

Intelligence artificielle : l'apport des paradigmes incarnés

Pierre De Loor¹, Alain Mille², Mehdi Khamassi³

1 Lab-STICC/CNRS – ENIB, Centre Européen de Réalité Virtuelle, 26 rue Claude Chappe 29200 Brest

2 Université de Lyon, CNRS Université Lyon 1, LIRIS, UMR5205, F-69622, France

3 Institut des Systèmes Intelligents et de Robotique, CNRS UMR 7222, Sorbonne Université, UPMC – Univ. Paris 6, 4 place Jussieu, 75005 Paris

Mots clés : constructivisme, éaction, intelligence artificielle, intelligence computationnelle, robotique développementale, neuro-robotique, interactionnisme, renforcement, motivation intrinsèque, simulation, interaction homme-machine

1. Introduction	1
2. Réflexions sur les limites d'une intelligence artificielle incarnée	4
3. Focus sur la cognition artificielle incarnée	7
3.1 De la cognition symbolique à la cognition incarnée et interactionniste	8
3.2 Les grandes familles d'approches incarnées	9
3.2.1 De l'autopoïèse artificielle à la morphogenèse	9
3.2.2 Robotique évolutionnaire	10
3.2.3 Apprentissage par renforcement	11
3.2.4 Neurorobotique	13
3.2.5. Apprentissage développemental	15
3.2.6. Perception active	16
3.2.7. Robotique souple	17
4. Conclusion	18
References	19

1. Introduction

Depuis l'apparition du terme "intelligence artificielle" (IA), la communauté est confrontée à des débats sur son objet ainsi que sur sa possibilité et son potentiel applicatif. Concernant son objet, s'agit-il de rendre une machine "intelligente" ou juste de fabriquer un outil capable de réaliser plus rapidement et plus efficacement des tâches que l'on attribue à l'intelligence ou encore de constituer un couplage cognitif homme-machine à l'intelligence répartie ? Concernant sa possibilité, les prouesses de certains algorithmes repoussent les limites de la calculabilité et de nombreux outils le démontrent. Quelle que soit l'approche retenue, peut-on parler toutefois d'intelligence ? Si l'on considère l'intelligence comme la faculté de ressentir d'agir, d'interagir ou d'être qui est propre à l'humain et plus généralement au vivant, alors le terme "d'intelligence

artificielle” est un oxymore qui peut expliquer le fait qu’il ait fait couler beaucoup d’encre. Une pierre d’achoppement repose sur la différence entre un programme informatique, s’appuyant sur un substrat abstrait (l’information) et s’exécutant sur une machine, statique par essence et le comportement biologique des êtres vivants, reposant sur un substrat concret organique, dynamique par essence. Nous reviendrons en détail sur cette question mais le fait est que, lors de l’énonciation de son célèbre test, Turing parlait bien de machines pensantes et identifiait déjà les défis de la compréhension du lien entre le physique et l’intellectuel (Turing 1950). Depuis, la cybernétique, sous la forme du connexionnisme, est revenue en force par le biais des progrès en neurosciences, notamment sur la plasticité synaptique et sur les règles de renforcement des connexions entre neurones sous-tendant la mémoire associative. La philosophie, la neurophysiologie et la psychologie ont également progressé et mis en avant l’importance de l’action, de l’interaction avec l’environnement dans la construction du sens, sans pour autant réviser clairement la notion d’*intelligence artificielle*. Enfin, les progrès récents en robotique, et la possibilité que celle-ci offre d’étudier la contribution du corps à la résolution de tâches particulières ont fait que le champ qu’on appelle aujourd’hui celui de “l’Intelligence Artificielle” s’est divisé en plusieurs approches. Une partie de la communauté ne se réfère plus à l’objectif initial de Turing d’une machine intelligente et préfère se focaliser sur la mise au point de solutions algorithmiques efficaces pour résoudre de nombreux problèmes tels que le traitement automatique des grandes quantités de données auxquelles est exposée la société d’aujourd’hui : l’intelligence revendiquée est dans la capacité à aller beaucoup plus vite, en prenant beaucoup plus d’informations en compte pour classer, recommander, décider, ... Une autre partie de la communauté a pour objectif d’essayer de contribuer aux problématiques des Sciences Cognitive en se servant des outils et des progrès mis à disposition par le domaine de l’IA - en les étendant, et parfois en les complexifiant ou en les simplifiant - comme modèles computationnels de la cognition chez l’homme ou chez l’animal. On peut enfin noter une autre mouvance qui vise à exploiter ces modèles computationnels pour concevoir des composants informatiques capables de faciliter un couplage assumé entre l’humain et son environnement informatique : ce couplage pose de nouvelles questions éthiques aujourd’hui discutées non seulement au sein de la communauté de la recherche mais, plus largement, dans la société.

