

Nouvelles perspectives en sciences sociales



Complexité des systèmes naturels et sous-détermination des théories : une possible limite de la modélisation

Henri Atlan

Volume 4, Number 2, avril 2009

Sur le thème de la complexité

URI: <https://id.erudit.org/iderudit/029890ar>

DOI: <https://doi.org/10.7202/029890ar>

[See table of contents](#)

Publisher(s)

Prise de parole

ISSN

1712-8307 (print)

1918-7475 (digital)

[Explore this journal](#)

Cite this article

Atlan, H. (2009). Complexité des systèmes naturels et sous-détermination des théories : une possible limite de la modélisation. *Nouvelles perspectives en sciences sociales*, 4(2), 35–45. <https://doi.org/10.7202/029890ar>

Article abstract

In Probabilistic Information Theory, as in Algorithm and Programming theory, there is no need to worry about how we understand or how we come to create meaning. Within each of these complex cases, we run into the same paradox: formal identity between maximum complexity and randomness (*i.e.* disorder within maximum statistical homogeneity). And, in both cases, the solution to this paradox consists of ignoring that sense and meaning exist *a priori* while supposing it does, which eliminates the hypothesis of randomness. It isn't until recently that paper in Algorithmic Complexity have attempted to solve this paradox, taking into consideration a definition of complexity that is meaning based. A first approach is to account for complexity in noise. A second, more recent approach, uses Multi-Agent Cellular Automata Systems to catch the emergence of functional meaning within the self-organizing networks. As a result of these approaches, we find a large under-determination of theories within the facts, and a small number of networks that allow us to retrace the origins and even quantify them. This under-determination of theory appears to be the most spectacular expression of what is natural complexity.

Tous droits réservés © Prise de parole, 2009

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

Complexité des systèmes naturels et sous-détermination des théories : une possible limite de la modélisation¹

HENRI ATLAN

École des hautes études en sciences sociales (EHESS), Paris

1. Complexité des systèmes artificiels et complexité des systèmes naturels : la place de la signification

Nous pouvons rencontrer trois notions différentes de complexité. Deux sont formalisées et désignent, respectivement,

¹ Avertissement au lecteur : Ce texte est un extrait d'un article déjà publié par Henri Atlan dans l'ouvrage dirigé par Françoise Folgelman Soulié, *Les théories de la complexité. Autour de l'œuvre d'Henri Atlan* (Paris, Seuil, 1991), reproduit récemment dans un livre d'Henri Atlan, *L'organisation biologique et la théorie de l'information* (Paris, Seuil, coll. « Librairie du XX^e et du XXI^e siècles, 2006) et dont la source était un colloque tenu dans le lieu prestigieux de Cerisy, en 1988.

En contradiction avec nos principes, mais de façon exceptionnelle, nous avons décidé de publier cet extrait du fait de son vif intérêt et de son actualité, même vingt ans après, alors que la simulation à base de réseaux d'automates pénètre dans les sciences humaines et sociales, comme NPSS s'en est fait l'écho, notamment dans le volume précédent.

Qu'Henri Atlan soit remercié, dans le même mouvement, pour cette contribution aux réflexions que souhaite susciter notre revue et plus généralement pour l'ensemble de ses travaux qui continuent de nous éclairer et, mieux encore, de nous interroger.

Merci également aux responsables des éditions du Seuil pour leur gracieuse autorisation de cette reproduction.

la difficulté pour une machine programmée à accomplir une tâche, et l'incertitude probabiliste (ou l'information manquante) sur une structure observée. La troisième désigne une intuition, non quantifiée, sur notre difficulté à comprendre un exposé, une idée. Il est clair que cette difficulté de compréhension ne peut pas être précisée tant que nous ne savons pas exactement quels types d'opérations nous faisons quand nous comprenons ou ne comprenons pas quelque chose. Autrement dit, cette complexité-ci concerne essentiellement la signification de ce qu'il s'agit d'analyser et de comprendre, que ce soit un observable dans la nature, un problème abstrait à résoudre ou un artefact fabriqué suivant une procédure automatisable. Et c'est la difficulté à formaliser la signification des mots, des phrases et des choses qui se retrouve ici comme difficulté à formaliser cette sorte de complexité. Par ailleurs, les deux autres sortes de complexité, algorithmique et probabiliste, ne sont formalisées que dans la mesure où elles ne s'occupent pas explicitement de questions de signification. Aussi bien la théorie de l'information probabiliste que la théorie des algorithmes de programmation atteignent leurs objectifs de quantification sans avoir à s'occuper de la question de savoir comment nous comprenons ni comment les significations sont créées. En fait, il est bien entendu que les significations existent, celles de l'information transmise dans une voie de communication, comme celles des instructions d'un programme d'ordinateur.

