



HAL
open science

L'hypothèse de la convergence de toute science dans l'ordinateur, par la force de la méthode

Arnaud Billion

► **To cite this version:**

Arnaud Billion. L'hypothèse de la convergence de toute science dans l'ordinateur, par la force de la méthode. Revue CONFLUENCE : Sciences & Humanités, 2024, 6. hal-04756417

HAL Id: hal-04756417

<https://hal.science/hal-04756417v1>

Submitted on 28 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

L'hypothèse de la convergence de toute science dans l'ordinateur, par la force de la méthode

Arnaud BILLION

IBM France Lab

Orsay, France

Résumé

L'article cherche à apporter une explication à la polysémie de la notion de donnée. Il formule, pour ce faire, l'hypothèse de la convergence : sous la force de la méthode scientifique, toutes les sciences s'atomiseraient d'elles-mêmes en informatique, et les divers types de donnée (jusqu'à la donnée numérique) correspondraient à divers stades de cette atomisation. Pour prouver cette hypothèse, la méthode scientifique est d'abord positionnée dans la société plutôt que dans sa pureté idéale. Ainsi concrètement, la science est sujette à une tension prescriptive que l'on reconnaît dans la méthode juridique, ce que manifeste l'histoire du terme « loi naturelle ». Cela posé, on peut s'interroger sur le rôle de la mathématique dans la translation de toute science dans un ordinateur du fait de la méthode. Une déperdition en rationalité est induite. Prenant un pas de recul, on peut enfin relever l'incapacité des critères scientifiques plus évaluatifs à freiner l'accomplissement informatique de toute science.

Discipline : droit de l'informatique

Mots-clefs : données, épistémologie, loi naturelle, méthode scientifique, ordinateur, prescriptivisme

Title: The force of the scientific method: The hypothesis of the convergence and computerization of all branches of science

Abstract

This paper aims to shed light on the polysemy of the concept of data. To this end, it formulates the "convergence" hypothesis: under the influence of the scientific method, it is assumed that all the sciences will atomize into data, and the various types of data (up to and including digital data) will correspond to different stages of this atomization. To prove this hypothesis, the author first situates the scientific method in society, rather than in its idealized form. In concrete terms, this method is viewed as an institution subject to prescriptive tension, just like law, as demonstrated by the history of the term "natural law." With this in mind, the article examines the role of mathematics in the translation of all branches of science into a computational system on account of the scientific method. It underlines the loss of rationality that this process entails. Finally, taking a step back, the article notes the inability of more evaluative scientific criteria to halt the complete computerization of all the sciences.

Discipline: cyber law

Keywords: data, epistemology, natural law, scientific method, computer, prescriptivism

Abstract

La pubblicazione cerca di spiegare la polisemia del concetto di dato. Per farlo, formula l'ipotesi della convergenza: sotto la forza del metodo scientifico, tutte le scienze si atomizzerebbero nell'informatica, e i vari tipi di dati (fino ai dati digitali) corrisponderebbero a diversi stadi di questa atomizzazione. Per dimostrare questa ipotesi, l'autore posiziona innanzitutto il metodo scientifico nella società, piuttosto che nella sua purezza ideale. In concreto, quindi, il metodo è un'istituzione soggetta a tensioni prescrittive, proprio come la legge, come dimostra la storia del termine "legge naturale". In questa prospettiva, si interroga sul ruolo della matematica nella traduzione di tutta la scienza in un computer come risultato del Metodo. Sottolinea la perdita di razionalità che si è verificata. Infine, facendo un passo indietro, sottolinea l'incapacità dei criteri scientifici più valutativi di fermare l'informatizzazione di tutta la scienza.

Disciplina : legislazione sull'informatica

Parole chiave : dati, epistemologia, legge naturale, metodo scientifico, computer, prescrittivismo

.
. .

Le GIEC, organe de l'ONU, est chargé d'examiner et évaluer les « données » scientifiques, techniques et socio-économiques sur le climat¹. Par le biais de traitements statistiques et de transformations, le même GIEC produit encore d'autres « données » sensément utilisées dans le cadre des politiques publiques de transition. Quelle qu'en soit l'influence, de telles données ne semblent pas être de ces sucs nourriciers de notre ère technologique, les fameuses « données-pétrole du XXI^e siècle ». Alors que les règlements de l'Union européenne élargissent toujours davantage la définition de la donnée, il est temps, dans l'optique d'éclairer les raisons de cette fâcheuse polysémie, de nous mettre à l'école informatique, pour comprendre ce dont la donnée est le nom.

La donnée informatique ne porte en principe pas de signification, ce qui l'éloigne des données au sens commun du terme. Elle est couramment associée à un excès hubristique lié à une erreur épistémique relevant du capitalisme (Rouvroy, 2014), ou à une révolution scientifique (Cattaruzza, 2019)... En somme, elle symboliserait un dévoiement de la rationalité scientifique (Lassègue, 1996). Ce serait alors sous l'influence de forces extrinsèques à leur méthode (forces telles que le marché ou le capitalisme) que les sciences s'informatiseraient à l'excès. Ainsi faudrait-il expliquer le passage de la recherche de causation à la généralisation de corrélations, manifeste en apprentissage-machine (Ganascia, 2022).

Cependant, la méthode scientifique n'a-t-elle pas partie liée avec ces forces vues comme extrinsèques, au moins du fait du caractère prescrit et prescriptif qui leur est commun... comme autant d'institutions saturées de règles (Billion, 2022) ? Retournant l'hypothèse de la rupture, nous pourrions formuler celle de la « convergence » : les disciplines scientifiques, grandes productrices de données informatiques, convergeraient elles-mêmes en direction et à l'intérieur d'un ordinateur, à cause de leur propre méthode.