Par conséquent, si l’on cherche à positionner les recherches en intelligence artificielle relativement aux sciences cognitives, nous pouvons distinguer de façon pragmatique, quatre grandes tendances : une approche ingénieure, une approche neuro-modélisatrice, une approche incarnée et une approche interactionnelle. Nous prenons le parti de ne pas nous livrer à une analyse critique approfondie de chacune d’elles dans cet article, mais simplement de les mentionner pour en comprendre le positionnement par rapport à l’IA incarnée que nous avons choisi d’analyser. Une première grande orientation concerne les chercheurs qui ont abandonné la référence à la cognition au sens de la pensée humaine ou naturelle. Dans ce cas, ils développent des algorithmes dont l’ambition est d’utiliser les performances de la machine pour résoudre un problème particulier et en optimiser les calculs. Le sens du problème est déporté du côté du programmeur ou de l’utilisateur. Ils prennent ainsi acte des conclusions sur les débats autour de la chambre chinoise (Searle 1980). La machine, en tant que système computationnel, n’a pas accès au sens. Pour ces chercheurs, l’objectif est la performance des algorithmes utilisés comme des outils adaptés à des tâches spécifiques « difficiles » : des marteaux un peu plus sophistiqués, voire très sophistiqués et performants dans certaines applications. Leur construction peut bien sûr s’appuyer sur toute technique algorithmique, y compris des techniques initialement inspirées de la compréhension de la cognition humaine. Cette orientation générale se manifeste par exemple, au travers des nombreux progrès récents en “apprentissage machine” ou “apprentissage statistique” pour la catégorisation de grandes quantités de données (e.g. Dulac-Arnold et al. 2012; Ganascia

et Velcin, 2004) ou en traitement automatique du langage (e.g. Roche 2011), avec de nombreuses applications dans les entreprises de pointes dans le secteur informatique et en particulier du web et des réseaux sociaux (Big Data).

Une deuxième grande orientation regroupe les chercheurs qui s'intéressent à une approche intégrative de simulation fondée sur les connaissances relatives au fonctionnement neurobiologique du cerveau, ou d'une façon plus générale sur le fonctionnement de systèmes biologiques complexes (e.g. Fix et al. 2007). L'ordinateur est un simulateur qui va "reproduire" l'ensemble des phénomènes connus, à une ou plusieurs échelles possibles en espérant que le résultat obtenu et les processus mis en œuvre (e.g. simuler les différentes étapes d'un processus biologique, Bertin et al. 2011) seront similaires à celui observé dans le vivant.

Il ne s'agit pas ici de présenter de façon exhaustive le large éventail de théories et paradigmes en Neurosciences Cognitives, allant des Neurosciences affectives (Panksepp 1998), incluant notamment la théorie des marqueurs somatiques (Damasio et al. 1996), à la théorie du cerveau social (Gazzaniga 1985). Il s'agit plutôt de souligner que nombre de théories computationnelles actuelles en Neurosciences ont vu le jour grâce à des approches venues de l'Intelligence Artificielle – e.g. pour la modélisation de l'apprentissage (Dayan et Abbott 2001) ou des bases de la théorie de l'esprit et du langage s'appuyant sur les neurones miroirs (Arbib 2003) – ou de la Physique – e.g. pour la modélisation du traitement de l'information dans les réseaux de neurones (Churchland et Sejnowski 1992) ou l'élaboration de théories Bayésiennes comme celle de l'inférence active (Friston 2005). Ceci donne lieu à de plus en plus de projets intégratifs de modélisation du cerveau au niveau système, pour tenter de rendre compte à la fois de l'interaction entre aires cérébrales et de leurs effets sur le comportement (e.g. Eliasmith et al. 2012).