Mais les théories mathématiques qui traitent de ces questions n'ont pas besoin, pour être opérationnelles, d'en rendre compte explicitement. Dans la théorie de l'information, le fait que messages et structures ont une signification est évident, mais reste implicite, car on n'y traite que de problèmes de codage et d'efficacité de transmission sans qu'il soit nécessaire d'envisager la signification effective des messages à coder et à transmettre. Et c'est le même formalisme qui est utilisé pour mesurer la complexité probabiliste d'une structure, en supposant qu'il existe une voie de communication entre la structure observée et l'opération d'observation. La mesure est la même pour toute page écrite en une langue donnée, quelle que soit sa signification, car elle ne

dépend que du nombre de signes contenus dans la page et de la distribution des fréquences d'utilisation des différents signes de l'alphabet, caractéristique de la langue utilisée. Mais le fait qu'il s'agisse d'une langue parlée ou écrite implique que les signes véhiculent des significations ainsi qu'on peut le vérifier en observant leur effet sur des récepteurs (auditeurs ou lecteurs) qui la comprennent. De la même façon, une structure, dont nous ne connaissons que la distribution des fréquences d'utilisation des composants dans une classe de structures équivalentes ne nous apparaît comme différente d'un simple amas de pièces jetées au hasard que parce que nous y supposons ou observons une ou plusieurs significations sous la forme de fonctions que cette structure accomplit. Mais ces significations, encore une fois, ne sont qu'implicites dans la théorie, qui se borne à mesurer des fréquences ou à calculer des probabilités sans les prendre en compte explicitement.

De même, il est évident qu'un programme d'ordinateur écrit pour résoudre un problème ou accomplir une tâche a une signification — précisément, résoudre ce problème ou accomplir cette tâche. Pourtant, la théorie de la complexité algorithmique est fondée sur des considérations logiques de longueur de calcul et de décidabilité qui n'ont pas à prendre en compte explicitement cette signification : la définition de la finalité du programme, c'est-à-dire du problème particulier à résoudre ou de la tâche à accomplir, n'est pas nécessaire pour faire fonctionner la théorie. Cette finalité est toujours évidente, et il n'est pas nécessaire de s'interroger sur son origine quand on analyse les propriétés logiques et opérationnelles du programme.

Le fait que la signification — de l'information transmise ou du programme à exécuter — ne joue aucun rôle explicite dans ces théories apparaît de façon spectaculaire dans la relation paradoxale qu'on y découvre entre complexité et hasard.

Suivant la définition la plus communément acceptée (Kolmogorov-Chaitin), la complexité algorithmique d'un problème à résoudre ou d'une tâche à accomplir est la longueur minimale de programme et de données nécessaires pour y parvenir.

En général, une suite structurée d'éléments, pourvu qu'elle soit assez longue, peut être produite par un programme qui réalise cette structure en étant plus court que la suite elle-même. Cela est évident dans tous les cas où un programme de longueur finie est capable de produire une suite de longueur infinie. Il en résulte le paradoxe qu'une suite infinie aléatoire — sans structure — représente la tâche la plus complexe à réaliser, puisque la seule façon de la produire est de la recopier et qu'il n'existe donc aucun programme plus court capable de la générer. Mais ce paradoxe n'en est un que par rapport à notre intuition globale de ce qu'est le complexe, où la signification et le sens semblent devoir occuper une place centrale à l'opposé du désordre statistique de l'aléatoire. Il n'empêche pas la théorie de fonctionner de façon utile dans son domaine d'application d'où, par définition, sont exclues de telles suites aléatoires puisqu'il est évident et implicitement admis qu'un programme écrit pour accomplir une tâche donnée ne peut pas être une suite aléatoire de signes².

