

LOGIQUE & CALCUL

Qu'est-ce qu'un objet complexe ?

La profondeur logique de Bennett, notion proposée il y a 35 ans, nous rapproche d'une compréhension scientifique de l'idée de complexité.

Jean-Paul DELAHAYE

Les différentes étapes de l'évolution de l'Univers décrites par la cosmologie semblent montrer qu'il se « complexifie ». De même, les écosystèmes terrestres comportent des éléments de plus en plus nombreux et variés interagissant de façon toujours plus subtile et délicate. À une échelle de temps plus réduite, on évoque la « complexification » des sociétés humaines, des échanges sociaux et commerciaux, des objets technologiques et des théories scientifiques. Mais qu'est-ce qui augmente quand on parle de complexité croissante ? Peut-on mesurer cette « complexité » ?

Nombre d'idées ont été proposées pour définir et évaluer la complexité des objets, des processus et des interactions. L'une des propositions, la profondeur logique de Charles Bennett, retient l'attention générale. Grâce à des résultats accumulés depuis son introduction en 1977, elle constitue la piste la plus sérieuse pour une définition scientifique robuste de complexité structurelle ou de richesse en organisation. La mise au point d'outils qui, dans certains cas, la mesurent, permet d'envisager des applications.

Sans surprise, la théorie du calcul est au cœur de la définition proposée par Ch. Bennett, physicien américain connu pour ses recherches en mécanique quantique et en thermodynamique. Qu'un physicien ait recouru à la théorie du calcul suggère que la formulation de Ch. Bennett, qui s'appuie sur

la notion de programme, s'impose et n'est pas juste une façon pour les informaticiens de se hausser du col !

Selon la vision de Ch. Bennett, la complexité structurelle d'un objet (qu'il dénomme profondeur logique) est la longueur du calcul, mesurée en cycles de machine ou plus généralement en ressources de calcul (temps, mémoire ou parallélisme), nécessaire pour qu'un programme court en produise une description explicite et détaillée.

