne sont repérables que dans des données massives, diverses, complexes. Mais l'intelligence artificielle est incapable de donner du sens au résultat de cette synthèse, d'interpréter. Le problème, c'est que dans un certain nombre de dispositifs – et ce n'est pas du tout de la science-fiction –, que ce soit dans le domaine du marketing ou de la sécurité, peut-être moins dans celui de la santé parce que là on sait qu'il faut interpréter sinon on risque véritablement des catastrophes mais lorsque la catastrophe est moins directement visible ou lorsqu'elle est géographiquement éloignée de notre territoire et bien on se permet de tuer sur base de corrélations détectées dans des métadonnées. Ça a été dit, au cours d'un symposium à Johns Hopkins University en 2014 par le général Hayden, ancien directeur de la NSA et de la CIA : "*We kill people based on metadata.*" Cette nouvelle stratégie « préemptive » – spéculative (on agit par avance sur base de spéculations algorithmiques) plutôt que préventive (on se désintéresse des causes) – de gestion de l'incertitude est aujourd'hui à l'œuvre dans nombre de secteurs. Je ne crois pas que ce soit de la science-fiction, je pense même que nous sommes en retard dans la description de ce qui est en train de se développer. Quant à la distinction entre corrélation et cause, elle est très importante, elle est très importante dans la mesure où, précisément, parce que l'opérationnalité devient le critère prépondérant d'évaluation de la félicité des dispositifs, la tendance est à se satisfaire de la corrélation sans rechercher des causes, ce qui peut aussi empêcher, dans une certaine mesure, de faire de la prévention.

Alain Ehrenberg : Je crois qu'il y a un contraste entre nos deux exposés. Il y a un contraste sur le gouvernement et sur le gouvernement de soi, c'est-à-dire je vois plutôt dans ces machines une sorte d'extension pour des exercices de l'autonomie, cela apparaît clairement avec la robotique sociale, et Antoinette Rouvroy voit plutôt un gouvernement de surveillance. Ce n'est pas contradictoire : ça dépend à quel niveau on étudie ces machines, sous quels angles ? Par ailleurs, au sujet du *nudge* et de l'économie comportementale, oui, c'est très très clair, comme c'est écrit d'ailleurs dans un rapport britannique sur les politiques du *nudge* : il s'agit de changer le comportement sans changer l'esprit, *changing behaviour without changing mind*. Donc, on a affaire à des pratiques qui évitent, un peu comme l'a dit Antoinette Rouvroy, de se gouverner soi-même. Là, il s'agit d'économiser de la décision, de la volonté, *etc*. En tout cas, il y a là une double tendance contradictoire : permettre le gouvernement de soi, éviter le gouvernement de soi. Alors, est-ce que c'est de l'aliénation ?

Antoinette Rouvroy : Pas nécessairement.

Alain Ehrenberg : Tout à fait : pas nécessairement. Le grand principe du *nudge* c'est ce que l'on appelle l'option « défaut » : c'est que lorsque l'on a pris un abonnement à une revue, il faut décider d'arrêter l'abonnement. Le *nudge* fonctionne là-dessus : il faut décider pour ne pas avoir le comportement adéquat. Avoir le comportement adéquat ne passe pas par une décision, on se laisse prendre par le système. Par ailleurs, je me demandais en vous entendant, si l'on a des travaux sur les usages des intelligences artificielles, les usages des données qui permettent d'avancer empiriquement sur ces questions-là et sur les questions d'aliénation ou d'émancipation : parce qu'il y a toujours des tensions entre les deux aspects.