On rencontre le même problème — avec la même solution — dans la définition de la complexité d'un objet naturel par la fonction entropie de Shannon. Comme toute mesure d'entropie, il s'agit là aussi d'une grandeur dont le maximum correspond à l'homogénéité statistique, au désordre de l'aléatoire. Et la seule différence entre notre expérience d'un désordre dans la nature et celle d'une complexité ainsi estimée provient de notre observation que celle-ci accomplit une ou plusieurs fonctions. Au minimum, à défaut d'observation directe — comme dans le cas d'une fourmilière en désordre où pourtant des fonctions d'organisation de l'espace, de division du travail et de stockage de réserves sont

² Remarquons pourtant qu'un programme minimal se présente comme une suite aléatoire d'après cette définition puisque, étant minimal, il ne peut pas être produit par un programme plus court que lui. La seule raison qui nous le fait considérer comme autre chose qu'une suite aléatoire est sa signification, c'est-à-dire le fait que nous savons, par ailleurs, qu'il est capable de produire une certaine fonction quand il est lu et interprété par une machine de Turing (voir Moshé Koppel et Henri Atlan, « Les gènes : programmes ou données? Le rôle de la signification dans les mesures de complexité », dans Françoise Fogelman Soulié (dir.), *Les théories de la complexité. Autour de l'œuvre d'Henri Atlan*, Seuil, Paris, 1991, p. 188-204).

observées —, nous supposons comme hypothèse de travail, sur la base d'arguments plus ou moins convaincants, que le désordre apparent est responsable de la réalisation d'une ou plusieurs fonctions. Autrement dit, nous attribuons une signification à ce que nous désignons comme complexe et non au simple (!) désordre, alors même que, formellement, la grandeur pour les mesurer — l'entropie — est la même.

C'est aussi vrai, d'ailleurs, d'une machine programmée où la signification d'un composant ou d'une instruction de programme concerne sa fonction, son rôle dans l'organisation fonctionnelle de l'ensemble de la machine. Mais l'origine de ces fonctions et, partant, de ces significations est très différente dans les deux cas. Dans une machine, la tâche à accomplir étant spécifiée de l'extérieur, c'est elle qui est source de significations en ce que c'est par rapport à elle qu'un composant ou une instruction apparaîtra comme absurde et vide de sens s'il ne contribue en rien à la réalisation de cette tâche tandis que sa signification apparaîtra dans sa contribution plus ou moins évidente et plus ou moins efficace à cette réalisation. Au contraire, dans une organisation naturelle, comme un organisme ou un écosystème, nous ne pouvons pas supposer qu'une finalité a été établie *a priori* par une conscience, soit à l'extérieur soit même à l'intérieur. Dans les deux cas, cela reviendrait à recourir à Dieu comme principe explicateur sous la forme soit d'un Créateur extérieur à la nature, soit d'une conscience cosmique embrassant tout; dans les deux cas, la valeur explicative est nulle, puisque personne ne peut savoir quelle fut l'intention de Dieu, si intention il y eut dans l'organisation d'une espèce vivante, d'un organisme ou d'un écosystème au contraire, bien entendu, d'une machine fabriquée par l'homme.

Dans les deux cas — complexité naturelle probabiliste et complication algorithmique —, nous rencontrons le même paradoxe : une identité formelle entre complexité maximale et aléatoire (c'est-à-dire désordre avec homogénéité statistique maximale). Et, dans les deux cas, la solution du paradoxe consiste à l'ignorer en supposant qu'un sens et une signification existent

a priori, ce qui élimine de ce fait l'hypothèse de l'aléatoire.

Ce n'est que très récemment qu'on a tenté de résoudre vraiment ce paradoxe par des travaux sur la complexité algorithmique tenant compte d'une définition de la complexité porteuse de signification ou porteuse de sens. Dans le travail déjà cité en collaboration avec Moshé Koppel, qui s'inscrit directement dans cette ligne, nous définissons formellement une telle complexité porteuse de signification, que nous dénommons « sophistication », qui s'avère nulle comme on s'y attend, et non maximale, dans le cas d'une suite aléatoire.