Le résultat d'un programme court, mais d'un long calcul

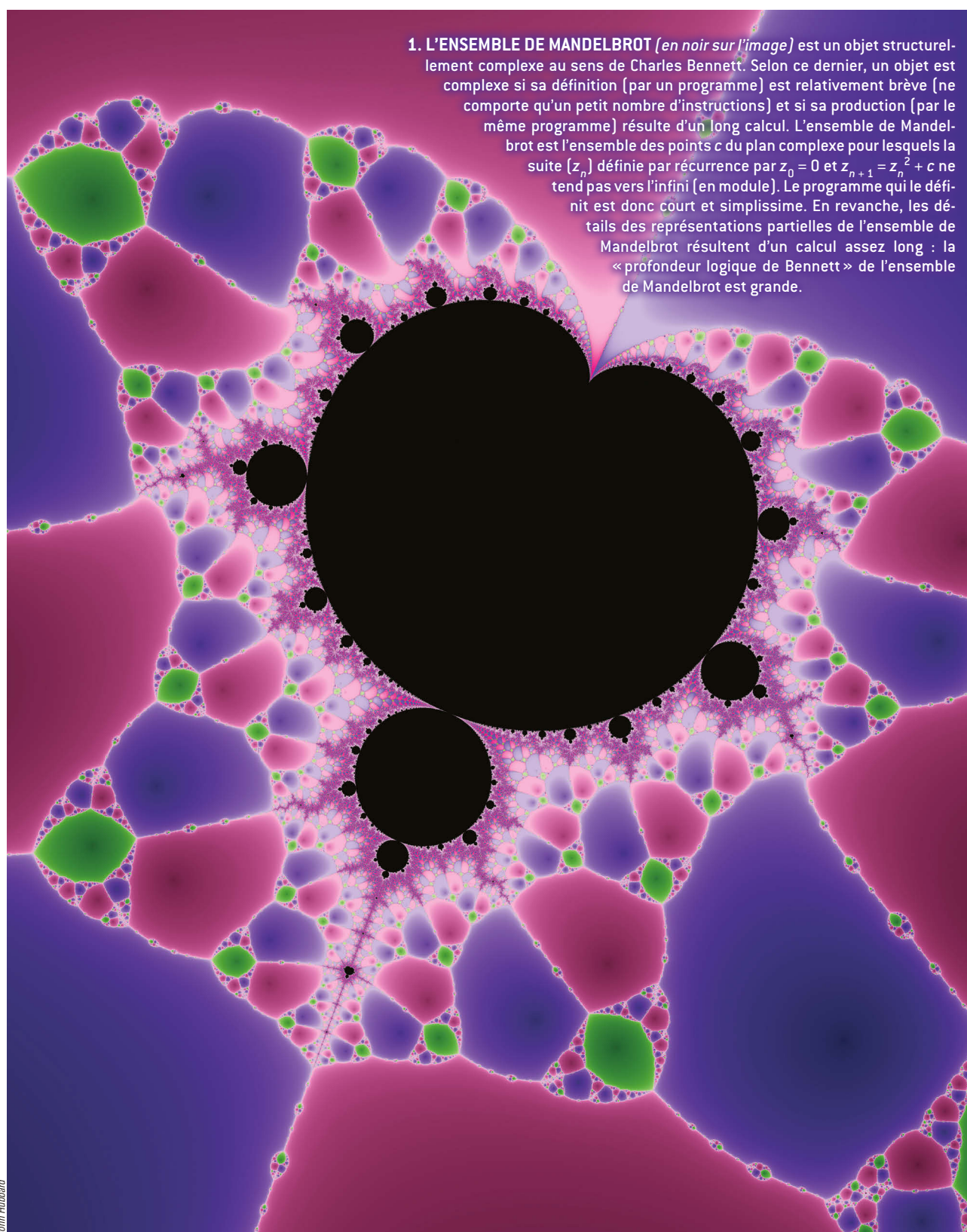
Quand la description complète d'un objet est produite rapidement par un programme court (comportant peu d'instructions), les objets ne sont pas complexes. C'est le cas par exemple pour un bloc de cristal, où le même motif élémentaire – la maille – se répète dans toutes les directions. C'est aussi le cas d'une suite numérique périodique comme 010101...01 (un million de fois). Ces deux objets ne sont pas complexes. Les objets de structure simple possèdent des programmes minimaux courts et ceux-ci produisent leurs résultats sans mettre en œuvre d'importantes ressources de calcul. De tels objets ont une faible « profondeur logique de Bennett ».

Il existe une seconde catégorie d'objets pauvres en structures. Une suite s'indiquant

le résultat d'un million de tirages à pile ou face est faiblement organisée et n'est donc pas structurellement complexe. Un programme court qui en permet l'écriture sera nécessairement de la forme « imprimer s ». Un tel programme s'exécute lui aussi rapidement, car il ne contient aucune boucle ou partie longue à faire fonctionner. Un objet purement aléatoire aura donc une faible profondeur logique de Bennett.

Pour être authentiquement organisé ou structurellement complexe, un objet doit pouvoir être décrit assez brièvement par un programme, ou quelque chose qui y ressemble, dont le fonctionnement est long. On a des raisons de croire que c'est le cas d'un microprocesseur, d'une ville, d'un mammifère, qui sont parmi les objets les plus structurellement complexes que nous connaissons.

Lorsqu'on attribue une grande complexité structurelle à un objet, c'est qu'on voit ou croit voir en lui les traces d'une longue histoire causale assimilable à un calcul. Cette histoire y a forgé les relations qu'entretiennent ses différentes parties et qui constituent sa complexité d'organisation. Si l'on accepte l'idée qu'il est beaucoup plus probable que l'objet qu'on qualifie de structurellement complexe provienne d'une dynamique longue, mais définissable assez simplement, plutôt que d'un processus bref qui l'aurait engendré par chance, alors l'idée de Ch. Bennett de prendre en compte le temps de calcul des programmes les plus



courts qui engendrent l'objet apparaît naturelle et conforme au bon sens.