Antoinette Rouvroy : Sur les usages, je pense notamment aux travaux de Dominique Boullier, de Dominique Cardon, d'Antonio Casilli qui sont sociologues. Par ailleurs nous avons un projet, avec le laboratoire de cyberjustice de l'université de Montréal, intitulé « autonomisation des acteurs de la justice par l'intelligence artificielle », dans lequel nous étudions toute une série d'applications, donc d'usages, de l'intelligence artificielle à tous les stades de l'administration de la justice, de la prévention des conflits et infractions à leur résolution judiciaire. Je suis d'accord avec vous aussi, concernant cette architecture du choix qui est « proposé » dans le *nudge*. Le *nudge* inverse la définition que Musil – dont on connaît les penchants déterministes – donnait de la liberté : le sentiment de faire volontairement ce qu'on veut involontairement. Le *nudge*, n'est-ce pas la prétention d'inciter les individus à faire involontairement ce qu'ils sont censés vouloir volontairement ? Le *nudge* est, à mon avis, beaucoup plus paternaliste que libertarien ! Il prive les individus de la possibilité de produire par eux-mêmes, y compris après-coup, la motivation de leurs actes, de (se) rendre compte de ce qui les fait agir, c'est-à-dire d'agir en sujets de droit.

Alain Ehrenberg : Je pense que c'est un monde de contrastes. D'une part, il y a toute cette tendance autour du *nudge* et de l'économie comportementale, de l'automatisation disons des comportements par un système d'incitations qui est l'option par défaut. Mais de l'autre côté, je prends l'exemple des robots, où c'est exactement l'inverse. Il y a tout un travail sur l'esprit. Il y a tout un travail de socialisation, d'augmentation

des compétences sociales, de reconnaissance des émotions pour des individus qui ont beaucoup de mal à se débrouiller avec ça. On a deux pôles.

Antoinette Rouvroy : Mais avec des usages très divers, ça dépend des usages.

Alain Erhenberg : ça dépend des usages, ça dépend aussi des programmes eux-mêmes bien entendu, mais je pense que l'on a affaire à un monde complexe et de contrastes qui vont dans les sens les plus opposés.

Antoinette Rouvroy : Tout à fait, ça se complète avec une phrase d'un des patrons de Google qui a dit que bientôt il sera impossible à n'importe quel individu de vouloir quelque chose qui n'a pas été prévu pour lui. Amazon a breveté un dispositif qui permet d'envoyer par avance des colis qui n'ont pas encore été commandés. Que « ça marche » ou pas n'est pas ici l'enjeu : à travers ces prétentions-là se perçoit l'idéologie technique à l'œuvre et suivant laquelle *tout serait dans les données*, pas seulement le passé, mais aussi l'avenir.

Co-présidents : *Un philosophe des sciences disait « qu'à force de mesurer, on croit que l'on mesure quelque chose ». Ceci étant, la médecine a toujours été personnalisée mais n'a jamais été parfaite et je pense que ces grandes données ne vont pas forcément donner réponse à tout.*

Antoinette Rouvroy : Oui, effectivement, on passe d'ambivalence à ambivalence mais il me semble juste précisément que l'usage des big data et des applications de type de personnalisation comme le « *quantified self* », *etc.*, font un peu le pont finalement entre ces deux visions opposées : personnalisation et *evidence based medicine*. En fait, ce n'est probablement pas encore le cas, mais on peut imaginer grâce à l'approche très fine, en très haute définition que permettent les big data à l'échelle de la donnée infra-personnelle et par l'effet de *feed-back loop* qui se produit dans les machines apprenantes, qu'on va enrichir l'*evidence based medecine* de tous les cas individuels et qu'on va pouvoir si vous voulez fragmenter en une myriade de... je me demande si on peut dire de maladies. Je me demande si on va encore pouvoir individualiser des maladies, si on va pouvoir nommer des maladies parce qu'à la limite, chaque personne va avoir des propensions tellement diverses que, à la limite, je ne sais pas si je vais rester un objet résilient comme une étiquette qui dirait un diagnostic.