La donnée informatique serait alors le niveau de décomposition atomique de l'activité scientifique. Ainsi s'expliquerait la polysémie du terme : c'est qu'on la considère à différents stades de décomposition. De la carte perforée héritée de Jacquard, à la donnée en proie à un calcul non explicable – typique de la dernière vague d'intelligence artificielle connexionniste – en passant par la donnée d'IA symbolique, règle à déduire dans un système expert, il y aurait simplement une différence de degré de « digestion » par la méthode, cette dernière éreintant progressivement toute intelligibilité. L'ordinateur est cette machine qui assure la transformation méthodique de toute science (tout savoir) en pure et non-signifiante énergie informationnelle. Avant l'ordinateur, on peut encore croire s'adonner à une recherche qualitative (Denzin et Lincoln, 2005) pour sauver ou mobiliser, en face du calcul, la rationalité évaluative (Guillermin, 2019). Mais on ne fait que préparer le moment de la programmation.

Nous verrons donc que c'est la méthode scientifique qui pousse la donnée dans l'ordinateur. La méthode se révèle alors, sous les atours d'une hygiène d'investigation rationnelle, technique sociale de computation. L'hypothèse de la convergence expliquerait pourquoi tous les chercheurs,

¹ L'auteur tient à remercier Éric Boillaud, Aurélien Grosdidier et Anaëlle Martin, pour la valeur de leurs corrections et le caractère décisif de leur amitié.

scientifiques et ingénieurs sont plus ou moins occupés à encoder ou paramétrer quelque logiciel : c'est parce que la méthode scientifique le réclame foncièrement.

Pour explorer cette hypothèse, nous devons tout d'abord contextualiser la méthode scientifique dans son indissociabilité aux dynamiques sociales prescriptives. Ensuite, voir que la notion de « loi naturelle » empiète sur « être » et « devoir être », ce que reflète dans l'ordinateur, le logiciel – qui n'est au fond que données. Dans un troisième temps, nous pourrions regarder en face la mathématisation des sciences jusqu'à l'informatique. Nous montrerons alors dans quelle mesure cela aboutit à une perte en rationalité. Dans une cinquième partie, nous verrons que les principes de scientificité moins calculatoires n'empêchent pas la transformation de la science en *data science*, puis en *data*.

La donnée scientifique dans son environnement socio-politique

L'objet de cette partie est d'inviter à penser la méthode scientifique dans la société, sous l'empire des forces par lesquelles cette dernière est agie, plutôt qu'en sa pureté originelle – qu'elle soit ancienne ou moderne, comme détachée de toute influence – ou idéale – sacralisée sous la forme d'un Graal académique vers lequel elle tendrait.

D'abord, on ne peut que constater le pilotage politique de la production scientifique, par l'intermédiaire de programmes de recherche, de son financement public direct, de ses incitations financières. Le cadre « Horizon » de l'Union européenne est exemplaire : l'activité scientifique, même lorsqu'elle est publique ou consiste en de la recherche fondamentale, ne se déploie pas de manière méthodologiquement pure. Or, les liens historiques et philosophiques entre science et politique sont profonds. La « gouvernance par les nombres » est montrée par Supiot (2015). Dans la lignée de la pensée de Michel Foucault, Rouvroy et Berns (2013) critiquent le principe de « gouvernementalité algorithmique » et le relie à la question du pouvoir par la donnée (Rouvroy, 2014). Rey (2017) dévoile la profondeur des racines scientifiques de la gouvernance politique. Pour Rappin (2014), les sciences du management sont cybernétiques. Les leçons de Max Weber (2003 [1922]) sur la bureaucratie accordent une place à la scientificité : le sociologue décrit une machinerie fondée sur la procédure, le formalisme et la mesure, traits communs aux méthodes scientifique, juridique moderne, mathématique et informatique. Dubarle pose, en 1948, la notion de « machine à gouverner » pour désigner l'ordinateur².

Ainsi, la méthode scientifique pourrait être elle-même une politique, une machine à gouverner. L'impact de la méthode sur le politique est bien présent, et l'effet de retour n'est guère douteux : la science veut organiser scientifiquement la société, y compris l'organisation scientifique, ce qui dessine le schéma d'une belle boucle de rétroaction cybernétique. Nous soupçonnons qu'il y a quelque chose de commun au scientifique et au politique. Or, cette influence de longue date de la

² DUBARLE, Dominique. 1948. Vers la machine à gouverner... La manipulation mécanique des réactions humaines créera-t-elle un jour « le meilleur des mondes »? *Le Monde*, 28 décembre. URL : https://www.lemonde.fr/archives/article/1948/12/28/la-manipulation-mecanique-des-reactions-humaines-creera-t-elle-un-jour-le-meilleur-des-mondes-les-premiers-grands-relais-du-cerveau-humain-le-depassement-du-systeme-nerveux-les-pro_3036529_1819218.html

mathématique sur les rapports sociaux, parallèle à la mathématisation des sciences naturelles, plaide en faveur de l'hypothèse de la convergence : elle résulterait de la force d'un principe interne à la méthode scientifique qui ferait avancer partout le nombre, en-dehors-même des laboratoires³. L'avalanche de données serait alors le fruit d'une effectuation politique sous la force de la méthode scientifique. Or, l'activité scientifique comporte une dimension prescriptive. C'est alors dans l'histoire du droit que nous verrons clairement la connaissance interprétable s'atomiser en données dans un « nominalisme dynamique » (Hacking, 2006) commun à l'informatique et au droit des Modernes (Billion, 2022). Pour comprendre ce que partagent les données informatiques, scientifiques et politiques, il faut s'intéresser au donné des juristes, à la science du droit.

