

Recension de l'ouvrage « A new kind of science » de Stephen Wolfram, éditeur Wolfram Media Inc., 1280 p., 2002

(Publié dans le *Bulletin de l'AFIA*, n°54, avril 2003, pp.52-55)

(Antoine CORNUÉJOLS)

Difficile de parler de cet ouvrage sans commencer par en évoquer l'aspect. D'abord le poids, 2,5kg. Autant qu'un portable, que le manuel de Mathematica du même auteur ou que le livre « Gravitation », livre mythique, à juste titre, sur la théorie de la relativité. Deux fois plus lourd que le « Newton » de Westfall. Ensuite l'intérieur : 1280 pages dont 349 de notes en point 8 qui forment en fait l'essentiel du livre, environ 1000 figures en niveaux de gris et en haute résolution, un index de 15000 termes, ... et pas de section de références. Selon un commentateur, la raison en serait qu'avec une seule page de plus, l'ouvrage s'effondrerait en un minuscule trou noir. Expérience à essayer dans son jardin, mais d'après « Gravitation », le risque est faible. En fait, Stephen Wolfram estime que tout compte fait peu de chose est vraiment à sauver, tout au moins à savoir, de la science d'avant lui, et comme ses notes sont sensées former une synthèse, forcément la meilleure, des différentes disciplines discutées dans l'ouvrage, il n'est pas utile de connaître les références antérieures.

« Gravitation » de Misner, Thorne et Wheeler (1973), livre de même poids rappelons-le, présente un paradigme : la physique comme émanation de la géométrie, expose merveilleusement, selon plusieurs perspectives, les outils conceptuels et techniques associés : tenseurs, invariances, etc., et dresse un bilan des résultats et des confrontations avec l'expérience. Formidablement pédagogique et très plaisant à lire et à étudier, il s'agit d'un ouvrage de référence sur un paysage scientifique déjà bien balisé.

Qu'en est-il de « A new kind of science » ?

D'abord, c'est l'œuvre d'un auteur, d'un chercheur solitaire qui croit avoir trouvé la pierre philosophale et qui ne veut courir aucun risque que l'histoire oublie de faire justice de cette éminente contribution. C'est donc l'exposé d'une thèse, mais qui au lieu de la placer dans un contexte, de reconnaître et saluer les « épaules de géant » qui en forment le socle, choisit au contraire de faire comme si toutes les idées exposées étaient dues, d'une manière ou d'une autre, à l'auteur qui seul en a saisi la portée.

Stephen Wolfram est physicien de formation, mais très vite il a abandonné les arcanes de la physique quantique pour se consacrer à l'étude des automates cellulaires avec l'intime conviction que ces systèmes, apparemment les plus simples, recèlent les clés de la compréhension de l'univers sous toutes ses manifestations. Des plus fondamentales : particules et forces, au plus philosophiques : la nature du libre-arbitre par exemple, en passant

par l'explication de processus cognitifs comme la perception ou l'analogie. Qu'est-ce qui peut l'autoriser à une telle stupéfiante extrapolation ?

Rappelons ce qu'est un automate cellulaire. Il s'agit d'un système dynamique généralement déterministe consistant en un réseau de machines à états finis identiques, ou cellules, qui changent de couleur d'un instant au suivant en fonction d'une règle de mise à jour. Cette règle est appliquée simultanément (en général) à toutes les cellules à des temps discrets. Une cellule particulière calcule sa nouvelle couleur en fonction de sa couleur actuelle et des couleurs de ses voisines. Il existe de nombreuses définitions de ces voisinages, ainsi que de règles de mises à jour. L'un des exemples les plus célèbres est celui du jeu de la vie mis au point par John Conway en 1970. Les cellules y sont placées sur une grille plane à maille carrée et calculent leur état : vivante ou morte (d'où le nom du jeu), en fonction de leur huit plus proches voisines selon une règle très simple. Si les cellules voisines vivantes sont exactement au nombre de deux, alors la cellule maintient son état courant. Si le nombre de voisines vivantes est de trois, la cellule sera vivante à la génération suivante quelque soit son état actuel. Dans tous les autres cas, la cellule sera morte à la génération suivante.