Cette tâche est gigantesque et fait d'ailleurs l'objet d'importants programmes de recherches tels que le 'Human Brain Project' en Europe ou le 'Brain Initiative Project' aux Etats-Unis, et souffre de deux lacunes : 1) L'incompréhension des liens entre le physique et l'intellect, déjà identifié par Turing comme étant « le » verrou à lever (ou plutôt un mystère à élucider). Ainsi rien ne permet de prévoir si les éléments actuellement connus sont nécessaires, suffisants ou inutiles. Ces approches auront sans doute le mérite de clarifier certaines de ces questions mais sont hélas souvent présentées comme "la" modélisation d'une machine pensante. 2) Le fond réductionniste sous-jacent oublie, plus ou moins volontairement, de s'intéresser à l'entité "intelligente" en relation avec son monde et en particulier ignore les travaux des sciences cognitives sur l'incarnation de la cognition, au sens de sa dépendance à l'environnement dont elle contribue à la constitution par ses actions (Varela 1993). Ainsi, même si la simulation de modèles issus des neurosciences computationnelles peut constituer une étape nécessaire et décisive avant la validation du mécanisme étudié par son implémentation dans un corps physique tel que celui d'un robot (Khamassi 2014), et même si l'ensemble des phénomènes contribuant au fonctionnement du cerveau pouvaient être reproduits en simulation, faudrait-il encore les immerger dans un environnement similaire à celui des êtres dont on voudrait reproduire la pensée. Ceci restant toujours sous la réserve d'un lien possible entre phénomènes physiques et phénomènes abstraits, nous y reviendrons.

La troisième grande orientation que nous retenons (et sur laquelle cet article mettra l'accent) concerne les chercheurs qui se sont rapprochés des paradigmes *incarnés*, traduction française adoptée par la communauté Informatique/Robotique pour le terme anglais « embodied ». Bien que ce terme pourrait paraître un oxymore dans le contexte robotique, il ne s'agit pas simplement de prendre en compte le corps du robot mais il s'agit plutôt d'une position philosophique selon laquelle l'interaction avec l'environnement est à la base et ne peut pas être détachée de la cognition. Dans un tel cas, les ambitions sont de reproduire des phénomènes de construction ou d'émergence d'une cognition par le biais d'interactions, en tentant d'utiliser les capacités des machines pour palier la complexité spatio-temporelle inhérente aux phénomènes naturels. Ici, il n'est pas question de simuler de la façon la plus précise possible le fonctionnement du cerveau mais de reproduire un

phénomène global de construction cognitive dans l'action. Les modèles sont alors loin de reproduire des performances humaines et on s'intéresse d'ailleurs plus à l'animal qu'à l'humain. Cependant, étant bio-inspirés, leur plausibilité quant à la reproduction de facultés propres aux organismes vivants (adaptation, viabilité, motivation, assimilation, compréhension, sens) est accrue. Une part de ces travaux tente de se détacher du terme 'intelligence artificielle' en lui préférant 'approche animat', 'robotique développementale' ou 'robotique épigénétique'. Ici, l'ordinateur ou le robot sont au service de la compréhension des mécanismes de la cognition et de la circularité individu - environnement. Ils servent de plateformes d'expérimentation, de test et de validation d'hypothèses souvent issues de la biologie et de la psychologie. Ces travaux, à leur tour, inspirent des développements de recherches sur la conception de mécanismes de *couplages* cognitifs entre les humains et des agents numériques, plongés dans une activité commune (apprentissage interactionnel et collaboratif pour la réalisation de l'activité sans modélisation préalable de cette activité). Il s'agit de la quatrième orientation que nous proposons de distinguer. La dernière grande orientation regroupe les chercheurs qui tentent d'améliorer les capacités d'interaction entre la « machine » et « l'humain », en revendiquant le fait qu'il s'agit d'une activité cognitive couplée. Ayant ainsi acté les avancées sur les relations entre corps, ontogenèse et cognition, et abandonné l'idée de 'machine pensante', ils tentent de développer des systèmes permettant une relation sensorimotrice et historique plus forte entre les utilisateurs et les machines (e.g., joint cognitive systems, Hollnagel & Woods 2005). Les nouveaux dispositifs immersifs, l'analyse des traces, les interfaces adaptatives, les environnements virtuels de formation ou les dispositifs de substitutions sensoriels sont des exemples de thèmes développés par ce courant de recherche. Il ne s'agit pas de créer une cognition artificielle mais de favoriser le couplage homme-machine, d'améliorer l'externalisation grâce à la technique. L'outil a toujours permis à l'homme d'améliorer ses capacités cognitives. On peut soit considérer qu'il artificialise l'intelligence humaine, ce qui pose la question de la frontière entre individu et outil (ma main est un outil qui participe à mes capacités cognitives, le marteau n'en serait que le prolongement, et pourquoi pas la machine ?), soit considérer qu'il est constitutif de l'intelligence humaine, auquel cas le terme d'intelligence artificielle n'est plus un oxymore. D'une certaine manière, ce quatrième courant est proche du premier mais y intègre les prémisses prises en compte par le troisième. On retrouverait alors l'idée initiale de l'intelligence artificielle mais en s'adressant au corps plus qu'à l'intellect (Froese et al. 2011).