Public : Cette vision du dispositif technologique sous forme de gouvernementalité algorithmique[1] presque omnipotente qui me fait penser à ce vers de Hölderlin : « *Là où croît le danger, croît aussi ce qui sauve* ». Où est « ce qui sauve » ? De plus il me semble que l'on est dans une vision ontologisante de ces technologies, ce qui est tout à fait intéressant, mais pour la médecine j'ai l'impression qu'il faut quand même concéder une part de pratiques théoriques singulières et considérer que la médecine n'est pas juste le traitement des big data de santé. Sinon on pourrait se dire que Twitter, Google avec leurs données (avec Google Flu par exemple) fondent la santé, mais est-ce que ça sert vraiment à quelque chose ? Du coup, ce qui m'interroge c'est cette idée qu'il y aurait une prise de décision directe des machines, c'est là où je conçois qu'il y a une limite et peut-être ce qui sauve. Est-ce que l'aspect de prise de décision indirecte, c'est-à-dire l'aspect documentaire du big data, de l'*evidence based medecine*, ne doit pas justement être limité, circonscrit à une activité, une pratique théorique précise comme celle de la médecine ou celle du droit par exemple ? Est-ce que ce n'est pas là que l'on peut faire jouer quelque chose qui permette de sortir de cette aliénation ?

Antoinette Rouvroy : Oui, j'aurais dû le préciser d'emblée : pour Foucault ces différents modes de gouvernement ne se succèdent pas dans le temps, mais se superposent. De même, ce que j'ai décrit ici, ce n'est qu'une couche de gouvernementalité qui se superpose et s'articule avec d'autres types de gouvernementalité qui continuent à exister bien entendu, la difficulté étant d'articuler entre elles, de percevoir que les liens « tiennent ensemble » tout ce « feuilleté » de gouvernementalités. J'ai évoqué une certaine forme de « complicité » ou de rapport de co-détermination ou de renforcement mutuel entre gouvernementalité algorithmique et gouvernementalité néolibérale. Concernant la prise de décision algorithmique, c'est une question difficile. Il y a tout un courant de designers et d'informaticiens travaillant à développer du *fair, accountable, transparent machine learning*. Le problème évidemment c'est que l'on ne considère pas l'interaction hommes machines, or nous continuons à exister malgré tout, et face à une recommandation algorithmique, que ce soit un juge (ou un médecin), le juge va parfois obéir à l'algorithme, parfois pas. Certains juges ignoreront systématiquement la recommandation algorithmique, d'autres s'y conformeront systématiquement, d'autres encore – et peut-être seront-ils les plus nombreux – tantôt l'ignoreront, tantôt s'y conformeront. Il faudrait pouvoir modéliser anticipativement la réaction de l'humain face à la décision machinique, mais c'est extrêmement compliqué. Si on est dans une entreprise par exemple ou dans une administration où les employés sont encouragés à prendre des initiatives et ne sont pas trop durement sanctionnés lorsque, de l'initiative qu'ils prennent, résulte un dommage ou un échec, et bien effectivement ils vont peut-être plus volontiers s'écarter de la volonté algorithmique. Lorsque c'est l'inverse, franchement, ils ne vont pas le faire parce qu'ils préfèrent reporter la responsabilité sur le dispositif. C'est une question qui est insoluble sans rentrer dans les pratiques ou les usages.

[1] Le terme est d'Antoinette Rouvroy

Pour ce qui est du salut, puisqu'il s'agirait de sauver le monde, comme je le disais, moi je ne suis pas du tout opposée aux algorithmes, tout ça ce sont des machines absolument passionnantes et fascinantes et qui pourraient bien contribuer à nous rendre bien plus intelligents que ce que nous sommes (!). Il y a des manières d'utiliser ces machines qui augmentent l'intelligence humaine en nous rendant sensibles à des régularités du monde que l'on n'aurait pas vues. Par contre, elles sont toxiques si elles sont utilisées afin d'optimiser des états de fait insoutenables ou, à maximiser les intérêts individuels ou des intérêts de firme qui incompatibles avec l'intérêt commun ou l'intérêt collectif. C'est malheureusement à cette fin qu'ils sont principalement déployés aujourd'hui mais ce n'est pas du tout une fatalité.

Liens d'intérêt les auteurs déclarent ne pas avoir de lien d'intérêt en rapport avec cet article