Philosophiques

Paul Thagard, *The Cognitive Science of Science,* MIT Press, Cambridge (MA)/Londres, 2012, 365 p.

Jean-Frédéric de Pasquale and Pierre Poirier

Volume 40, Number 1, Spring 2013

URI: https://id.erudit.org/iderudit/1018387ar
DOI: https://doi.org/10.7202/1018387ar

See table of contents

Publisher(s)

Société de philosophie du Québec

ISSN

0316-2923 (print)
1492-1391 (digital)

Explore this journal

Cite this review

de Pasquale, J.-F. & Poirier, P. (2013). Review of [Paul Thagard, *The Cognitive Science of Science,* MIT Press, Cambridge (MA)/Londres, 2012, 365 p.] *Philosophiques*, *40*(1), 238–243. https://doi.org/10.7202/1018387ar

Bien qu'elles n'aient pas la fougue des textes de jeunesse (1830-1845), les nouvelles traductions d'Anne-Marie Pin offrent au lecteur francophone l'occasion d'apprécier les développements de la pensée de Feuerbach qui s'inscrit de plus en plus dans le primat de la réalité sensible et naturelle, tout en opposant à la morale du devoir ou de l'utilité — autrement dit la morale chrétienne ou bourgeoise — la nécessité de tirer des leçons de la sagesse des Anciens.

EMMANUEL CHAPUT
Université de Montréal

Paul Thagard, *The Cognitive Science of Science,* MIT Press, Cambridge (MA)/Londres, 2012, 365 p.

Il y a quelques mois paraissait dans la revue *Science* un article décrivant Spaun, un cerveau artificiel comportant 2,5 millions de neurones, capable de réaliser huit tâches cognitives, dont une version simplifiée d'un test d'intelligence[17]. Cette prouesse des *neurosciences théoriques*, héritières du connexionnisme d'antan, constitue une bonne occasion de se demander si celles-ci ont quelque chose à apporter à la philosophie. *The Cognitive Science of Science* de Paul Thagard pourrait être la réponse à cette question, et c'est dans cet esprit que nous présenterons son ouvrage. Thagard est l'un de ceux qui (comme les Churchland) ont poursuivi le projet d'une philosophie neurocomputationnelle des sciences et, dans ce livre où il collabore, entre autres, avec certains des membres du laboratoire derrière Spaun, il apporte à l'étude des sciences la nouvelle révolution neurale en sciences cognitives : l'application des neurosciences théoriques aux fonctions de haut niveau du cerveau — dans ce cas-ci, celles qui sont essentielles à la pratique scientifique. Malgré une certaine volonté de multidisciplinarité, les neurosciences théoriques occupent en effet la place centrale dans la version des sciences cognitives de la science proposée par Thagard. Il a développé une approche de la science par simulations neuronales, laquelle permet à la fois de *décrire* les mécanismes cognitifs en jeu dans la pratique scientifique et, par le fait même, grâce à une conception instrumentale de la rationalité scientifique, de mieux cerner les *normes* de la science — deux tâches qui ont traditionnellement occupé les philosophes.

Pour comprendre ces modèles, il faut d'abord dire un mot sur les neurosciences théoriques qui les animent. Cela donnera en même temps une idée des progrès faits par celles-ci et de leur capacité à réanimer les vieux débats philosophiques sur les modèles neuronaux, car de la résolution de ces débats dépend en partie la valeur des modèles de la science de Thagard. Il

17. C. Eliasmith, T. C. Stewart, X. Choo, T. Bekolay, T. DeWolf, Y. Tang, D. Rasmussen (2012). « A Large-Scale Model of the Functioning Brain », *Science*, 338(6111), 1202-1205.