Avant de préciser comment l'idée de Ch. Bennett se traduit mathématiquement, formulons encore un argument en sa faveur à l'aide d'une expérience de pensée. Imaginez que l'on découvre, écrits sur un tronc d'arbre, un milliard de chiffres binaires consécutifs du nombre π matérialisés par des traits longs (pour les 1) et courts (pour les 0). Que faut-il raisonnablement en penser ?

(a) Que cette suite, qui n'en est qu'une parmi $2^{1000000000}$ autres de même longueur, se trouve là par hasard, du fait des intempéries ou du travail aveugle de quelques insectes. Ce serait par pure chance que les traits écrivent un milliard de chiffres binaires de π .

(b) Qu'un processus particulier, physique ou impliquant des humains et un ordinateur, a effectué le calcul de ce milliard de chiffres, qui a ensuite été recopié là. Dans ce cas, l'origine des chiffres serait attachée à un programme assez court ou, ce qui revient au même, à quelques lois physiques dont la description serait brève comparée au milliard d'éléments inscrits sur le tronc d'arbre.

Si vous optez pour (a), vous choisissez une conception proche de celle des biologistes d'avant Pasteur, qui défendaient l'idée de la génération spontanée, soutenant par exemple que des rats naissent parfois d'un tas de chiffons humides abandonnés sans qu'aucun rat n'y soit préalablement passé. L'hypothèse (b) correspond à l'idée plus raisonnable, du moins aux yeux de la science moderne, que l'origine des rats dans un tas de chiffons est la même que celle des autres rats : ils sont nés de leurs parents, et sont, comme eux, le produit de l'évolution biologique qui, depuis plus de trois milliards d'années, a élaboré des formes vivantes aux structures de plus en plus riches et délicates. Cette évolution fonctionne selon les lois de la physique et de la chimie, lesquelles font interagir les atomes et molécules, et ces interactions sont assimilables à un calcul.

Pour les rats du tas de chiffons, comme pour le milliard de chiffres de π de notre expérience de pensée, la complexité struc-

