

LE DÉFI ANTI-REPRESENTATIONNALISTE DYNAMICISME ET THÉORIE COMPUTATIONNELLE DE L'ESPRIT

Mathieu MAGNAUDET

*Thèse du laboratoire de sciences cognitives EA487 (Université Bordeaux 2) en collaboration avec le CREA (École Polytechnique, CNRS), présentée le 4 mars 2006.
Mél : mathieu.magnaudet@eurisco.org*

Thesis title — The anti-representationalist challenge: dynamicism and the computational theory of mind.

Titolo della tesi — La sfida antirappresentazionalista: dinamicismo e teoria computazionale della mente.

Titulo da tese — O desafio anti-representacionalista: dinamicismo e teoria computacional da mente.

Titulo de la tesis — El reto antirepresentacionalista: dinamicismo y teoría computacional de la mente.

Composante clé des sciences cognitives contemporaines, l'hypothèse représentationnelle, entendue au sens de la pertinence explicative d'un certain concept de représentation mentale, demeure un objet de débat passionné. Au cours des années 80, le courant connexionniste avait déjà entrepris une première forme de critique du représentationnalisme cognitif. L'analyse montre cependant que ce n'était pas tant la propriété de représentation qui était remise en cause, que certaines déterminations propres à l'hypothèse fodorienne du langage de la pensée. Il en est bien autrement des nouvelles approches dites dynamiques de la cognition que certains n'hésitent pas à qualifier, pour les défendre comme pour les critiquer, d'anti-représentationnalismes cognitifs (voir par exemple : Bechtel, 1998 ; Beer, 2000 ; Chemero, 2000). Cette mouvance recouvre cependant des positions théoriques très diverses, de sorte qu'il est en vérité fort difficile d'identifier avec précision ce qui en fait la spécificité, et sur quels aspects au juste elle se sépare d'une théorie computationnelle de l'esprit. L'ambition de ce travail est donc double. Il s'agit d'une part d'apporter un peu de lumière à ce débat, d'étudier avec autant de soin que possible les concepts en jeu, afin de mieux marquer les lignes d'opposition. Et il s'agit d'autre part d'examiner la possibilité d'une théorie non représentationnelle des phénomènes cognitifs. En somme, au-delà des effets d'annonce, peut-on véritablement parler d'une théorie dynamique, non représentationnelle, de la cognition et, le cas échéant, quelle est sa portée véritable ?

La première partie de notre travail est consacrée à la clarification du concept de système computationnel, ainsi qu'à l'étude de l'articulation entre hypothèse computationnelle et hypothèse représentationnelle. À l'inverse de certaines affirmations un peu hâtives, et en nous appuyant sur les travaux de Giunti (Giunti, 1997), nous montrons qu'il est tout à fait possible de dégager un concept rigoureux de computation et de système computationnel formel. Nous explicitons ainsi l'indépendance du concept formel de computation à l'égard de toute forme de sémantique. Nous soulignons par ailleurs que l'indépendance des concepts de

computation et de représentation, parfaitement reconnue par les pères fondateurs des sciences cognitives, est précisément ce qui a conduit à une complexification particulière du concept de système computationnel, le concept de système symbolique physique, qui articule explicitement hypothèse computationnelle et hypothèse représentationnelle. On peut donc soutenir en principe, et sans contradiction logique, que le système cognitif est un système computationnel, non représentationnel. Nous montrons cependant que la proposition converse est plus difficile à décider et dépend de la forme prise par l'hypothèse représentationnelle. La version de la théorie représentationnelle de l'esprit développée par Fodor est ainsi étroitement liée à une forme de computationnalisme. En première approximation, il est donc raisonnable de supposer que le rejet de l'hypothèse computationnelle, prôné par le dynamicisme, doit au minimum interdire certaines formes d'hypothèses représentationnelles.