À la source du droit et de la science : la loi naturelle

La notion de « loi naturelle⁴ » (norme de comportement indiscutable/fonctionnement spontané du réel) est, tout comme « donnée » (renseignement instructif/élément d'effectuation logicielle), employée pour désigner une chose et son contraire. Nous entrevoyons la faiblesse de catégories plus larges, symptôme de la convergence. Nous montrerons dans ce chapitre les conditions d'une saisie de la méthode scientifique par des mécaniques prescriptives (sa compatibilité logicielle pour ainsi dire).

L'expression « loi naturelle » est d'abord employée par Descartes pour désigner les régularités naturelles qui sont objets de l'investigation scientifique. Montesquieu (2002 [1748]), dans le premier chapitre du premier livre de *L'esprit des Lois*, reprendra cette définition : « Les lois, dans la signification la plus étendue, sont les rapports nécessaires qui dérivent de la nature des choses ». Cependant, Hobbes (2002 [1651]) lui donnait un sens performatif et normatif : un ensemble de principes ou théorèmes propres à assurer la sécurité de l'individu. Le philosophe joue de l'ambivalence de la notion, à cheval sur les disciplines descriptives et prescriptives. L'histoire du droit moderne, c'est cette recherche d'un prescrit plus rationnel, conforme aux lois de la nature, et peut-être à la volonté divine (Villey, 2017).

Ainsi, les développements de la notion de loi naturelle, en particulier dans les milieux savants du XVII^e siècle, incitent à penser qu'il y a quelque chose de commun des deux côtés de la barrière épistémologique : les sciences descriptives-naturelles cherchent des lois pour expliquer le monde, les sciences prescriptives juridico-informatiques pour le régir. Or des deux côtés, on emploie l'outil mathématique. Leibniz élabore scientifiquement des lois entendues au sens juridique (Paksy, 2018), cherchant à faire coïncider droit volontaire et droit naturel en mobilisant l'intelligence mathématique (Griard, 2006). Sa théorie du droit ressemble à la théorie de l'informatique, associant systémique, déductivisme et dichotomie. Les informaticiens ont fini par programmer ses monades. Aujourd'hui le droit (exception faite de la conception du juge) a migré en *legal tech* dans

³ Ceci sans empêcher que la mathématique puisse par ailleurs échapper en elle-même à la machine de Turing pour demeurer un outil d'exploration intellectuelle.

⁴ Nous allons ici consentir à un rapprochement des notions de loi et de droit naturel (ou de nature), considérant qu'elles n'ont jamais été vraiment stabilisées dans l'histoire de la pensée juridique.

la mesure où on y a reconnu de la prescription opérable⁵. L'ordinateur totalement déployé serait une société bien réglée selon la loi. C'est la profonde finalité du jusnaturalisme des modernes (Billion, 2021).

Ainsi, la notion de « loi naturelle » suggère que dire ce qui *est* suppose d'établir comment on *doit* le dire, et que par effet de retour, un nouveau *décrit* se voit entraîné par un autre *prescrit*. L'être et le devoir être ne se laissent alors pas aussi aisément arracher l'un à l'autre (Brague, 2016). La méthode scientifique n'est-elle pas une solution à l'énigme de la loi naturelle, à la fois formulée comme la description juste de ce qu'il faut faire pour bien décrire, et vécue par le scientifique comme une machine prescriptive indiscutable ? Quant à la donnée, elle est au repos (*at rest*) dans l'ordinateur condensé descriptif, et en cours de traitement (lors du *runtime*) machine prescriptive. Et voilà en quoi consiste la programmation logicielle : prescrire des transformations contrôlées à ce que l'on a décrit.

Mathématisation des sciences jusqu'à leur informatisation

Nous pouvons désormais nous demander dans quelle mesure les sciences qui adoptent la méthode pourraient échapper à la computation : c'est remettre en question la place en science, de la mathématique.

Wigner (1960) s'étonnait de la puissance et de l'ambivalence de cette dernière dans la mise en évidence, ou la sélection, des lois naturelles. Simone Weil, selon Lafforgue (2009), pensait que le fait algébrique réduit le monde à des schémas et équations et conduit à se passer de toute discipline de démonstration, de toute pensée. Or, dans un ordinateur, la mathématique est subsumée toute entière sous une algèbre booléenne. La convergence suggère alors un effet d'aspiration de la dimension calculable et programmable des disciplines par le bas, jusqu'au signal binaire. Il faut cependant bien distinguer selon le degré de calculabilité de chaque discipline. Les plus littéraires subissent bien une attraction mathématique : nous avons montré ailleurs que la rationalisation du droit dès le XVII^e siècle est la longue préparation de son informatisation ; en linguistique, Roberto Busa, disposant d'un ordinateur, se mit à recomposer numériquement la théologie de l'Aquinate (Stagnaro, 2017). Mais c'est dans les sciences naturelles, au cœur desquelles la mathématique s'impose de longue date, et qui mobilisent la méthode scientifique la plus pure, que l'attraction informatique se fait irrésistible. Listons les traits de cette méthode qui hâtent l'informatisation de ses objets. Lassègue (2022) insiste sur le lien entre logique mathématique, déduction et programme. La calculabilité est un trait majeur de la méthode scientifique et compte au rang des critères d'informatisation.