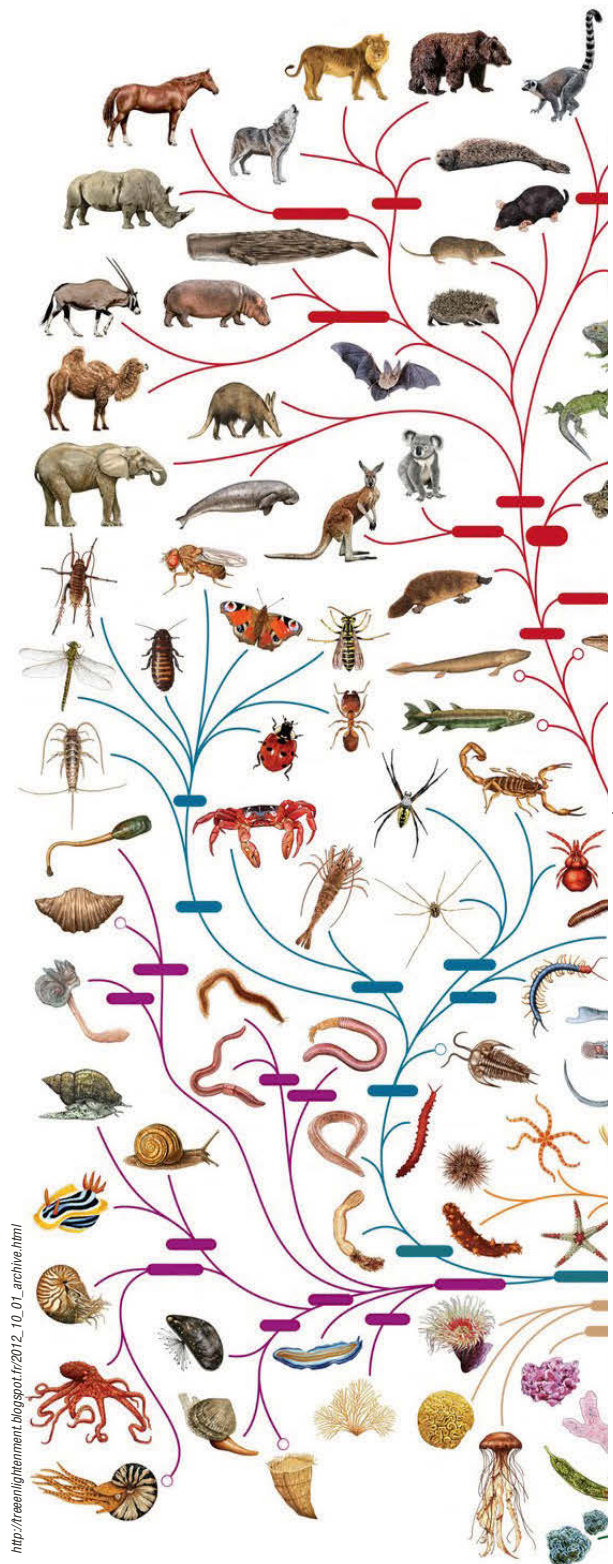
turelle provient de toute évidence d'un processus causal long, mais définissable assez brièvement. En termes informatiques, on dira que ces objets sont le résultat du calcul « d'un programme court ayant fonctionné longtemps ». À la condition d'adopter une conception large de ce qu'est un calcul, l'idée de Ch. Bennett correspond à l'option (b) et elle est donc naturelle. Cette façon large de concevoir le calcul est assez facile maintenant, puisque les développements de l'informatique nous montrent au quotidien que tout processus dynamique peut être simulé par un calcul (ou, à l'inverse, peut être un moyen d'opérer des calculs).

Définition mathématique

L'idée générale de Ch. Bennett est rendue mathématiquement précise de plusieurs façons, dont les plus élémentaires conduisent aisément à des théorèmes. La première version mathématique de la définition de Ch. Bennett consiste à poser que la profondeur logique $P[s]$ de la suite finie s de 0 et de 1 (tout objet physique de taille finie se ramène, par numérisation, à une telle suite) est le temps de calcul (mesuré en nombre de pas de calcul) que le programme le plus court (on le note s^*) qui produit s met pour fabriquer s (s'il existe plusieurs programmes de même taille, on choisit le plus rapide).

Bien sûr, le temps de calcul dépend du langage de programmation utilisé. Cependant, à la condition de choisir un langage assez général (et ceux utilisés en informatique le sont : *C*, *Java*, *Basic*, *Python*, etc.), la définition de la complexité structurelle proposée par Ch. Bennett varie peu. Cette invariance est bien évidemment centrale et, grâce à elle, la notion de profondeur logique peut être prise au sérieux.

La détermination du plus court programme s^* engendrant s est très difficile et ne réussit pas toujours ! Ce n'est pas très étonnant, car les logiciens ont montré que le passage de s à s^* est non calculable : il n'existe aucun algorithme permettant, pour toute suite s , de connaître le programme s^* et son temps de calcul $P[s]$. Cette indécidabilité algorithmique provient de celle



2. LES OBJETS AYANT UNE RICHE ORGANISATION ne surgissent pas de rien et, selon Charles Bennett, ils vérifient une « loi de croissance lente ». Il faut une longue suite de causalités, assimilables à de nombreux calculs, pour engendrer ou décrire de tels objets. Ch. Bennett a démontré que la



profondeur logique d'un système en évolution vérifie la loi de croissance lente (d'après laquelle la complexité structurelle ne peut augmenter rapidement), ce qui constitue un argument fort en faveur de sa définition. Cette idée s'appliquerait en particulier à l'évolution des organismes vivants.

démontrée en 1936 par Alan Turing et qui concerne l'arrêt des programmes : aucune méthode d'analyse ne permet de savoir à l'avance si un programme va se mettre à boucler.