adopte deux développements récents dans le domaine qui sont particulièrement importants ici : les représentations holographiques réduites (HRR) et le cadre d'ingénierie neurale NEF. Les HRR permettent la construction de représentations neurales (des vecteurs d'activité neuronale) structurées, et notamment compositionnelles. En plus de l'opération de somme de deux vecteurs, deux opérations des HRR rendent cette construction possible : la convolution circulaire (⊗), que Thagard décrit comme une sorte d'entortillement de deux vecteurs, et l'opération inverse de corrélation circulaire (#). Ces deux opérations sont complexes, mais il suffit ici de dire que la première permet de joindre deux vecteurs en un seul de même dimension (A⊗B=C) et la seconde de retrouver (à peu près) un des deux vecteurs initiaux (C#A≅B). Un tel formalisme permet de dépasser certaines limites liées à la construction de représentations neurales structurées. Ainsi, les HRR permettent de distinguer entre la représentation neuronale de « Marie aime Jean » (marie⊗sujet + aime⊗verbe + jean⊗complément) et celle de « Jean aime Marie » (jean⊗sujet + aime⊗verbe + marie⊗complément). Un des reproches fait à la tentative de Churchland de rendre compte de la représentation scientifique en termes neuronaux étant précisément son incapacité à rendre compte de la nature propositionnelle des théories[18], c'est là un solide avantage de Thagard. NEF, l'autre outil qui fait de l'approche de Thagard une nouvelle forme de neurophilosophie des sciences, est un cadre d'ingénierie neuronale qui permet de construire analytiquement des réseaux de neurones d'un haut niveau de réalisme biologique. Celui des anciens modèles connexionnistes laissait souvent à désirer ; par exemple, leurs neurones formels communiquaient par des nombres réels dits « taux d'activation ». Dans les nouveaux modèles plus réalistes que permet NEF, les neurones communiquent entre eux par des impulsions, conformément à nos théories neurologiques les plus avancées sur la question. Là encore, cet ajout à la boîte à outils de Thagard règle un problème majeur des modèles neuroépistémologiques, soit l'interprétation mal définie de leurs « unités de traitement ». Ici, ces unités correspondent clairement à des neurones biologiques. Armé de ces avancées, il est possible de construire des modèles neuronaux biologiquement réalistes capables d'effectuer des tâches cognitives complexes, et même, pense Thagard, celles qui sont propres à la pratique de la science.

La première partie du livre traite de trois processus centraux de l'explication : la production d'explications à partir des connaissances actuelles, la génération de nouvelles hypothèses explicatives, et l'évaluation des explications. Thagard passe en revue les façons classiques en sciences cognitives de rendre compte de l'explication, puis, après avoir critiqué plus fortement les approches déductives et probabilistes de l'explication, propose un modèle neural de l'abduction capable de représenter des règles de type « a⊗antécédent

18. C. Glymour, « Invasion of the Mind Snatchers », *Minnesota Studies in Philosophy of Science*, R. Giere (dir.), University of Minnesota Press, 1992.

+ cause⊗relation + b⊗conséquent ». Ce modèle peut, lorsqu'on lui présente l'événement b associé à un signal de surprise, rechercher la règle qui permet d'expliquer b. En ce qui concerne l'évaluation, Thagard met à jour sa théorie antérieure, la théorie de la cohérence explicative, en lui ajoutant une implémentation biologiquement réaliste (par l'intermédiaire de NEF) d'ECHO, un modèle qui utilise des réseaux récurrents pour résoudre des problèmes de cohérence[19]. Thagard utilise ECHO pour comprendre l'évolution des opinions sur le réchauffement climatique. Après avoir présenté ses modèles de processus centraux liés à l'explication, il défend la notion centrale de son épistémologie, la cohérence explicative, contre une objection adaptée d'un argument qu'on retrouve dans le débat sur l'antiréalisme en philosophie des sciences, celui de « l'induction pessimiste ». On observe que plusieurs théories possédant une bonne cohérence explicative se sont révélées fausses et on pourrait faire l'induction que cette cohérence n'est pas une bonne indication de la vérité des théories. Thagard soutient cependant que la cohérence explicative est un bon signe de la vérité lorsque les théories augmentent leur cohérence de deux façons : elles s'*élargissent,* c'est-à-dire expliquent une plus grande variété de faits, mais aussi s'*approfondissent* en parvenant à expliquer les couches successives de mécanismes. De telles théories semblent en effet historiquement plus résistantes, ce qui permet de soutenir qu'elles ont de bonnes chances d'être approximativement vraies, justifiant ainsi le modèle de la cohérence comme modèle normatif.

La seconde partie du livre s'intéresse au processus de découverte en sciences. Thagard soutient que la créativité résulte de la combinaison de représentations, conjecture qu'il valide empiriquement par une analyse de découvertes scientifiques et d'inventions. Il propose un mécanisme pour cette combinaison, la convolution circulaire, puis procède à deux études de cas, pour identifier les processus et situations menant à des combinaisons utiles. La troisième partie traite du changement conceptuel. Après un chapitre qui examine le changement conceptuel en sciences par l'analyse de l'évolution historique de divers concepts, l'auteur s'intéresse à trois cas philosophiquement pertinents qui présentent des changements conceptuels radicaux : la sélection naturelle, le rapport de la médecine occidentale à l'acupuncture et l'évolution de la psychiatrie. Dans la deuxième étude de cas, Thagard identifie les schémas explicatifs de la médecine chinoise et sa structure conceptuelle, montrant qu'elle est incommensurable avec la médecine occidentale : il n'y a pas de traduction directe d'un schème conceptuel à l'autre, et leur cohabitation générerait un conflit de classifications. Néanmoins, l'Institut national de la santé américain (NIH) a réussi à établir un consensus favo-