La science moderne est friande de méthode quantitative. Au temps de l'intelligence artificielle et des grands modèles, les approches numériques et probabilistes sont centrales dans les sciences naturelles, et l'on utilise des ordinateurs pour les mettre en œuvre. Ces derniers servent aussi à organiser l'activité scientifique en la quantifiant : notation des publications dans des revues scientifiques, classement des universités, pondération de l'influence des chercheurs dans les

⁵ Au moins depuis 1764 (Beccaria, 1991, p. 67) : « En présence de tout délit, le juge doit former un syllogisme parfait [...] rien n'est plus dangereux que l'axiome commun selon lequel il faut consulter l'esprit de la loi. »

réseaux sociaux, politiques d'optimisation des budgets de la recherche. Le nombre se déploie partout dans les sciences, ainsi qu'en informatique.

C'est à mettre en rapport avec la tendance à la formalisation⁶ du langage scientifique et des lois qu'il permet de décrire. La formalisation mathématique renforce la tendance à l'informatisation de la science, puisqu'elle prépare la programmation des équations. Toute avancée de la formalisation dans l'expression d'une loi naturelle préfigure un prochain effort d'informatisation : les ontologies informatiques dans la science du web sémantique (OWL et RDF) sont un moyen de faire communiquer tous les savoirs dans un ordinateur ; les avancées en sciences des matériaux enrichissent bien vite les modèles et logiciels de conception 3D.

Entre formalisation et informatisation se trouve la logique, une logique de plus en plus formelle depuis Newton. Le critère de réfutabilité introduit par Popper (1973 [1934]) s'inscrit dans le paradigme de la programmabilité en théorie de l'information. Or, l'activité de débogage logiciel manifeste le fond de la réfutation scientifique. Lorsqu'un *bug* est détecté, le programmeur formule une hypothèse sur la cause du problème, puis la teste en insérant des traceurs dans le code, ou bien il lance un programme qui court-circuite cette étape de formulation. Le repérage de l'erreur correspond à une validation de l'hypothèse et en cas de réfutation, une nouvelle hypothèse est formulée. Il faut recourir ici à un procédé scientifique – la dichotomie – pour isoler les parties du code sur lesquelles le test est effectué. Le débogage apparaît comme un pur processus hypothético-déductif de réfutation. La déductibilité logique, propre à la méthode scientifique et typique de l'informatique, emporte avec elle dans l'ordinateur le raisonnement hypothético-déductif et la notion même d'expérience, tant ces notions sont liées et indissociables (Lassègue, 1996).

D'autre part, la démonstration logique est une pure procédure. En témoignent avant tout l'utilisation extrêmement répandue dans les laboratoires de ces machines bureaucratiques que sont les ordinateurs, l'intégration informatique et l'automatisation toujours plus large de la chaîne d'investigation scientifique⁷ (Liu, 2022), y compris les processus de revue scientifique, mais aussi maintenant leur rédaction (par « IA » générative). Le chercheur, en bon *data scientist*, commence à se repositionner sur la curation des données d'entrée, et l'optimisation des algorithmes d'apprentissage. Or, la procéduralisation de l'expérience conduit à isoler, découper et ordonnancer les problèmes dans le sens de la calculabilité. Quant à la donnée orchestrée dans une procédure d'investigation scientifique, elle révèle une décomposition mécaniste de cette dernière.

La mathématisation des théories à travers une modélisation dont l'essence est le calcul (Leibniz, 1714) peut être considérée comme un premier pas vers la programmation. La donnée constitue le grain à moudre de ces modèles, et dans un ordinateur, un modèle peut être lui-même une donnée. Outre cela, les données dans l'ordinateur sont structurées – de même qu'elles structurent l'éventuelle connaissance sous-jacente – comme le reflet d'un structuralisme scientifique plus général. Or, l'analyse structurale a pour but de produire des modèles. Certes, ces derniers sont d'abord des modèles de connaissance, mais la polysémie du « modèle » en informatique, aussi

⁶ Que nous assimilons ici à la normalisation et même à la standardisation.

⁷ LIU, Jia. 2022. Researchers Design AI-Chemist System to Facilitate Chemical Experiments. *Newsroom*, 20 octobre. URL : https://english.cas.cn/newsroom/cas_media/202210/t20221020_321893.shtml

large que celle de la « donnée », laisse planer le même doute : n'avons-nous pas bien vite affaire, avec l'obtention de modèles d'abord pleins de richesse sémantique et de perspectives d'intelligibilité, à de simples éléments opérables en proie à un puissant principe de computation ? Dans le même sens, le structuralisme mathématique (matrice de modèles) s'intéresse aux relations entre les structures formelles plutôt qu'aux objets en eux-mêmes. Il présuppose l'existence de structures universelles, définies par leurs relations et des opérations. Dans les sciences expérimentales, il a fait florès. Or, le structuralisme est tout à fait informatique, et explique l'efficacité opératoire des modèles d'apprentissage automatique, méthode en grande partie basée sur la détection et la modélisation de relations au sein des structures – par exemple, entre des caractéristiques d'entrée et les sorties. Plutôt que de se concentrer uniquement sur les propriétés ou la signification intrinsèques des données, on veut dans cette discipline informatique capturer les structures relationnelles qui définissent la façon dont les données sont organisées ou interagissent.

Structuration appelle systématisation. La systémique moderne révélée par le mouvement cybernétique plonge ses racines au moins jusqu'à la pensée de Descartes, et aboutit au « système » d'information. La donnée est une facette du système d'information, au mieux connaissance de l'interaction, au pire simple principe d'effectuation, mais pas connaissance de l'objet ni du sujet. Du point de vue de la connaissance, la mise en système de concepts est une forme de nominalisme dont le logiciel ou la représentation informatique est le modèle abouti.