Cependant, il ne faut pas tirer de cette impossibilité d'un calcul exact systématique pour tout s de $P(s)$ la conclusion qu'on ne peut pas évaluer $P(s)$. Les algorithmes de compression de données sont des outils qui calculent des versions approchées de s^* et leur utilisation permet l'évaluation de $P(s)$.

De la même façon, la complexité de Kolmogorov de la suite s , qui est la longueur (mesurée en bits) de s^* , n'est pas calculable, mais, grâce aux algorithmes de compression de données, elle est évaluable. Cette possibilité a permis des avancées dans la classification automatique des textes, des partitions musicales et des séquences génétiques. La non-calculabilité n'est pas un obstacle rédhibitoire : nous faisons des mathématiques tout en sachant qu'il n'y a pas de système parfait d'axiomes qui élimine toutes les propositions indécidables.

Complexité aléatoire et complexité structurelle

La complexité de Kolmogorov d'une suite, notion très importante en informatique théorique (et maintenant en physique et en biologie), ne pouvait pas mesurer la complexification évoquée en cosmologie et en biologie. En effet, une suite aléatoire où chaque élément est tiré à pile ou face a une complexité de Kolmogorov maximale (parmi les suites de même longueur) alors que, nous l'avons déjà remarqué, aucune structure n'y est présente en général.

La complexité de Kolmogorov mesure le « contenu en information » d'une suite s , c'est-à-dire la quantité minimale d'information qu'il faut mémoriser pour être capable de reproduire s (même si cette information est sans intérêt, comme c'est le cas pour une suite de tirages à pile ou face). La complexité structurelle, que la profondeur logique de Bennett veut évaluer, est liée à la richesse d'organisation, au nombre de corrélations et d'interactions apparentes ou cachées entre parties de s , à la quan-

tité d'interdépendances des parties, toutes choses absentes dans une suite aléatoire de 0 et de 1.

Cette distinction que nous faisons tous sans en avoir nécessairement conscience entre complexité aléatoire (ou contenu incompressible d'information que mesure la complexité de Kolmogorov, la taille de s^*) et complexité structurelle (celle que Ch. Bennett veut évaluer, et qui en première approximation est le temps de calcul de s^*) est essentielle pour que puisse progresser une théorie générale de la complexité. C'est faute de l'avoir comprise plus tôt et de s'être aperçu que l'impossibilité d'un calcul exact des versions formelles de ces notions (complexité de Kolmogorov et profondeur logique) n'interdisait pas de les utiliser, que l'on a cherché – et que l'on cherche encore parfois – à formuler des définitions particulières de la complexité qui ne s'appliquent que dans des contextes limités.

Des versions mathématiques plus avancées de la profondeur logique de Bennett ont été proposées. Elles ne se contentent pas de faire intervenir le programme le plus court produisant s , mais tous les programmes produisant s qui ne peuvent être compressés de plus de k bits ou même tous les programmes produisant s , mais en accordant plus de poids aux programmes les plus courts qui sont considérés comme des origines plus probables de l'objet s . Ces versions ont de meilleures propriétés que $P(s)$, dont celle d'autoriser la démonstration d'une « loi de croissance lente » (la complexité structurelle ne peut pas croître rapidement). Malheureusement, ces définitions sont plus délicates à mettre en application que la définition de base $P(s)$.