2. Approches de la création de signification dans les systèmes vivants

Quant à la complexité naturelle, je rappellerai très brièvement quelques tentatives de prendre en compte d'une façon ou d'une autre le rôle des créations de significations dans les processus d'auto-organisation.

Dans les organismes vivants, la signification de l'information s'exprime dans une fonction physiologique observable dont on n'a aucun moyen de savoir si elle a été conçue et réalisée avec une intention précise qu'il soit possible de définir de façon non circulaire (autre que la vision pour voir, la respiration pour respirer, etc.). C'est pourquoi divers formalismes furent utilisés pour tenter de modéliser des mécanismes d'auto-organisation par lesquels des systèmes non intentionnels, non finalisés de l'extérieur peuvent s'organiser eux-mêmes de façon telle que la signification de l'information y soit propriété d'une dynamique. Je me contenterai de citer deux séries de travaux par lesquels j'ai tenté de poser le problème de la création de signification dans des systèmes biologiques. La première concerne ce que j'ai appelé le principe de complexité par le bruit.

La théorie probabiliste de l'information y est étendue en sorte d'établir des conditions nécessaires pour une auto-organisation avec accroissement de complexité, c'est-à-dire création d'information. La question de la création de signification y est ici traitée en négatif, en prenant avantage de ce que la quantification de

l'information dont il s'agit ici ne prend pas explicitement en compte sa signification. L'auto-organisation fonctionne comme suite de désorganisations rattrapées en réorganisations. La complexité par le bruit, dans ce formalisme, est l'expression d'une affirmation par double négation : la destruction par le bruit (première négation) d'une information d'où la signification est absente (seconde négation) peut être équivalente, à condition de changer de niveau d'intégration et d'observation, à la création d'une complexité nouvelle. Celle-ci implique la création de significations sans lesquelles la désorganisation par le bruit ne pourrait être accompagnée de réorganisation.

Dans une seconde approche, plus récente, en collaboration avec Gérard Weisbuch et Françoise Fogelman Soulié³, nous utilisons des simulations de réseaux d'automates pour tenter de surprendre, de façon positive cette fois, l'émergence de significations fonctionnelles dans les réseaux d'automates à propriétés auto-organisatrices. Il est bien connu que de nombreux types de réseaux d'automates présentent des propriétés d'auto-organisation structurale en ce que leur dynamique les fait évoluer de conditions initiales homogènes vers des attracteurs où des structures spatio-temporelles macroscopiques peuvent être observées. C'est tout un champ de recherches très actif actuellement, où il s'agit de comprendre comment s'effectuent le passage du local au global, la détermination de propriétés macroscopiques émergentes (du réseau) à partir d'effets de coopérativité entre propriétés microscopiques (des automates individuels). Mais, de plus, ces réseaux peuvent faire apparaître des exemples d'auto-organisation fonctionnelle tels que la simulation d'une machine, construite en partie au hasard, capable de classer et de reconnaître des formes sur la base de critères, qui sont eux-mêmes autogénérés : les formes reconnues et les critères de reconnaissance sont observés *a posteriori*, car ils sont les produits de processus auto-organiseurs et n'ont pas été programmés explicitement. Ce n'est

³ Henri Atlan, Esther Ben-Ezra, Françoise Fogelman Soulié, Didier Pellegrin et Gérard Weisbuch, « Emergence of Classification Procedures in Automata Networks as a Model for Functional Self-Organization », *Journal of Theoretical Biology*, n° 120, 1986, p. 371-380.

qu'*a posteriori* que le mécanisme du système de reconnaissance peut être (parfois) compris. Comme dans l'observation et l'analyse d'un système vivant naturel, non programmé par l'homme, le critère de signification doit être déchiffré après coup, parfois difficilement car il est le produit global, non programmé explicitement, d'un grand nombre d'interactions locales, et les significations qu'il crée ainsi dans son activité classificatrice peuvent sembler *a priori* étranges du point de vue d'un être rationnel doué d'une intention planificatrice.