Ainsi numérisation, formalisation, modélisation, déduction, structuration et systématisation ne produisent ensemble rien d'autre que de la donnée informatique. Les étapes théoriques de la démarche scientifique – observation, formulation d'hypothèse, expérimentation, interprétation du résultat – ne sont, sous l'hypothèse de la convergence, que des pis-aller tant que l'on ne dispose pas d'ordinateur. L'ordinateur présent, on passe aux « choses sérieuses » : la programmation logicielle, c'est-à-dire l'activation mathématique de la donnée.

Obtentions informatiques et perte de rationalité scientifique

Ainsi, toute discipline scientifique s'informatise à mesure qu'elle convoque ses propres principes de scientificité. Ces derniers sont-ils germes de prudence épistémique, ou menace de réductionnisme scientiste (Rouvroy, 2014) ? La mathématique est-elle *in fine* outil de compréhension ou vecteur universel d'opérationnalisation ? Quant aux données, sont-elles la colonne vertébrale ou les sucs digestifs de la rationalité scientifique ? Sous l'hypothèse de la convergence, la perte en rationalité de la science prend le visage de son informatisation.

La numérisation de l'activité scientifique est caricaturale en science des données. Elle induit le plus souvent une surpondération des valeurs moyennes et le lissage des singularités des phénomènes étudiés, par la dissolution des observables dans les mélanges statistiques. C'est tracer un chemin épistémique très étroit, qui interdit par principe de tirer profit des valeurs atypiques et extraordinaires⁸. La formule mathématique est devenue suite de nombres, et derrière elle les

⁸ La détection d'anomalies, quant à elle, permet de valoriser les valeurs extrêmes, mais celles-ci le sont toujours en regard d'une norme, et ces exceptions ne font que conforter la règle.

théories scientifiques les plus diverses se voient réduites à de simples équations ou modèles calculables (Lassègue, 2008). Au-delà, les données scientifiques programmables sont les plus valorisées parce que l'on manque le plus souvent de temps, de budget ou de compétence pour formuler de meilleures hypothèses afin de bâtir de nouvelles théories... Mais aussi parce qu'il est toujours tentant de tester même les moins bons modèles lorsqu'on dispose des moyens informatiques adéquats. Ainsi, toute modélisation vise plus ou moins consciemment une machine de Turing ayant évacué la majeure partie du non calculable.

D'abord le formalisme hilbertien, quant à lui, s'accommode du caractère préconceptuel de la connaissance, que l'on a mis en lien avec une « pensée » finitiste, typique de l'informatique (*ibid.*). Pour Wiener (1950), l'actuation de la machine peut s'opérer par des données simples ramassant sous elles toute complexité. Chez Turing, le programme exécutable est bien une compilation de signes (Grinbaum, 2016). La théorie mathématique de la communication de Shannon et Weaver (1998), à la base des langages de programmation, évacue tout regard sémantique sur l'information. Gödel a explicité ce passage de la démonstration signifiante à travers la formule, jusqu'au signe, et il fut contemporain en cela de l'invention de l'informatique. Formaliser, c'est aplatir la signification et chercher à unifier toutes les disciplines scientifiques dans un contexte d'informatisation (Hunault, 1999).

Que penser ensuite de la logicoisation des sciences, permise par cette formalisation ? L'atomisme logique défendu par Russell (1989) dès 1918 et le premier Wittgenstein (1921) promet une description du monde en un langage idéal composé uniquement de combinaisons logiques de faits élémentaires. Mais si la connaissance scientifique peut être logiquement traitée, sous forme de données, c'est que l'on a enclos l'investigation dans un champ logique déterminé pour garantir la validité des inférences, au détriment de la signification... ceci dans une éthique parfois très discutable (tests sur animaux, manipulations génétiques, etc.) qui pourrait être rapprochée d'un manque de rationalité circonstancielle⁹.

La procéduralisation induite est attestée déjà chez Descartes, qui applique sa méthode à la démonstration arithmétique pour atteindre une rigueur et une certitude de la conclusion jamais revendiquée avant lui (De Buzon, 2005). L'« enquête pure » cartésienne (Williams, 1978) évoque alors l'algorithme ou l'arbre de décision, objets informatiques permettant d'avancer des prémisses vers une conclusion d'une manière ni subjective et ni évaluative, que Rouvroy (2014) n'est pas loin de taxer d'obscurantiste. La multiplication historique des silos en science (Marin, 2021) est un fantasme de procéduralisation du vivant visant sa mise en ordre. La procéduralisation implique un discontinuisme scientifique contestable (*ibid.*) : la division disciplinaire empêche d'envisager les problèmes dans leur globalité et postule une capacité de synchronisation organisationnelle *ex-post* qui fait bien souvent défaut en pratique. À mesure qu'elle conduit à l'informatisation, la

⁹ L'objection que ces considérations relèveraient du champ de l'éthique et non de la science, puiserait ses racines dans une conception de la science justement réduite au programme informatique : si l'on croit que les facteurs premiers de la démarche scientifique relèvent de la méthode mathématique, il faut bien admettre que toute autre considération raisonnable, comme le découragement du chercheur, est secondaire et pseudo ou extra-scientifique.

dichotomie évacue les epsilons, ces valeurs qui ne peuvent être ni 0 ni 1¹⁰. La discrétisation du continu serait-elle aboutie au plan des données informatiques ?

D'autre part, la conception de modèles en apprentissage machine ne repose pas sur le triptyque idéal théorisation-expérimentation-observation propre à la méthode hypothético-déductive, mais sur des formalismes inférentiels automatisés, et les modèles obtenus en science des données réduisent les groupes et collections au plus petit dénominateur signifiant, au détriment de la prise en considération des singularités exceptionnelles. Est menacée de ce régime, toute activité scientifique utilisant un ordinateur. Or, sur le chemin épistémique du singulier non régulier pourraient se trouver de formidables connaissances et pratiques... Réciproquement, sur le chemin du régulier non singulier, la méthode ne dit pas comment reconnaître seulement les hypothèses qu'il sera raisonnable de tester sans déclencher, par exemple, la fin du monde¹¹. Il faut aussi rappeler qu'on n'obtient les résultats réguliers qu'incidemment avec la méthode. Son objet immédiat n'est pas de donner les résultats intéressants, mais de prodiguer les moyens de chercher des modèles calculables (molécules, artefacts techniques, régularités sociologiques, etc.).