2. Réflexions sur les limites d'une intelligence artificielle incarnée

Les travaux que nous allons détailler retiennent notre attention parce qu'ils s'inscrivent dans une approche non cognitiviste et se réclament souvent de l'approche incarnée. Ils correspondent en cela à la troisième des grandes orientations de l'IA que nous avons mentionnées précédemment. Néanmoins, la question se pose des limites voir des impossibilités pour ce type d'approche au regard de la gradation qui existe au sein même du paradigme incarné. En particulier, le lien entre matière organique et intelligence, déjà soulevé par Turing est à interroger. Il en va de même pour le pré-donné qu'impose le programmeur à son programme et qui peut aller à l'encontre d'une autonomie au sens de la clôture opérationnelle du paradigme de l'énaction ou à l'encontre d'une évolution possible des entités artificielles pour s'adapter à une niche écologique. Enfin, dans une visée externaliste, que faire de la relation corps/environnement si ce corps n'est pas biologique ? Force est de constater qu'aucune de ces approches ne couvre complètement le courant de la cognition incarnée et pour cause, celui-ci embrasse à la fois la phénoménologie, le constructivisme, la biologie, les neurosciences, l'anthropologie, la psychologie écologique... il faut reconnaître un certain courage (ou une certaine inconscience !) des chercheurs informaticiens ou roboticiens qui tentent d'avancer dans cet ensemble complexe de considérations qui souvent les

éloignent de leurs communautés, voir les met en difficulté face à des pairs qui peuvent se targuer de résultats plus impressionnants que permettent les approches plus traditionnelles ; encore faut-il s'accorder sur ce qu'on entend par "résultats".

Une des tâches de l'informaticien est de proposer des modèles et des algorithmes qui fonctionnent, là où parfois les discours théoriques peuvent sembler confus ou abstraits, au prix d'un réductionnisme souvent décrié. C'est tout l'intérêt de l'informatique pour les sciences cognitives. Les travaux mentionnés ci-dessous ne partent pas de l'hypothèse d'une représentation pré-donnée de l'environnement et du raisonnement des entités artificielles. Les représentations, si on peut les appeler ainsi et sans présupposer qu'elles sont données à l'avance au programme, sont façonnées par le biais d'interactions (renforcement, apprentissage) voir n'existent pas explicitement (automate autopoïétique) ou sont sous-jacentes à l'apparition d'un comportement (robotique souple, neurorobotique). On peut tenter une graduation du niveau de complexité de ces comportements ou de ces cognitions. Par exemple, la robotique souple (Pfeifer et al. 2012) reste pour le moment focalisée sur l'acquisition de comportements élémentaires et ce parce qu'elle n'intègre pas un système actif, doté de circularité et capable de *prendre du temps* pour que ce qui s'est passé dans le passé puisse influencer sur l'instant présent. A l'inverse, les expériences en robotique évolutionnaire montrent une capacité de mémorisation d'un système dynamique, toujours sans représentations explicites, la mémoire étant la perpétuelle influence du passé sur le présent de tels systèmes (e.g. Wood et Di Paolo 2007, De Looer et al. 2014; Doncieux et Mouret, 2014). Malgré les limites spécifiques à chaque proposition, que nous allons identifier de manière plus approfondie dans la suite de cet article, elles construisent chacune à leur manière quelque chose qui sera dépendant de l'histoire de leurs interactions avec leur environnement. On peut faire un lien avec la notion de construction du monde perçu par une créature.