On peut voir là une façon de simuler en quoi la nature, par opposition à l'œuvre d'un architecte, semble travailler par « bricolage », comme le dit François Jacob. Remarquons pourtant que les techniques préscientifiques apparaissent aussi comme du bricolage (Claude Lévi-Strauss). Leur efficacité aurait ainsi peut-être la même origine que celle du bricolage de la nature. Enfin, la réalisation récente que nos activités de découvertes scientifiques, elles aussi, doivent être souvent attribuées à la raison rusée de Mêtis (*cunning reason*)⁴, qui évoque elle aussi ce bricolage, semblerait ainsi faire retourner notre science et notre fabrication d'artefacts dans le giron de la nature. Après tout, nous aurions dû nous y attendre. Seul le Dieu de Descartes ne ruse pas et ne veut pas nous tromper. C'est parce que les physiciens ont cru quelque temps que leur tâche était de découvrir son activité d'Architecte et de l'imiter que leur activité a semblé ne relever que de celle de Thémis, de l'architecte et du géomètre classiques, qui mesurent et planifient, éliminent l'imprévu et ignorent les jeux du vivant.

3. Sous-détermination des théories par les faits ou des modèles par les observations

Le calcul par réseaux d'automates à seuil sur le modèle des réseaux neuronaux imaginé autrefois par McCulloch et Pitts⁵ s'est

⁴ Marcel Detienne et Jean-Pierre Vernant, *Les ruses de l'intelligence. La Mêtis des Grecs*, Flammarion, 1974; Yehuda Elkana, « A Programatic Attempt at an Anthropology of Knowledge », Everett Mendelson et Yehuda Elkana (dir.), dans *Sciences and Cultures*, Dordrecht, Reidel, 1981, p. 1-76.

⁵ Voir Henri Atlan, *L'organisation biologique et la théorie de l'information*, Paris,

beaucoup développé récemment dans le cadre des théories néo-connexionnistes. Mais son utilisation nous permet aussi de simuler des phénomènes de régulation immunitaire et, de façon générale, de régulation biologique où sont impliqués plusieurs processus fonctionnant de façon couplée. Là, il n'est pas nécessaire de faire appel à des milliards ni même à des centaines d'éléments en interaction; de petits réseaux, de l'ordre d'une dizaine d'automates ou moins, sont très utiles en même temps qu'ils font apparaître une richesse de comportement aussi grande, relativement à leur taille, que les grands réseaux neuronaux. En particulier, on y reconnaît déjà une large sous-détermination des théories par les faits, et la petite taille de ces réseaux permet d'en analyser clairement l'origine et même de la quantifier. Cette sous-détermination des théories apparaît ainsi comme l'expression probablement la plus spectaculaire de ce qu'est la complexité naturelle.

L'analyse de ces petits réseaux neuronaux montre que cette propriété n'est pas spécifique, comme on l'a cru, des phénomènes humains psychologiques, linguistiques ou autres, et ne résulte donc pas d'un rapport de la théorisation aux faits découlant d'aspects mystérieux de la relation corps-esprit. Beaucoup plus simplement, elle provient de l'augmentation rapide du nombre de théories possibles, différentes mais équivalentes quant à leur pouvoir prédictif de faits d'observation, dès que ces faits sont déterminés par les effets couplés de plusieurs facteurs ou processus. Chaque théorie est représentée par une structure de connexions donnée d'un réseau dont les automates représentent les différents facteurs ou processus en cause. En général, les faits observés ne sont représentés que par les états stables d'un réseau, ses attracteurs. Ses états transitoires, beaucoup plus nombreux, sont souvent inobservables, au moins dans les conditions naturelles, en l'absence de montage expérimental *ad hoc*. Or de nombreuses structures de connexions différentes peuvent générer les mêmes attracteurs, et le nombre de telles structures augmente très vite avec le nombre d'automates interconnectés. On peut voir