Ce qui est surprenant c'est qu'une structure et des règles aussi simples donnent naissance à des évolutions très complexes. Le jeu de la vie étant devenu un programme très courant, fréquemment utilisé comme nuit d'écran, tout le monde est devenu familier des formes qui naissent, se métamorphosent, se déplacent et parfois meurent sur l'écran (voir les ouvrages « The recursive universe » de William Poundstone, 1985 et « Laws of the game » de Manfred Eigen et Ruthild Winkler, 1981, remplis d'illustrations et de détails sur ce jeu. Voir aussi le site de Paul Rendell et celui de David Griffeth psoup.math.wisc.edu). Ce que l'on sait peut-être moins est qu'il a été démontré que ce jeu peut simuler une machine de Turing et a donc la puissance d'une machine universelle. On profitera de cette occasion pour citer l'extraordinaire livre de Berkelamp, Conway et Guy « Winning ways for your mathematical plays », dans lequel, entre autre mille fascinantes facettes, on trouve une esquisse de cette preuve.

Wolfram, dès le début des années quatre-vingt, a étudié des automates encore plus simples puisqu'ils sont définis sur un réseau linéaire dans lequel le voisinage d'une cellule comporte seulement ses deux plus proches voisines. Cela signifie que l'état d'une cellule est déterminé par l'état de trois cellules (dont elle-même) à la génération précédente. Il y a donc $2^3=8$ contextes possibles et $2^8=256$ règles d'évolution envisageables. L'intérêt de ce type d'automates linéaires est qu'il est aisé d'analyser leur évolution en superposant ligne après ligne sur une même figure les générations successives des cellules. C'est ce qu'a réalisé Wolfram de manière systématique et ce qu'il a obtenu ainsi a transformé sa vie et devrait, pense-t-il, révolutionner la science. L'observation des évolutions des différents automates (partant tous du même état initial) suggère en effet de distinguer plusieurs classes de

comportements. Selon Wolfram, il y aurait la classe 1 de ceux qui évoluent vers un attracteur et dont les cellules conservent leur état après un certain nombre de générations, la classe 2 de ceux qui évoluent vers un processus périodique, la classe 3 de ceux qui exhibent des « trajectoires » qui ont un caractère aléatoire et dont l'apparence indique une transmission d'information à longue distance entre les cellules, et enfin la classe 4 de ceux qui seraient à la frontière entre les comportements des classes 2 et 3 avec des régions apparemment aléatoires jouxtant des régions localement structurées. Deux automates en particulier ont fasciné Wolfram, correspondant aux règles 30 et 110 (voir dans le livre la manière astucieuse qu'utilise Wolfram pour décrire ses règles). La première correspondrait à la classe 3 et serait une sorte de génératrice de hasard (on s'amuse d'ailleurs d'apprendre qu'elle est effectivement utilisée pour cela dans Mathematica). Aucun test connu ne permettant d'y distinguer des régularités. La seconde, la règle 110, étant plus extraordinaire encore car, comme le jeu de la vie mais d'une définition plus simple, elle serait équivalente à une machine de Turing. Le lecteur de « A new kind of science » pourra admirer, grâce à des figures superbement bien choisies et rendues, l'extraordinaire bestiaire de comportements possibles.

Wolfram est donc stupéfait que partant de systèmes déterministes aussi simples une telle richesse de comportements soit possible et puisse inclure quelque chose ayant une puissance de calcul universelle. D'autres avant lui l'avaient vu, citons Emile Post en 1943 avec les fameuses « règles de production », Stan Ulam et John von Neuman, inventeurs des premiers automates auto-reproductibles, John Conway bien sûr, et, en France, Françoise Fogelman, Yves Robert, Maurice Tchuente et Henri Atlan étudiant la dynamique de réseaux booléens au début des années quatre-vingt. Cependant Wolfram n'hésite pas à y discerner là deux conséquences qu'il élève au rang de lois fondamentales.

Premièrement, des règles discrètes extrêmement simples peuvent produire des comportements et des structures aussi complexes qu'il est possible d'atteindre. C'est l'essence de ce qu'il appelle « le principe d'équivalence computationnelle ». Il n'est donc pas besoin d'expliquer le complexe par le complexe. Notons au passage que Wolfram assimile ici le fait que le simple puisse produire du complexe de tout ordre pour laisser entendre que tout complexe doit nécessairement être produit par du simple.

Deuxièmement, les règles simples produisant du complexe sont très abondantes plutôt qu'exceptionnelles comme on aurait pu le penser.