Le recours à la propriété de représentation mentale soulève cependant une difficulté particulière. Il est en effet bien connu que le problème majeur posé par les théories représentationnelles de l'esprit est celui de la naturalisation de la propriété de représentation. De nombreux débats se sont noués autour de cette problématique, et de nombreuses formes de théories naturalistes de la représentation ont été proposées¹. On doit pourtant reconnaître qu'aucune ne s'avère pleinement satisfaisante. Il y a donc une pleine légitimité à s'interroger sur la pertinence d'une approche représentationnelle des phénomènes cognitifs. Nous avons donc voulu expliciter les raisons qui militent en faveur d'un représentationnalisme cognitif ; notre ambition n'était pas tant, en effet, d'en invalider telle ou telle forme, que d'examiner la possibilité d'un anti-représentationnalisme. Or, ceci exige que l'on éprouve l'aptitude d'une théorie non représentationnelle de la cognition à rendre compte des phénomènes qui justifient le recours à la propriété de représentation. Nous nous sommes donc attachés à dégager les éléments théoriques et empiriques qui sont invoqués en faveur d'une hypothèse représentationnelle. Nous avons en particulier isolé deux classes de comportements que nous qualifions, à la suite de Clark et Toribio (Clark et Toribio, 1994), de « gourmands en représentations ». Ces derniers se caractérisent soit par une forme de découplage avec l'environnement, c'est-à-dire qu'ils ne sont pas dirigés par des stimuli immédiatement présents, soit par une sensibilité à des catégories non nomiques, c'est-à-dire qu'ils sont gouvernés par une catégorie d'objets qui ne possèdent aucune détermination physique commune. Ces deux classes de comportements constituent, nous semble-t-il, une première base d'évaluation pour toute théorie du comportement cognitif qui se présenterait comme non représentationnelle.

La deuxième partie de notre travail est consacrée à l'étude des approches dynamiques, leurs fondements théoriques et leurs critiques du représentationnalisme cognitif. En nous appuyant sur une analyse du concept physico-mathématique de système dynamique, nous montrons que l'hypothèse selon laquelle le système cognitif est un système dynamique est triviale et ne dit rien sur la nature des systèmes cognitifs ; mieux encore, elle est en vérité parfaitement compatible avec une théorie computo-représentationnelle, sauf à réduire le concept de système dynamique à celui de système d'équations différentielles, ce qui ne se justifie pas d'un point de vue mathématique. Ainsi, contre l'opinion de Van Gelder (Van Gelder, 1998), et à la différence de l'hypothèse du système symbolique physique de Newell et Simon

¹ On peut citer, à titre d'exemples, la théorie des rôles conceptuels, la théorie de la causalité asymétrique, ou la théorie téléo-informationnelle.

(Newell et Simon, 1976), nous soutenons que l'hypothèse dynamique ne saurait être considérée comme une loi de structure qualitative, puisqu'elle échoue précisément à circonscrire la classe des systèmes cognitifs. L'étude de quelques théories cognitives, d'inspiration dynamique (cybernétique, fonctionnalisme dynamique, théorie des patrons dynamiques, théorie de l'enaction) montre cependant que certaines hypothèses sur la nature des phénomènes cognitifs peuvent s'avérer incompatibles avec, ou peu propices à, une interprétation computationnelle. Elles justifient en revanche pleinement le recours au formalisme spécifique de la théorie mathématique des systèmes dynamiques. Nous proposons donc d'employer l'expression d'« hypothèse cognitive dynamique » pour désigner toute hypothèse sur la nature des systèmes cognitifs dont la mise en œuvre requiert le recours au formalisme spécifique de la théorie des systèmes dynamiques.

Parmi ces hypothèses, nous en étudions deux en particulier, l'enaction et la cognition expérientielle, qui proposent une critique circonstanciée du représentationnalisme cognitif. L'analyse montre cependant que ces deux théories ne peuvent être à proprement parler qualifiées d'anti-représentationnalisme. Elles retiennent en effet un concept minimal de représentation et contestent essentiellement une forme de réalisme représentationnel attaché, selon elles, au cognitivisme classique. En vérité, on trouve difficilement l'expression d'un anti-représentationnalisme radical dans la littérature dynamiciste sur les fondements des sciences cognitives. Pour cela, il faut se tourner vers un courant particulier de l'intelligence artificielle, l'approche animat ou intelligence artificielle située, où se manifeste une véritable volonté de développer une approche non représentationnelle des phénomènes cognitifs. C'est l'étude de ce courant qui occupe la dernière partie de notre travail.