Certes, mais encore faudrait-il savoir ce que l'on entend par "perception" pour une entité artificielle. La question fondamentale qui est alors posée est celle-ci : "ces modèles sont-ils suffisants pour rendre compte de phénomènes de perceptions jugés spécifiques au vivant et/ou à l'homme tels que l'expérience vécue ?". En particulier, une machine ne ressent ou n'éprouve pas. Le problème est que la conscience phénoménale sous-jacente à ces notions est pour bonne part largement incomprise. Certes, il est assez raisonnable de considérer que les émotions, par exemple, ne sont pas du registre de l'artificiel mais encore faudrait-il identifier pourquoi, et montrer en quoi les machines ne le pourraient-elles pas un jour ? On pourrait évoquer l'absence de système physiologique et hormonale porteur d'émotions voir de conscience mais ce serait ignorer le fait que l'on ne sait pas comment ce système procède pour que "cela fasse quelque chose" à l'organisme biologique qu'il constitue et tel que nous le connaissons à la première personne. La question du remplacement possible de ce système par un autre n'est alors pas complètement absurde. Plus précisément, prenons le cas des fondements de la théorie de l'énaction : c'est parce qu'il est opérationnellement clos (organisé selon une structure qui permet de maintenir cette organisation malgré les perturbations que peut susciter son environnement), qu'un organisme est considéré comme cognitif (Maturana et Varela 1980 ; Varela et al. 1993). Un système autopoïétique est opérationnellement clos et est donc cognitif. Le sens que prend l'environnement du système autopoïétique est propre à lui-même et aux interactions qu'il mène avec son environnement pour préserver son organisation. Il n'est pas accessible à la troisième personne et est éprouvé à la première personne, par l'organisme et par lui seul. Aucune machine artificielle ou ordinateur n'est autopoïétique à l'heure actuelle (elle ne se répare pas totalement ; même si certains chercheurs explorent des formes basiques et minimales d'autoréparation et d'autoreproduction chez des robots, Zphysiykov et al. 2005). Ce constat pourrait amener à la conclusion qu'aucune machine ne pourra éprouver quelque chose. Néanmoins, F. Varela a bel et bien généralisé la notion d'autopoïèse à celle d'un système non nécessairement biologique par le biais du concept de clôture opérationnelle (Varela et al. 1993). Un système, au sens général du terme, peut être