Ensuite, la critique du structuralisme nous enseigne à nous méfier de l'effet de prisme, et de la confusion entre l'objet et la méthode d'étude. Ainsi, le structuralisme mathématique, qui préside à l'apprentissage automatique pour découvrir des structures, contourne en science la réalité de la qualification des objets auxquels il s'intéresse – comme en droit, le système normatif formel éreinte la signification des normes. Le structuralisme fait le pari que ces qualités intrinsèques sont insignifiantes. Plus loin, la structuration d'un programme en bases de données permet de déplacer le traitement dans un monde pseudo-platonicien de formes pures, en évacuant les questions tenant à la qualité intrinsèque et à la signification.

Ainsi, chaque critère de scientificité a ses travers ; mobilisés ensemble, ils réduisent le champ des possibles scientifiques à celui du programmable informatique. Or, même si nous devons appeler « rationnel » un nouveau programme informatique au motif qu'il est calculable, il faudrait bien nommer « irrationnelles » ses contreparties irrémédiables dans un ordinateur, les logiciels déployés étant, contrairement aux algorithmes étudiés, des artefacts composites éminemment conventionnels. La rationalité perdue, c'est au moins celle liée au caractère non-calculable de l'activité scientifique, ainsi que ses à-côtés et contreparties nécessaires, en creux de la diffusion d'innovations dans un monde qui ne sait pas (ne pourra jamais ?) les réguler ; le caractère gouvernant de logiciels scientifiques sensément exploratoires ; la multiplication des dépendances et difficultés d'administration de programmes toujours plus complexes et fragiles ; l'augmentation de la dette énergétique et environnementale corrélative à l'existence même d'une nouvelle donnée numérique, etc. La donnée dans l'ordinateur est comme la preuve par l'absurde que l'accumulation des règles logiques et des modèles calculables ne sera jamais que de la dette technique et cognitive.

¹⁰ De nouveaux *chips* de type analogique et quantique sont inventés, ce qui permettrait d'y échapper... Mais s'il fallait les appeler à travers des interfaces booléennes, se poserait la question de savoir si quelque chose de propre au réductionnisme binaire ne viendrait pas contaminer les résultats qu'ils obtiendraient.

¹¹ RAPPAPORT, Sheila K. 2008. Malgré les déclarations optimistes. La fin du monde a sans doute eu lieu. *Scientists of America*. URL : <https://scientistsofamerica.com/index.php?texte=114>

Convergence finale de la science des données et de la science par les données

Peut-on espérer, malgré ce qui précède, que les principes de scientificité les plus évaluatifs – ceux qui ne contribuent pas directement au calcul – puissent tenir lieu de renforts à la science qualitative ? La prudence épistémique peut-elle s'accorder à la fois de la logique et du formalisme ? Ce serait retrouver l'hypothèse de l'informatique comme une rupture de scientificité. Des exemples nous en éloigneront pourtant. Nous allons voir comment la méthode reste aveugle et sourde à la science qualitative, au doute et à l'indécision.

Un recensement portant sur 179 publications scientifiques appliquant l'apprentissage machine à des problèmes de chimie (Hagg et Kirschner, 2023) montre que cette technique de computation a colonisé de multiples domaines de la chimie théorique – résolution de problèmes abstraits liés à la modélisation moléculaire, prédiction de structures moléculaires, simulation de réactions chimiques, analyse de données spectroscopiques, étude des interactions atomiques... – et de la chimie appliquée – formulation de produits chimiques, étude de qualité des matériaux, optimisation des processus chimiques, prédiction de propriétés, qualité, en chimie analytique, etc. C'est dire que la science des données, pratique de programmation la plus explicitement scientifique, tend à épuiser la démarche scientifique elle-même. Elle en couvre tendanciellement tous les champs et tous les modes. Pourquoi ? Parce que l'effet de colonisation des critères évaluatifs et interprétatifs par les critères calculables et prescriptifs repousse toujours plus loin le moment de la prudence épistémique, comme une coquetterie éthique *a posteriori*. Ou, peut-être : la science qualitative joue comme l'argument publicitaire de l'informatisation tendancielle. Encore nous faut-il montrer de quelle manière.

Prenons l'exemple du critère, apparemment évaluatif, de reproductibilité de l'expérience. La reproductibilité exige des protocoles d'expérimentation, ce qui passe par l'algorithmisation de cette dernière. Le gage de la reproductibilité est ainsi, au fond, celui de la programmabilité. Allons plus loin : le phénomène reproductible à coup sûr dans des conditions identiques, c'est celui qui se voit simulé ou modélisé dans un ordinateur, lieu dans lequel les conditions sont parfaitement stables, et les effets conformes aux inférences. En-dehors d'un ordinateur, sur la pailleasse, reproduire suppose bien la neutralisation des variables non directement étudiées ; mais la neutralisation d'une variable particulière peut rendre les effets observés sensibles à une autre variable tout aussi indésirable. Cet inconvénient peut disparaître en programmation logicielle.