Une série d'avancées dans la compréhension et l'utilisation de la profondeur logique de Bennett ont eu lieu depuis sa définition en 1977. Il faut les voir comme la confirmation que l'idée de Ch. Bennett est pertinente et que, même si rien n'est très facile dans ce domaine soumis à de nombreuses « indécidabilités », il faut poursuivre son étude. La profondeur de Bennett, ne serait-ce que d'une manière abstraite, nous aide à mieux cerner la complexité dont notre compréhension intuitive

Évaluation expérimentale de $K(s)$ et $P(s)$

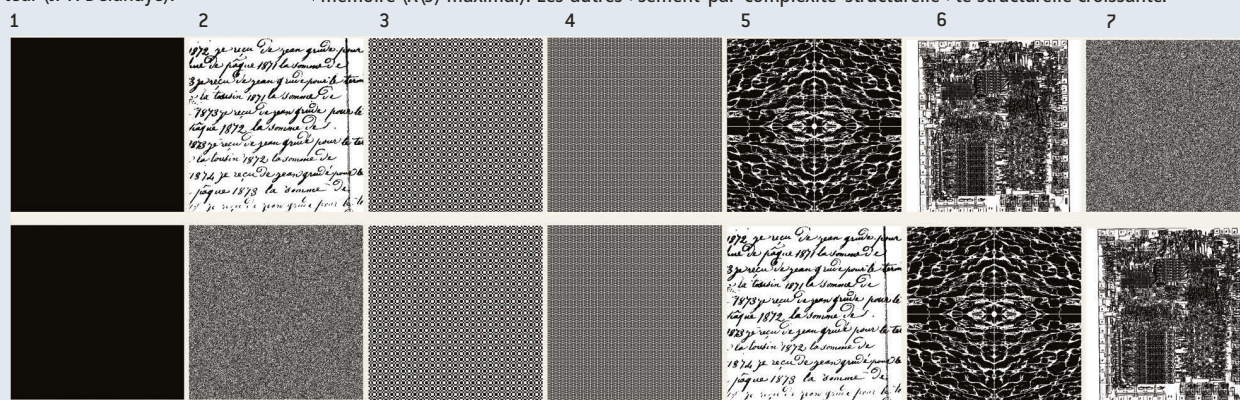
Un programme de compression de données permet d'évaluer la complexité de Kolmogorov et la profondeur logique de Bennett. La version compressée d'un fichier doit être vue comme un court programme engendrant le fichier : sa taille indique une valeur approchée de la complexité de Kolmogorov, et le temps nécessaire à la décompression du fichier est assimilable au temps de calcul pour reproduire le fichier initial, donc à une évaluation de sa profondeur logique. Les expériences ont été réalisées par Hector Zenil, Cédric Gaucherel et l'auteur (J.-P. Delahaye).

Sept images de même format ont été compressées avec un algorithme de compression sans perte : l'image retrouvée après décompression est exactement celle avant compression. La première série indique le classement des images par ordre croissant de taille du fichier compressé, donc par complexité de Kolmogorov $K(s)$ croissante. Sans surprise, on trouve que l'image toute noire (image 1) a le contenu en information le plus petit, et que l'image composée de pixels tirés aléatoirement (image 7) est celle qui exige la plus grande quantité de mémoire ($K(s)$ maximal). Les autres

images sont classées à peu près comme on s'y attend : le manuscrit (image 2) demande assez peu de mémoire, car il y a beaucoup de blanc ; un réseau périodique (image 3) ; une courbe de Peano (image 4) ; une image irrégulière, mais avec deux axes de symétrie (image 5) ; un microprocesseur (image 6).

La seconde série d'images reprend les mêmes sept images, mais cette fois par ordre croissant de temps de décompression, ce qui donne des valeurs approchées de la profondeur logique de Bennett $P(s)$, donc un classement par complexité structurelle

croissante. L'image 7 du premier classement, qui est parfaitement aléatoire, est maintenant parmi les premières, conformément à l'idée qu'un hasard parfait est totalement sans structure. Le microprocesseur est bien identifié comme le plus complexe structuralement. La courbe de Peano est, sans surprise, toujours considérée comme contenant des structures un peu plus riches que le motif périodique. Le texte écrit et le motif symétrique changent de position, donnant au total un classement compatible à ce qui, intuitivement, correspond à une complexité structurelle croissante.



doit maintenant devenir mathématique et scientifique.