19. Pour mémoire, la théorie de la cohérence explicative de Thagard est décrite par sept règles spécifiant par exemple qu'une hypothèse est cohérente avec les données qu'elle explique, et que deux hypothèses qui expliquent conjointement certaines données sont cohérentes entre elles.

rable, dans certains cas, à l'acupuncture, en se concentrant sur l'évaluation des effets empiriques : l'incommensurabilité n'est donc pas un obstacle cognitif infranchissable à l'évaluation des théories scientifiques. Dans la quatrième partie enfin, deux essais présentent des pistes de recherche en sciences cognitives de la science. Dans le premier, Thagard examine les bases neurales des valeurs et rejette l'idée d'une science sans valeurs, considérant qu'elle est psychologiquement impossible à mettre en œuvre. Le second soutient que la notion de concept peut être expliquée par celle de « pointeurs sémantiques », des représentations vectorielles symboliques possédant des propriétés qui les rendent aptes à expliquer la nature des concepts en psychologie : 1) ils ont une sémantique superficielle ; 2) ils fonctionnent comme des pointeurs informatiques en ce qu'ils permettent l'accès à des représentations multimodales stockées en mémoire à long terme qui, elles, constituent une sémantique profonde ; et enfin : 3) ils permettent des inférences cognitives complexes. Thagard examine trois concepts scientifiques au moyen de cette notion. Par exemple, le concept de force possède une sémantique profonde sensori-motrice (les actions de « pousser » et « tirer ») à laquelle on peut accéder, mais ce concept possède aussi un caractère quasi-symbolique pouvant servir aux inférences complexes (F=ma). Il s'essaie finalement à une réfutation directe de la thèse de Machery selon laquelle la notion de concept n'a pas assez d'unité pour être conservée en science, en montrant que les pointeurs sémantiques unifient les trois théories principales des concepts (prototypes, exemplaires et théories).

C'est de toute évidence un livre ambitieux, et nous ne pouvons ici passer en revue chacune de ses thèses philosophiquement provocatrices. Trois problèmes retiennent cependant notre attention.

Le rôle de la convolution dans la créativité par combinaison. Parce que 100 % des découvertes analysées impliquent (plus ou moins trivialement) la formation de nouvelles combinaisons de représentations sous forme de *propositions* (comme lorsque Harvey découvre que *Le cœur est une pompe*), mais que seulement 60 % d'entre elles impliquent en outre la création de nouveaux *concepts* par combinaison de représentations (comme celui de « sélection naturelle » créé par Darwin), Thagard opte pour l'hypothèse voulant que la créativité concerne avant tout la création de nouvelles *propositions* par combinaison. La trivialité apparente de cette hypothèse est contrée par le mécanisme qui sous-tendrait la combinaison des représentations en question, la convolution. Or la convolution à elle seule ne permet pas d'exprimer la distinction entre les propositions « A cause B » et « B cause A » : a⊗cause⊗b = b⊗cause⊗a, car la convolution est commutative et associative. On a vu qu'ensemble les trois opérations associées aux HRR permettent une solution à ce problème, mais celle-ci utilise des rôles de type linguistique (sujet, verbe, etc.) qui n'ont pas d'équivalents directs pour les états perceptuels ou émotionnels. Adopter cette solution dans le cadre du modèle de la créativité pourrait donc remettre sérieusement en question la

prétention de Thagard à offrir un mécanisme de composition qui peut s'appliquer à des représentations multimodales ou émotionnelles. On peut imaginer des manières de sauver une forme de l'hypothèse combinatoire, une fois qu'on accepte de complexifier le mécanisme de combinaison, que ce soit en restreignant sa portée aux représentations qui sont des pointeurs sémantiques ou en adaptant la technique des rôles aux représentations multimodales et aux émotions. Mais toutes mènent à l'abandon de l'idée que le seul processus de combinaison en jeu est la convolution.