Un autre exemple est celui de l'amélioration d'une théorie à travers la prise en compte de nouveaux résultats. C'est véritablement un programme informatique qui peut assurer le lien entre de nouvelles données et un nouveau modèle, formulation réductrice d'une théorie scientifiquement améliorée. La possibilité de réfuter une théorie, marque poppérienne de sa scientificité, suppose que cette dernière soit suffisamment formalisée, à tout le moins qu'elle ne produise pas de prédictions invérifiables. Or, un ordinateur prend idéalement en charge ces deux aspects. Autrement dit, la mise en évidence idéale des conditions expérimentales qui pourraient prouver la fausseté d'une théorie, c'est l'obtention par l'ordinateur des valeurs à partir desquelles les

variables commencent à ne plus produire le résultat reconnu comme correct. Le programme informatique apparaît d'abord comme un moyen de tester la validité d'une théorie scientifique, faisant de l'ordinateur une véritable machine de défaisabilité.

Dès lors, les injonctions à trouver un équilibre, au sein de la méthode scientifique, entre le pôle calculable et le pôle évaluatif, ne sont jamais proférées que depuis des chaires très platoniciennes ; celles où l'on n'a pas directement affaire aux nécessités de régler un ordinateur. La donnée se présente bien comme le cheval de Troie de l'informatique au sein des méthodes qualitatives, ces dernières n'étant capables *in fine* que de proposer de nouveaux modèles calculables, comme chez Silver et Lewins (2020)... Quoiqu'elle produise aussi ce phénomène moins régulier, moins formalisable et peu vérifiable – bref moins scientifique – qu'est la conviction du chercheur.

Conclusion

Ce qui précède nous permet de désigner ce dont la donnée informatique est le nom : l'atomisation des sciences (naturelles, juridiques et autres) en données numériques, sous la force de leur propre méthode... Expliquant au passage la polysémie de la notion de données : elle est convoquée à divers stades de décomposition. Face à nos arguments, l'hypothèse de l'informatique comme rupture extrinsèque à la rationalité scientifique semble bien fragile. C'est bien plutôt la méthode, s'accomplissant en informatique, qui fait converger toute science dans l'ordinateur. Celle-ci toutefois n'est qu'un programme descriptif/préscriptif, et l'on peut se demander de quoi ce programme est le nom. Peut-être la dynamique de sociabilisation, voire d'hominisation, réclame-t-elle un rapport à la loi toujours plus exact ? Quoi qu'il en soit, l'ordinateur nous apparaît ici comme une preuve par l'absurde de l'irrationalité de notre rationalité scientifique. Appliquant la méthode, le scientifique ne sait jamais ce qu'il provoquera, et l'agent numérique est le parangon de cette indifférence aux effets.

C'est alors au-delà de la méthode et à l'extérieur de l'ordinateur que nous envisageons un meilleur sens de la rationalité. Ayant compris que la méthode scientifique n'est, pas plus que la règle de droit, prescription, mais description proposée d'une juste chose à faire, nous pourrions envisager ces notions dans de plus riches et libres acceptions.

Bibliographie

BECCARIA, Cesare. 1991 [1764]. *Des délits et des peines* [*Dei delitti e delle pene*]. Trad. de CHEVALLIER, Maurice. Paris : GF-Flammarion.

BILLION, Arnaud. 2021. *Éthique de l'informatique et droit naturel : à la recherche de la solution technique équitable*. In : BARBÉ, Vanessa, KERLÉO, Jean-François et PADOVANI, Julien. *L'Éthique à l'épreuve de la crise*. Le Mans : L'Épitoge, pp. 279 s.

BILLION, Arnaud. 2022. *Sous le règne des machines à gouverner. Le droit entre intelligence artificielle et raison naturelle*. Bruxelles : Bruylant.

BRAGUE, Rémi. 2016. *Modérément moderne*. Paris : Flammarion.

- CATTARUZZA, Amaël. 2019. *Géopolitique des données numériques. Pouvoir et conflits à l'heure du big data*. Paris : Le cavalier bleu.
- DE BUZON, Frédéric. 2005. Mathématiques et dialectique : Descartes ramiste ? *Les Études philosophiques* [En ligne], vol. 4, n° 75, pp. 455-467. DOI : 10.3917/leph.054.0455.
- DENZIN, Norman K. et LINCOLN, Yvonna S. 2005. *The SAGE handbook of qualitative research*. 3^e éd. Londres/Thousand Oaks : Sage.
- GANASCIA, Jean-Gabriel. 2022. Enjeux épistémiques de la science des données. *Annales des Mines* [En ligne], vol. 3, pp. 45-48. DOI : 10.3917/rindu1.223.0045
- GRIARD, Jérémie. 2006. Conformer le droit volontaire au droit naturel. *Revue des sciences philosophiques et théologiques* [En ligne], vol. 90, n° 3, pp. 479-494. DOI : 10.3917/rspt.903.0479
- GRINBAUM, Alexei. 2016. *Gouverner le code qui nous gouverne*. Conférence donnée à la 2^e journée de la CERNA le 1^{er} février. URL : https://www.academia.edu/21395452/Gouverner_le_code_qui_nous_gouverne
- GUILLERMIN, Mathieu. 2019. Non-neutralité sans relativisme ? In : BRIÈRE, Laure, LIEUTENANT-GOSSELIN, Mélissa et PIRON, Florence (dir.). *Et si la recherche scientifique ne pouvait pas être neutre ?* [En ligne]. Québec : Science et bien commun. URL : <https://corpus.ulaval.ca/server/api/core/bitstreams/5b79722a-cb07-498b-8f50-e4466e55bad1/content>
- HACKING, Ian. 2006. Making Up People. *London Review of Books* [En ligne], vol. 28, n° 16, 17 août. URL : <https://www.lrb.co.uk/the-paper/v28/n16/ian-hacking/making-up-people>
- HAGG, Alexander et KIRSCHNER, Karl N. 2023. Open-Source Machine Learning in Computational Chemistry. *Journal of chemical information and modeling* [En ligne], vol. 63, n° 15, pp. 4505-4532. DOI : 10.1021/acs.jcim.3c00643
- HOBBS, Thomas. 2002 [1651]. *Léviathan - Traité de la matière, de la forme et du pouvoir de la république ecclésiastique et civile*. [En ligne]. Trad. de FOLLIOU, Philippe. URL : http://www.uqac.quebec.ca/zone30/Classiques_des_sciences_sociales/index.html.
- HUNAULT, Cyril. 1999. *La réduction des entités chez Bertrand Russell la construction logique du « monde physique »*. Mémoire de maîtrise en philosophie. Nantes : Université de Nantes.
- LAFFORGUE, Laurent. 2009. *Simone Weil et la mathématique* [En ligne]. Paris : Bibliothèque nationale de France, 23 octobre. URL : <https://www.laurentlafforgue.org/textes/SimoneWeilMathematique.pdf>
- LASSÈGUE, Jean. 1996. La méthode expérimentale, la modélisation informatique et l'intelligence artificielle. *Intellectica* [En ligne], n° 22, pp. 21-65. DOI : 10.3406/intel.1996.1515
- LASSÈGUE, Jean. 2008. Turing, entre le formel de Hilbert et la forme de Goethe. *Matière première. Revue d'épistémologie et d'études matérialiste*, n°3, pp. 57-70.
- LASSÈGUE, Jean. 2022. Le droit automatisé et le problème de la délibération collective. *Daloz IP-IT*, pp. 21-23.