opérationnellement clos. Le système nerveux est opérationnellement clos. Varela utilise même des automates cellulaires pour illustrer la notion de couplage et de clôture opérationnelle suffisante pour produire un monde au système¹. Ce qui apparaît ici est donc la question du support matériel de la clôture opérationnelle. La difficulté réside dans le fait que les phénomènes qui nous intéressent ici (ressentir, éprouver, comprendre) sont considérés comme émergents, en partie, de la clôture opérationnelle. Les questions de cette émergence, essentiellement autour de la notion de conscience, ont été remarquablement examinées dans (Steiner 2013) qui propose qu'elles ne puissent survenir que parce que l'organisme est immergé dans un environnement ; il parle alors d'"immurgence". L'autopoïèse biologique s'appuie sur un substrat organique, le système nerveux sur des neurones et un système physiologique. L'automate cellulaire, sur un monde abstrait ou virtuel. Ils peuvent tous être immergés dans "leur" monde. C'est donc le statut des mondes virtuels (ou artificiels) qu'il faut évaluer pour connaître la possibilité d'une intelligence artificielle. Partons du fait que lorsque l'on simule un phénomène biologique on n'obtient pas ce phénomène. Par exemple, simuler une table dans un monde virtuel ne donne pas une table. Cependant, on peut faire rebondir une balle virtuelle sur une table virtuelle. En d'autres termes, on peut reproduire quelque chose qui est de l'ordre de l'organisation ou des principes du monde réel dans un monde virtuel. Ainsi on peut simuler la transformation d'une molécule par le changement d'une valeur numérique comme on peut simuler une catalyse ou une apoptose. On simule des tailles de plus en plus grandes de parties du système nerveux. Il est donc possible de simuler l'organisation du vivant tout en sachant parfaitement que *ce n'est pas le vivant*. Imaginons que l'on simule une clôture opérationnelle (le débat n'est pas clos sur cette possibilité). Est-ce qu'au contact d'un environnement (lui-même simulé), son comportement sera considéré comme cognitif et surtout y aura-t-il un pendant virtuel à la notion d'expérience vécue, une sorte d'*expérience vécue virtuelle* ? A l'inverse, est-ce que c'est le substrat au sein duquel prend naissance l'organisation du système qui est indispensable à ce phénomène qu'on appelle conscience phénoménale ? Peut-on considérer que le système nerveux, système opérationnellement clos, produit ces phénomènes indépendamment de la matière sur laquelle il repose ? Ces questions semblent soit absurdes et dérangelantes, soit vertigineuses. Il faut savoir y répondre, ou tout du moins proposer une réponse provisoire satisfaisante, pour avancer en Sciences Cognitives.

Une autre question sous-jacente est celle du lien ou des interactions possibles entre ces modèles virtuels et le monde réel ? Ici, il ne s'agit pas de débattre sur l'existence du monde réel, nous partons du principe, souvent inspiré par les constructivistes radicaux, qu'il existe, même si on ne le perçoit pas 'en soi', car il est possible de partager la perception que chacun de nous en a (même si chacun de nous ne le ressent pas nécessairement de la même façon) grâce à un consensus social (von Foerster 2003). Reprenons le cas de la table virtuelle : elle n'existe pas dans le monde réel. On peut cependant, grâce aux technologies actuelles, la montrer à des utilisateurs, ils peuvent même, grâce à des dispositifs haptiques (e.g. Millet et al. 2009), la toucher. Il est donc possible de faire interagir un monde virtuel avec le monde réel et par conséquent, possible d'imaginer une relation entre ces deux mondes. L'important réside dans les affordances que l'agent peut avoir par rapport au monde qui l'entoure, c'est-à-dire qu'il perçoit les différents objets ou constituants de ce monde à travers les actions et interactions qu'il peut effectuer sur eux (Gibson 1977). Bien évidemment, nous ne serions pas en mesure de savoir si une entité artificielle ressent ou non quelque chose. Mais si cette entité mime suffisamment bien le comportement humain, ce serait au même titre que nous n'accédons pas à ce que ressent notre voisin. Revenons sur terre car il faut savoir mettre une limite aux hypothèses créatrices mais vertigineuses. Constatons simplement qu'elles permettent de poser des questions et ici, de montrer en quoi les approches incarnées de

la cognition artificielle permettent d'aborder des points fondamentaux qui sont bien plus vastes que les petits problèmes techniques qu'elles doivent néanmoins résoudre en chemin, même si, l'air de rien, ces 'petits' problèmes sont sources de réflexions scientifiques riches et mènent parfois à la formulation de nouvelles hypothèses. Elles peuvent être le support de débats philosophiques ardues. Les auteurs de cet article ne sont pas philosophes mais ne négligent pas pour autant ce débat. Ils sont confrontés à ce qui fait la beauté et la complexité des sciences cognitives : ils doivent tenter de faire des ponts entre des disciplines et l'approche incarnée de l'intelligence artificielle est un moyen de créer ces ponts. Ils ne prétendent pas qu'une intelligence artificielle est possible, ils se posent la question et, au regard des différents travaux qui existent dans le domaine de l'IA et de leur compréhension des autres champs des sciences cognitives, ils estiment que la famille des approches incarnées est prometteuse d'un éclaircissement de ces questions.