Pour Wolfram, cela amène à penser que l'univers entier n'est que la manifestation du fonctionnement d'un automate très simple (au maximum quatre lignes de code en Mathematica a-t-il conjecturé dans une interview). Cela le conduit aussi à traquer les automates sous-jacents dans tout un ensemble de phénomènes naturels, cognitifs et culturels. Après avoir montré dans les premiers chapitres l'extraordinaire richesse du monde des

automates (Wolfram en explore une énorme variété au-delà des automates linéaires qui servent de base à sa thèse), c'est ainsi qu'il consacre les chapitres 7 à 10 à l'exploration des formes naturelles (que l'on trouve dans la nature, tels les arrangements de couleurs sur les coquillages), les lois fondamentales de la physique et les processus de la perception et de l'« analyse ».

Pour ma part, j'ai trouvé la partie consacrée aux phénomènes cognitifs faible, mal informée et la moins intéressante. Il est visible que, même après quelques coups de téléphones nocturnes à Terry Sejnowski, Wolfram ignore en grande partie ce que sont les sciences cognitives et l'intelligence artificielle. Il en est essentiellement resté à l'image de l'IA des années 70 avec les grands programmes, tels ARCH, AM, ACT* ou SOAR, qui se décrivaient au niveau d'une architecture cognitive et à celle d'une IA désormais entièrement fondée sur l'étude des réseaux connexionnistes. Les modélisations de la perception, essentiellement visuelle, qu'il propose sont attrayantes, mais pas plus que beaucoup d'autres. Il parle également d'apprentissage artificiel sans en utiliser le nom en assimilant apprentissage et compression de données, ce qui d'ailleurs n'est pas très éloigné de la perspective actuelle. Cependant il ignore tout des résultats obtenus sur l'analyse théorique de l'apprentissage durant ces dernières décennies. Il semble en particulier penser qu'il peut exister un système universel et optimal de recherche du meilleur modèle sous-jacent à des données. Mais cela est interdit par le no-free-lunch theorem. Globalement, il ne voit pas que l'apprentissage est un « problème inverse » dans lequel il s'agit de chercher un modèle qui ait pu produire les données. Or les données seules ne peuvent suffire à déterminer le modèle optimal, il faut ajouter des connaissances *a priori* extérieures aux données. Pour en terminer sur ce chapitre, Wolfram affirme que l'intelligence est avant tout liée à l'utilisation efficace et à bon escient d'une énorme mémoire. C'est intéressant, même si d'autres avant lui, comme Roger Schank, l'avaient affirmé avec plus de force encore. Malheureusement il en reste essentiellement là, sans donner la clé permettant cette utilisation si puissante de la mémoire. D'un certain côté, je me demande si, curieusement, la contribution la plus féconde à l'étude de l'IA ne serait pas à chercher dans le chapitre 9 sur la physique fondamentale dans lequel il décrit des réseaux causaux pour rendre compte de la structure de l'espace et des forces entre particules. Même si d'éminents physiciens ont montré l'insuffisance de ce modèle en ce qui concerne la physique (voir par exemple la critique de Scott Aaronson dans *Quantum Information and Computation*, vol.1, 2001), il n'est pas impossible que l'on puisse détourner ce modèle pour l'étude de certains phénomènes de transmission de l'information dans les systèmes cognitifs.

Reste à discuter ce qu'entend Wolfram quand il parle de « nouvelle science » et en quoi cela pourrait concerner par exemple les chercheurs en IA. Wolfram considère que la science traditionnelle est avant tout occupée à chercher des normes, à décrire les contraintes que respectent les systèmes naturels. De ce fait, elles passeraient à côté de la vraie explication des

phénomènes qu'il identifie évidemment à l'identification de l'automate sous-jacent. Selon lui, il faudrait donc arrêter de chercher des modèles globaux dont la complexité est forcément du même ordre que celle du phénomène étudié pour se tourner vers la recherche des règles des systèmes dynamiques sous-jacents qui sont d'après les deux lois fondamentales de Wolfram nécessairement simples. Par ailleurs, Wolfram propose _non, je me trompe, Wolfram ne propose jamais, il affirme_ que pour juger de la complexité d'un processus on se repose essentiellement sur notre capacité d'analyse visuelle puisqu'en effet, toujours selon lui, les techniques d'analyse sont en nombre limité, essentiellement équivalentes dans leur pouvoir d'analyse et que notre système visuel est sans doute le meilleur.