LEIBNIZ, Gottfried Wilhelm. 1714. *La Monadologie. Précédé de La Monadologie de Leibniz / suivie de La philosophie de Leibnitz*. La Flèche : Librairie générale française.

MARIN, Pascal. 2021. Penser l'humain au temps de son effacement dans le vivant et la machine. *Bulletin d'anthropologie philosophique* [En ligne], tome 105, pp. 487-532. DOI : 10.3917/rspt.1053.0487

MAZZOCCHI, Fulvio. 2015. Could Big Data be the end of theory in science? A few remarks on the epistemology of data-driven science. *EMBO reports* [En ligne], vol. 16, pp. 1250-1255. DOI : 10.15252/embr.201541001

MONTESQUIEU, Charles de Secondat de. 2002 [1748]. *De l'esprit des lois* [En ligne]. Chicoutimi : Les Classiques des sciences sociales. URL : http://classiques.uqac.ca/classiques/montesquieu/de_esprit_des_lois/partie_1/esprit_des_lois_Livre_1.pdf

PAKSY, Mate. 2018. Leibniz et les questions de l'ontologie juridique : la science, les règles et le concept du droit. *Astérior. Philosophie, histoire des idées, pensée politique* [En ligne], n° 19, pp. 1-23. DOI : 10.4000/asterion.3573

POPPER, Karl. 1973 [1934]. *La logique de la découverte scientifique [Logik der Forschung]*. Trad. de THYSSEN-RUTTEN, Nicole et DEVAUX, Philippe. Paris : Payot.

RAPPIN, Baptiste. 2014. *Théologie de l'organisation*. Nice : Ovidia.

REY, Olivier. 2017. *Quand le monde s'est fait nombre*. Paris : Stock.

RUSSELL, Bertrand. 1989. *Écrits de logique philosophique*. Trad. de ROY, Jean-Michel. Paris : Presses Universitaires de France.

ROUVROY, Antoinette. 2014. Des données sans personne : le fétichisme de la donnée à caractère personnel à l'épreuve de l'idéologie des Big Data. *Le numérique et les droits et libertés fondamentaux. Étude annuelle du Conseil d'État*. Paris : La documentation française, pp. 407-422.

ROUVROY, Antoinette et BERNS, Thomas. 2013. Gouvernamentalité algorithmique et perspectives d'émancipation. *Réseaux* [En ligne], n° 177 1, pp. 163-196. DOI : 10.3917/res.177.0163

SHANNON, Claude et WEAVER, Warren. 1998. *The mathematical theory of communication*. Chicago: University of Illinois Press.

SILVER, Christina et LEWINS, Ann. 2020. Computer-Assisted Analysis of Qualitative Research. In: LEAVY, Patricia (dir.). *The Oxford Handbook of Qualitative Research*. Oxford: Oxford University Press, pp. 912-951.

STAGNARO, Angelo. 2017. *The Italian Jesuit Who Taught Computers to Talk to Us. National Catholic Register* [En ligne]. URL: <https://www.ncregister.com/blog/the-italian-jesuit-who-taught-computers-to-talk-to-us>.

SUPIOT, Alain. 2015. *La gouvernance par les nombres. Cours au Collège de France, 2012-2014*. Nantes/Paris : Institut d'Études Avancées de Nantes/Fayard.

VILLEY, Michel. 2017. *La formation de la pensée juridique moderne*. Paris : PUF.

WEBER, Max. 2003 [1922]. *Économie et société*. Paris : Pocket.

WIENER, Norbert. 1950. *The human use of human beings. Cybernetics and society*. Cambridge: The Riverside Press.

WIGNER, Eugene. 1960. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics* [En ligne], vol. 13, n°1, pp. 1-14. DOI : 10.1002/cpa.3160130102

WILLIAMS, Bernard. 1978. *Descartes. The project of pure enquiry*. Harmondsworth: Penguin.

WITTGENSTEIN, Ludwig. 1921. *Logisch-Philosophische Abhandlung*. Leipzig: Wilhelm Ostwald.