Parmi les tentatives de définition d'une mesure de complexité structurelle, certaines, issues de la théorie du calcul, sont différentes de celle de Ch. Bennett. Elles sont en compétition avec l'idée de Ch. Bennett et il est donc nécessaire de comprendre s'il s'agit de notions liées.

La notion de « sophistication » introduite par l'Israélien Moshe Koppel et le Français Henri Atlan, et la notion de « complexité effective » des Américains Seth Lloyd et Murray Gell-Mann se fondent sur l'idée d'une séparation possible entre la composante aléatoire et la composante organisée d'un objet. L'idée des définitions proposées par ces chercheurs sous des formes mathématiques qui en masquent parfois le côté naturel est la suivante. Quand on examine

un programme produisant un objet, il est souvent possible d'y reconnaître deux parties : l'une représente ce qui fixe et décrit la structure de l'objet, l'autre est ce qui habille la structure par des éléments de moindre importance, voire aléatoires.

La partie non aléatoire

Considérons par exemple une suite finie de paires de chiffres binaires 00 ou 11 dont la succession est aléatoire :

$s = 00\ 11\ 11\ 00\ 00\ 00\ 11\ 00\ 11\ 00\ 00\ 11\ 00\ 11\ 11$. Le plus court programme s^* produisant cette suite comportera deux parties assez faciles à identifier, car il sera de la forme : « Imprimer, en doublant chaque chiffre, la suite 011000101001011. »

La structure régulière de s est exprimée par la partie « Imprimer en doublant chaque

chiffre ». La composante désordonnée se trouve, quant à elle, dans la partie descriptive et incompressible de 0 et de 1 qui termine le programme.

Lorsqu'on aura affaire à une suite s plus structurée que celle de notre exemple, la longueur de la partie non aléatoire sera supérieure. C'est cette longueur qui, selon M. Koppel, H. Atlan, S. Lloyd et M. Gell-Mann, mesure la richesse en organisation, la complexité structurelle de la suite s .

La partie du programme où l'on a inscrit la séquence 011000101001011 est d'autant plus longue que la séquence s comporte beaucoup d'éléments aléatoires. C'est la partie aléatoire de la séquence et sa longueur est, en gros, la complexité de Kolmogorov de s .

Des difficultés techniques rendent délicate la formulation mathématique de cette idée de séparation en deux parties d'un pro-

gramme court donnant s , mais ce que nous venons d'expliquer est l'essence de l'idée de la sophistication et de la complexité effective.

Ces définitions étant sans doute aussi naturelles que la définition de la profondeur logique de Bennett, la question était posée de savoir s'il existe un lien entre ces deux visions en apparence très différentes de la complexité structurelle. Une réponse a été donnée en 2010 dans un remarquable article de Nihat Ay, Markus Müller et Arleta Szkola, de l'Institut Max Planck de Leipzig. Ces chercheurs y établissent formellement qu'une grande complexité effective entraîne toujours une grande profondeur logique, confirmant donc le bien-fondé de la définition de Ch. Bennett.

Les recherches récentes ont conduit à un second progrès de la théorie, avec la mise au point de versions calculables de la profondeur logique. Ces versions font des hypothèses simplificatrices et restrictives sur les processus de calcul, mais aboutissent à des définitions débarrassées de la gênante incalculabilité de $P[s]$.

Un troisième motif de satisfaction pour ceux qui pensent que le concept de profondeur logique de Bennett ou les concepts voisins sont une clef pour comprendre le monde réel a été de le voir reconnu dans des travaux scientifiques liés à ce qui, dans le monde, est une sorte de grand calcul : l'évolution biologique.