facilement⁶ que si N est ce nombre, le nombre de structures possibles est $2N^2$ si chaque connexion ne peut prendre que deux valeurs, 0 si elle n'existe pas, 1 si elle existe; et, de façon générale, pN^2 si chaque connexion peut être pondérée et que la force de chaque connexion peut avoir p valeurs. Le nombre d'attracteurs différents possibles d'un réseau de N automates ne peut pas être plus grand que le nombre d'états, soit $2N$ s'il s'agit d'automates binaires et qN si chaque automate peut occuper q états. Ainsi, de façon générale, le nombre de structures de réseaux est beaucoup plus grand que celui des états et, *a fortiori*, des attracteurs, pour un nombre donné d'automates. On peut concevoir les cas où q soit beaucoup plus grand que p (plus précisément, si $q > pN$) de telle sorte que le nombre d'états possibles soit plus grand que le nombre de structures possibles du réseau. Cela implique une précision très grande dans l'observation, en sorte que le nombre de faits discriminables devienne aussi très grand. Ces cas se rapprochent alors de ceux où les états observables peuvent être décrits par des systèmes d'équations différentielles avec de petits nombres de variables aux solutions continues (c'est-à-dire où le nombre d'états possibles est *a priori* infini). Dans ces conditions, évidemment, la sous-détermination des théories par les faits peut être réduite et même disparaître. Mais ces conditions sont beaucoup plus rarement réunies dans l'observation de systèmes naturels que dans celle de systèmes artificiels montés au laboratoire de façon telle que les variables soient séparées en variables pertinentes observées et non pertinentes (contrôlées) et que la précision la plus grande soit atteinte dans l'observation des variables pertinentes. C'est la raison pour laquelle la modélisation par systèmes d'équations différentielles est souvent peu utile quand il s'agit de systèmes naturels comportant plusieurs processus couplés, car elle exigerait, pour que toute sa force prédictive soit utilisée, des données empiriques sur les états observables et les interactions qui ne sont

⁶ Henri Atlan, « Automata Network Theories in Immunology: Their Utility and Their Under-Determination », *Bulletin of Mathematical Biology*, vol. 51, n° 2, 1989, p. 247-253.

en général pas accessibles. La modélisation par réseaux d'automates discrets permet de ne pas renoncer à toute modélisation quantitative pour autant, mais le prix à payer est que le nombre des solutions possibles est le plus souvent beaucoup plus petit que celui des modèles réalisables avec un nombre donné d'automates. Autrement dit, la dynamique des réseaux d'automates constitue une forme dégénérée de celle des systèmes d'équations différentielles et la sous-détermination des théories par les faits est une conséquence de cette dégénérescence. En ce que cette sous-détermination provient de la méconnaissance des données empiriques qui seraient nécessaires pour la lever, elle exprime bien la complexité de systèmes naturels imparfaitement connus.

Bibliographie

- Atlan, Henri, « Automata Network Theories in Immunology: Their Utility and Their Under-Determination », *Bulletin of Mathematical Biology*, vol. 51, n° 2, 1989, p. 247-253.
- Atlan, Henri, *L'organisation biologique et la théorie de l'information*, Paris, Hermann, 1972, Seuil, 2006.
- Atlan, Henri, Esther Ben-Ezra, Françoise Fogelman Soulié, Didier Pellegrin et Gérard Weisbuch, « Emergence of Classification Procedures in Automata Networks as a Model for Functional Self-Organization », *Journal of Theoretical Biology*, n° 120, 1986, p. 371-380.
- Detienne, Marcel et Jean-Pierre Vernant, *Les ruses de l'intelligence. La Métis des Grecs*, Flammarion, 1974.
- Elkana, Yehuda, « A Programmatic Attempt at an Anthropology of Knowledge », Everett Mendelson et Yehuda Elkana (dir.), dans *Sciences and Cultures*, Dordrecht, Reidel, 1981, p. 1-76.
- Koppel, Moshé et Henri Atlan, « Les gènes : programmes ou données? Le rôle de la signification dans les mesures de complexité », dans Françoise Fogelman Soulié (dir.), *Les théories de la complexité. Autour de l'œuvre d'Henri Atlan*, Seuil, Paris, 1991, p. 188-204.