La contrainte de l'architecture. Thagard lui-même soutient que la cohérence explicative est une norme qui s'applique aux théories scientifiques. Une science cognitive de la science devrait donc intégrer dans un tout cohérent les modèles de chacun des processus cognitifs impliqués dans la pratique de la science. Or si Thagard développe des modèles pour une variété de processus scientifiques, on ne voit pas bien comment ceux-ci peuvent s'intégrer les uns aux autres. Une neuroépistémologie qui souscrirait au précepte méthodologique d'intégration et de complétude offrirait ce que l'on nomme une *architecture cognitive neurale*. Faute de quoi on pourrait avoir, par exemple, un modèle de la découverte générant des hypothèses causales qui ne peuvent être évaluées par les moyens décrits par le modèle de l'évaluation des explications. Et de fait, il semble qu'un problème de cet ordre se présente : l'échelle de temps et le format représentationnel du modèle de la production d'explications proposé dans le livre semblent difficilement compatibles avec ceux du réseau ECHO (et du réseau de neurones à impulsions qui le remplace, NECO) qui sert à déterminer la meilleure explication. L'absence d'une interface entre ces deux modèles est d'autant plus gênante que la structure du problème de cohérence que doit résoudre ECHO est entrée à la main par l'expérimentateur, ce qui, comme l'a noté Glymour[20], réduit la testabilité du modèle. Ce problème disparaîtrait si l'on unifiait la modélisation des processus dans un système intégré assurant l'ensemble des fonctions en jeu. Puisque Thagard travaille avec l'équipe qui a développé l'architecture des pointeurs sémantiques, la prochaine étape de son entreprise sera peut-être de procéder à cette unification.

La cognition distribuée. Enfin, une épistémologie naturalisée de la science qui, pour souscrire à ce principe d'unification des explications, intégrerait ses modèles en une architecture cognitive *neurale* rencontrerait nécessairement la question du caractère situé et distribué de la cognition scientifique. Convenons qu'un phénomène cognitif est situé si et seulement si son explication ne peut faire abstraction de l'environnement du sujet (qu'il s'agisse de ses outils, de son environnement social, ou simplement d'un papier et d'un crayon). Bien qu'écrivant à l'occasion que la connaissance scientifique est « de plus en plus une affaire de cognition distribuée » (p. 190), Thagard fait souvent comme si tout était « dans la tête » : l'essentiel du travail de modéli-

20. *Idem.*

sation et d'analyse recourt à des représentations mentales, identifiées à des configurations d'activité neurale. On peut se demander si c'est là, comme il semble le penser, une première approximation justifiée (la relation entre l'architecture cognitive neurale et son environnement étant simplement additive), ou si, en se concentrant ainsi sur les processus neuraux du chercheur, on ne risque pas de dénaturer fondamentalement l'entreprise scientifique, ce qui serait le cas si la cognition scientifique était située, au sens précisé ci-dessus.

En somme, on lira *The Cognitive Science of Science* comme un aperçu d'un futur possible, et souhaitable, de l'étude rationnelle de la science, un futur où les sciences cognitives contribueraient à l'étude des mécanismes et normes responsables d'un des phénomènes centraux de la culture humaine. Et bien qu'il soit douteux que ce futur comprenne toutes les propositions de l'ouvrage de Thagard, on verra certainement en elles des précurseurs sérieux de cette entreprise.

JEAN-FRÉDÉRIC DE PASQUALE
Université du Québec à Montréal

PIERRE POIRIER
Université du Québec à Montréal

Brian Leiter, *Why Tolerate Religion?* Princeton, Princeton University Press, 2013, 192 p.

Sous un titre provocateur qui suggère que l'auteur souhaite remettre en cause la tolérance accordée aux cultes et aux croyants, le bref ouvrage de Brian Leiter pose une question non seulement légitime mais encore urgente dans le contexte des démocraties libérales qui mettent en œuvre des politiques multiculturelles : pourquoi accorder aux croyants des privilèges sous la forme d'exemptions ou d'accommodements raisonnables que l'on n'octroie pas d'ordinaire aux non-croyants ? La thèse de Leiter est qu'il n'est pas justifié de traiter de manière différente les revendications religieuses et celles qui sont fondées sur des croyances ou des pratiques sociales non religieuses ; la liberté religieuse doit recevoir la même protection que la liberté de conscience en général, ni plus ni moins. En outre, que ce soit dans le cas de revendications religieuses ou dans celui de revendications communautaires laïques, il est illégitime d'accorder des exemptions à la loi, qui doit s'appliquer également pour tous. Bien que le problème ne soit pas nommé, ce qui est visé dans l'argumentation de Leiter, ce sont les théories multiculturalistes qui justifient la pratique juridique de l'exemption uniquement pour les groupes organisés, communautés religieuses ou culturelles, et non pour les individus.

Le chapitre I est consacré à la justification philosophique de la tolérance. Leiter commence par distinguer les arguments instrumentaux et les arguments de principe. Il identifie trois arguments instrumentaux qu'il rejette : l'argument dit « hobbesien » du *modus vivendi ;* l'argument de l'inadéquation