Antoine Danchin, de l'Institut Pasteur, Guillaume Baptist, de l'Université de Grenoble, et John Collier, de l'Université de Durban en Afrique du Sud, défendent tous les trois indépendamment que la profondeur logique de Bennett permet d'éviter certaines considérations trop naïves sur la complexification des êtres vivants et éclaire l'idée d'un progrès dans les lignées d'organismes vivants, progrès qui ne serait que l'accumulation de résultats de calculs.

Divers travaux de nature expérimentale ont conduit à des débuts de validation pratique de l'idée de Ch. Bennett dont les fondements intuitifs ne font pas de doute, mais dont il restait à montrer qu'elle a aussi un potentiel d'application.

Un premier travail a consisté à mesurer vraiment le temps de calcul du plus court

programme produisant une image. Il a été mené par Hector Zenil, de l'Université de Sheffield, Cédric Gauchere, du Département d'écologie de l'Institut français de Pondichéry en Inde, et moi-même. La méthode évalue le temps de décompression nécessaire pour des images représentant divers objets plus ou moins structurés. Ce temps de décompression est assimilable à $P[s]$, car le fichier compressé de s est une version approchée de s^* et le calcul pour passer du fichier compressé au fichier complet est assimilable au calcul que s^* effectue pour produire s , c'est-à-dire $P[s]$.

Expérimentations

Les résultats de ces mesures de temps de calcul ont été utilisés pour ordonner les images par complexité structurelle croissante (voir l'encadré page ci-contre). La classification obtenue est proche de celle attendue, avec au début les images parfaitement simples ou totalement aléatoires, et à la fin du classement les images des objets les plus riches en structures.

Un second travail fondé sur l'énumération massive de petites machines de Turing a été mené par Fernando Soler-Toscano, de l'Université de Séville, H. Zenil, Nicolas Gauvrit, de l'Université Paris VII, et moi-même. Il a permis d'évaluer la complexité de Kolmogorov et la profondeur logique de Bennett pour de courtes séquences de 0 et de 1. Ces calculs montrent, comme on s'y attendait, que les deux notions de complexité sont indépendantes et que la profondeur logique de Bennett est faible aussi bien pour les suites simples (dix fois 0, par exemple) que pour les suites d'apparence aléatoire.

L'idée de la complexification, objet de nombreux commentaires et d'autant de spéculations, devient ainsi un thème scientifique et mathématique fondé sur une théorie générale rigoureuse et qui, dans les cas les plus simples, produit des outils de mesure. On peut même rêver d'aboutir à une compréhension théorique profonde du développement de la complexité de l'Univers et du vivant, qui rendrait définitivement caduques les élucubrations des créationnistes et de l'*Intelligent Design*. ■

L'AUTEUR



J.-P. DELAHAYE est professeur à l'Université de Lille et chercheur au Laboratoire d'informatique fondamentale de Lille (LIFL).

BIBLIOGRAPHIE

H. Zenil, J.-P. Delahaye et C. Gauchere, *Image characterization and classification by physical complexity*, *Complexity*, vol. 17(3), pp.26-42, 2012.

N. Ay et al., *Effective complexity and its relation to logical depth*, *IEEE Trans. Inform. Theory*, vol. 56, pp. 4593-4607, 2010.

J.-P. Delahaye, *Complexité aléatoire et complexité organisée*, Quæ, 2009.

Ch. Bennett, *Logical depth and physical complexity*, dans *A Half-Century Survey on the Universal Turing Machine*, pp. 227-257, Oxford University Press, 1988.

M. Koppel, *Complexity, depth, and sophistication*, *Complexity*, vol. 1, pp. 1087-1091 1987.

SUR LE WEB

F. Soler-Toscano et al., *Correspondence and independence of numerical evaluations of algorithmic information measures*, 2012 : <http://arxiv.org/abs/1211.4891>

Ch. Bennett, *What increases when a self-organizing system organizes itself ? Logical depth to the rescue*, 2012 : <http://dabacon.org/pontiff/?p=5912>

P. Grassberger, *Randomness, Information, and Complexity*, 2012 : <http://arxiv.org/abs/1208.3459>