Pour finir, rappelons que la dimension ingénierie (créative) de ces travaux porte en elle-même un potentiel de contribution aux Sciences Cognitives. L'objet n'est pas de créer une intelligence artificielle pour lui assigner un but ou un rôle particulier. L'objectif est de comprendre le mystère de la cognition par le biais de sa reproduction ou de sa modélisation, suivant en cela la phrase du physicien Richard Feynman : « What I cannot create, I do not understand. ». Si l'on considère que c'est la clôture opérationnelle qui est nécessaire à la cognition, ce n'est pas la même chose que de considérer que c'est le substrat organique sur lequel elle repose. Cette question reste entière à notre avis et ce d'autant plus si l'on aborde la cognition dans ce qu'elle a de plus intime telle que la conscience phénoménale. Néanmoins, au regard des programmes de recherche qui sont lancés de par le monde, il devient urgent d'avoir une réflexion sur l'impact qu'aurait un succès ou un échec d'un tel programme de recherche sur les fondements de pans entiers de l'histoire de l'humanité. En particulier, il faudrait être en mesure de définir des limites quant à l'usage qui serait fait d'une intelligence artificielle éprouvante. Ce débat n'est pas si éloigné de celui qui existe au sujet du transhumanisme. Nous faisons ici un pas vers la troisième famille d'approches mentionnée au début de cet article sur laquelle nous reviendrons lors de la conclusion.

3. Focus sur la cognition artificielle incarnée

Il nous a semblé intéressant d'entrer davantage dans les détails des approches de la cognition artificielle incarnée (troisième orientation présentée plus haut) pour que le lecteur puisse mieux appréhender les idées et réflexions que nous venons d'exposer. En effet, cette orientation garde en elle l'essence de la place de l'Intelligence Artificielle en Sciences Cognitives en mettant la machine à contribution pour reproduire, comprendre et tester le phénomène appelé cognition. Bien évidemment, les trois autres orientations que nous avons mentionnées en introduction la nourrissent continuellement, mais elle est à notre sens la plus rigoureuse et raisonnable au regard des avancées des Sciences Cognitives et de la compréhension des paradigmes incarnés de la cognition. Sans doute plus modeste et moins visible que les autres courants qui sont souvent médiatisés au vu des résultats, des sommes investies ou des fantasmes qu'ils génèrent (avec entre autres, le transhumanisme), elle n'en est pas moins prometteuse d'une contribution significative à long terme à notre compréhension des mécanismes de grandes fonctions cognitives telles que l'apprentissage, l'interaction sociale, la perception, la motricité, la cognition spatiale, le raisonnement, le langage ou encore la conscience. Notons que les cinq premières de ces fonctions ont été identifiées comme permettant de couvrir l'ensemble des travaux des laboratoires de recherche français du domaine de la "Robotique Cognitive" dans le rapport PIRSTEC de perspectives sur les Sciences et Technologies cognitives en 2010 (<http://pirstec.scicog.fr/ressources/accesfichier/157.pdf>).

Après une partie sur l'historique du développement des approches incarnées en Intelligence Artificielle, caractérisant donc le troisième courant énoncé ci-dessus, la suite de l'article résumera

les aspects techniques, le 'comment ça marche' des principales propositions qui se réclament de ce courant. Nous tenterons d'en identifier à chaque fois les principales caractéristiques, des exemples notables, et les contributions potentielles aux grandes questions des Sciences Cognitives sur le long-terme. Aucune d'elles ne prétend résoudre complètement la question posée par Turing. Mais chacune d'elle tente d'y apporter une contribution, en prenant sérieusement en compte les considérations théoriques de la cognition incarnée. Le lecteur pourra alors procéder à un retour sur la discussion précédente pour mieux en appréhender la portée. En effet, il nous a semblé plus pertinent d'aborder la discussion fondamentale avant les « détails » techniques pour garder un fil conducteur « serré » et pour que les deux parties de l'article puissent se lire indépendamment.