Le succès de l'intelligence artificielle est avancé comme un argument essentiel en faveur de l'hypothèse du système symbolique physique, c'est-à-dire de l'hypothèse selon laquelle le fait d'être un tel système constitue une condition nécessaire et suffisante du comportement intelligent. À ce titre, elle constitue une discipline de choix pour l'examen de la question du représentationnalisme cognitif. Après avoir spécifié la forme particulière d'explication que propose l'IA, nous distinguons deux formes d'anti-représentationnalisme correspondant aux deux étapes que sont la conception et l'explication du comportement d'un système artificiel. Nous montrons ainsi que l'approche fondatrice de R. Brooks se caractérise par un rejet de la méthode représentationnelle de l'IA classique dont nous proposons une caractérisation (voir par exemple Brooks, 1991). Les limites de cette approche, qui ne s'inscrit pas dans le courant dynamiciste, ont cependant été rapidement relevées (pour une critique circonstanciée, voir Vera et Simon, 1993). Il apparaît notamment que l'explication du comportement des créatures de Brooks peut s'effectuer en termes représentationnels. Il semble en effet possible d'attribuer une sémantique cohérente aux différents signaux qui transitent au sein du système. En outre, on a souvent reproché à Brooks de s'en tenir à des comportements trop simples pour être qualifiés de cognitifs. Ce n'est pas le cas en revanche des systèmes conçus par R. Beer (voir principalement Beer, 2003).

Beer s'inscrit explicitement dans le débat sur le représentationnalisme cognitif et propose une série d'expérimentations qui vise à montrer la pertinence d'une approche dynamique et non représentationnelle du comportement cognitif. L'un des intérêts majeurs de son approche réside dans le fait qu'il s'attaque directement aux comportements gourmands en représentations évoqués plus haut. Les technologies employées, réseaux de neurones récurrents et algorithmes génétiques principalement, vont un peu plus loin encore en direction d'un anti-représentationnalisme de la

conception. Mais la nouveauté réside surtout dans l'anti-représentationnalisme de l'explication que ses travaux permettent de développer. L'analyse fine de ces modèles révèle en effet qu'il n'est pas possible de respecter le principe de localisation des représentations. En d'autres termes il n'y a aucun ensemble d'états internes identifiables sur lequel il soit possible de définir une sémantique cohérente. Si l'on accorde que les comportements modélisés méritent le qualificatif de cognitif, alors on doit admettre qu'il s'agit véritablement de modèles de comportements cognitifs minimaux non représentationnels. Nous concluons en soulignant les limites de ces résultats et en notant cependant leur convergence avec certains travaux, en neurosciences cognitives notamment, qui proposent une redéfinition du concept de représentation en termes de contrôle. Une étude reste à conduire pour identifier quelles sont les déterminations du concept de représentation qui demeurent alors pertinentes.

Références bibliographiques

- Bechtel W. (1998). Representations and cognitive explanations : Assessing the dynamicist challenge in cognitive science. *Cognitive science*. 22(3), 295-318.
- Beer R. D. (2000). Dynamical approaches to cognitive science. *Trends in cognitive science*. 4(3), 91-99.
- Beer R. D. (2003). The dynamics of active categorial perception in an evolved model agent. *Adaptive behavior*. 11(4), 209-243.
- Brooks R. (1991). Intelligence without representation. *Artificial intelligence*. 47(1-3), 67-90.
- Chemero A. (2000). Anti-representationalism and the dynamical stance. *Philosophy of science*. 67(2), 553-557.
- Giunti M. (1997). Computation, dynamics, and cognition. Oxford: Oxford University Press.
- Newell A., Simon H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*. 19(3), 113-126.
- Van Gelder T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and brain sciences*. 21(5), 615-665.
- Vera A. H., Simon H. A. (1993). Situated Action: A symbolic interpretation. *Cognitive science*. 17(1), 7-48.

L'auteur

Mathieu Magnaudet a effectué cette thèse au laboratoire de sciences cognitives EA487 (université Bordeaux 2) et au CREA (école polytechnique, CNRS). Il a bénéficié pour ce travail d'une allocation de recherche MENRT. Après avoir enseigné durant deux ans l'épistémologie et la philosophie cognitive à l'université Bordeaux 2 en tant qu'ATER, il travaille actuellement comme chercheur pour Eurisco international à Toulouse.