Je vois au moins trois problèmes à cette position. La première est que c'est une chose d'affirmer qu'il existe une règle simple sous-jacente et une autre de la trouver. On se retrouve à nouveau face à un problème inverse redoutable. Wolfram n'est pas le seul à s'être fait le chantre de l'émergence de la complexité à partir du simple. Pour le moment, personne ne sait comment s'y prendre de manière non empirique. Wolfram lui-même ne recule pas, dans son livre, à une exploration systématique de toutes les règles possibles dans certains cas (il a ainsi exploré le nombre hallucinant de 4 294 967 296 règles pour les automates mobiles, chap.3). Mais cette approche est évidemment vouée à l'échec dans le cas général. La seconde tient à la nature de ce que l'on tient pour une explication ou ce que signifie comprendre. Est-ce que connaître, à supposer qu'il existe, l'automate sous-jacent à un phénomène signifie qu'on le comprend ? Certes d'un certain côté on est ainsi capable de le reproduire, mais on n'est plus capable de faire des prédictions dans les cas complexes intéressants, il manque les normes et les contraintes justement (voir sur le sujet de la nature de l'explication, les livres « Prédire n'est pas expliquer » de René Thom et « L'aveuglante proximité du réel » de Michel Bitbol, tout deux dans la collection Champs Flammarion). De plus, et c'est le troisième problème de la nouvelle science, Wolfram insiste sur le caractère justement imprévisible des comportements des automates intéressants. Il faut donc attendre et observer, mais combien de temps ? Et puis pour quoi faire à la fin ? Vérifier que l'on a reproduit l'univers ou le cerveau ? On retrouve d'ailleurs là un débat qui a agité les sciences cognitives il y a une dizaine d'années lorsque l'émergence des réseaux de neurones sapait l'approche traditionnelle fondée sur l'idée de représentation et de raisonnement. Décrire en termes de lois élémentaires (au niveau neuronal) peut-il remplacer ou à tout le moins rendre inutile une description au niveau mental ou psychologique ? Il y a là certainement un débat qui touche crucialement l'étude de la vie artificielle (nous examinerons dans une prochaine rubrique le livre « Vie artificielle. Où la biologie rencontre l'informatique » de Jean-Philippe Rennard, Vuibert Informatique, 2002). Pour finir, comme beaucoup d'autres commentateurs (chercher ANKOS (A New Kind Of Science) sur google), je ne discerne pas vraiment d'idées nouvelles dans le livre de Wolfram, ni de révolution scientifique en germe. Je ne pense pas non plus que ce livre doive provoquer

l'intérêt des chercheurs en IA toutes affaires cessantes. Cependant, c'est un livre riche. D'abord parce qu'il réexamine de nombreux problèmes en essayant d'en rendre compte avec des automates, or cela est souvent très éclairant, presque comme une renaissance. Par exemple les discussions sur le codage, sur la compression de données, sur la nature de la complexité ne sont pas toujours parfaitement satisfaisantes sur le plan académique mais elles font réfléchir à neuf. Les illustrations étayent souvent magistralement ces discussions. Et si on prend le recul nécessaire pour ne pas être dupe de ces « démonstrations par l'œil », on ne peut qu'admirer l'ingéniosité qu'elles recèlent. Ensuite, le livre vaudrait sa lecture rien que pour les notes de fin d'ouvrage qui reprennent des grandes disciplines et, à côté de synthèses souvent brillantes, même si Wolframocentriques, fourmillent de détails sur des concepts, des techniques, des algorithmes. Également, pour ceux qui comme moi sont des utilisateurs heureux de Mathematica, on y trouve des centaines d'exemples de petits programmes magnifiques. Finalement, « A new kind of science » est un très beau livre, superbement réalisé. Cela le met à part et en fait presque un objet d'art.

À la fin de leur célèbre article sur la structure de l'ADN en 1953, Watson et Crick en trois lignes « It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material » annonçaient la révolution de la biologie du génome. Wolfram a lui choisi de marteler son message en 1280 pages et 2,5 kg. L'avenir dira si sa portée sera inversement proportionnelle ou non à cette insistance.

Je tiens à signaler aux lecteurs motivés des ouvrages très intéressants sur des sujets proches : « Cellular automata machines » de Toffoli et Margolus, MIT Press, 1987, et surtout « The origins of order . Self-organization and selection in evolution » (1993) et « Investigations » (2000) de Stuart Kauffman qui valent vraiment le temps passé à les étudier.