3.1 De la cognition symbolique à la cognition incarnée et interactionniste

Les fondations sur lesquelles s'appuie initialement l'Intelligence Artificielle (IA) - à savoir, un système de traitement de symboles souvent assimilé à une machine de Turing - ont fortement influencé et porté le paradigme dominant des Sciences Cognitives qui assimile le fonctionnement du cerveau à celui d'un ordinateur. Pourtant, avant la naissance même de l'IA, la cybernétique proposait, avec la notion de contre-réaction et d'auto-organisation, une autre vision des capacités cognitives et du cerveau plus bio-inspirée. Cette proposition fut plus ou moins délibérément mise de côté pour des raisons techniques et humaines (Ganascia 1996). Cependant, l'approche dominante allait être largement controversée, en particulier par des philosophes comme Searle ou Dreyfus (Dreyfus 1984). La question sous-jacente était celle de l'impossibilité pour une machine de rendre compte du sens et de la question de l'origine des représentations utilisées par les programmes d'IA et l'impossibilité de formaliser la pensée (Searle 1980). Sur cette question, Winograd devait prendre un positionnement très clair, reprenant les travaux de Dreyfus en "rebondissant" pour annoncer d'une certaine façon l'approche nouvelle de l'IA (Winograd 1989, 2006; voir aussi la traduction en français de Lassègue 1993). Parallèlement, des roboticiens tels que Brooks, affirmaient que les meilleures représentations de l'environnement étaient l'environnement lui-même et qu'il fallait se poser la question d'une intelligence sans représentations (Brooks 1991). C'est à ce moment qu'est née l'approche *animat* cherchant à simuler sur un ordinateur ou sur un robot des comportements adaptatifs simples, ancrés dans un flux d'informations sensori-motrices, et inspirés de la manière dont les animaux parviennent à agir pour se maintenir dans des zones internes et externes de viabilité (Wilson 1990; Meyer et Guillot 1994). La biologie venait en renfort et un travail conséquent produit par Francisco Varela a été de faire le lien entre les phénomènes biologiques se produisant à l'échelle de la cellule et la cognition de haut niveau grâce aux nouvelles techniques d'imagerie cérébrale utilisées en Neurosciences. Il en vint à considérer qu'il est impossible de séparer l'organisme de son environnement pour comprendre ce qu'est "son monde" car il s'agit d'un monde construit et spécifique à la dyade environnement-organisme (Varela et al. 1993). La notion d'*autopoïèse*, faculté unique du vivant à se produire lui-même et sa capacité à compenser les perturbations provenant de son environnement, est la racine de la cognition qui est alors vue comme un comportement. L'"interaction" avec l'environnement nécessite une "frontière" (membrane, corps, ...), la "vie" émergeant quand des échanges se font au travers de cette frontière/membrane qui se met en place. Dans ce cadre, le cerveau n'est qu'un moyen plus complexe, sélectionné naturellement, pour assurer la survie des espèces multicellulaires. Ainsi, le corps, l'action, le couplage entre l'organisme et l'environnement, l'ontogenèse sous-jacente sont des racines de la cognition et c'est tout le courant de la cognition incarnée qui va être lancé par ce type de considérations. Les accointances seront nombreuses avec le constructivisme en psychologie, la phénoménologie en philosophie et tous les domaines des sciences cognitives ont désormais des représentants de l'approche incarnée (même la linguistique).

3.2 Les grandes familles d'approches incarnées

Dans la suite de cet article, nous allons tenter d'expliquer les grands principes de différentes propositions, ou familles de propositions, en Intelligence Artificielle et en Robotique qui se réclament du courant incarné en Sciences Cognitives.

3.2.1 De l'autopoïèse artificielle à la morphogenèse

Les paradigmes incarnés de la cognition ont été largement influencés par les travaux de Maturana et Varela portant sur la notion d'autonomie du vivant et en particulier sur sa capacité à se "produire lui-même" et à résister aux perturbations de son environnement (Maturana et Varela 1980, Varela

1989). L'organisme n'est pas figé mais composé d'éléments en perpétuelle transformation, dont l'organisation se maintient. Ceci fait dire à des auteurs comme John Stewart que la cognition, c'est la vie (Stewart 1996). La question pour l'informaticien ou le roboticien, est alors de savoir si ce phénomène d'autopoïèse est reproductible ou simulable. L'arrière-plan épistémologique est alors de définir le statut de la simulation informatique, reposant sur un monde virtuel, relativement à son pendant biologique reposant sur un monde physique. Plusieurs modèles d'autopoïèses ont été proposés. Ils diffèrent par les mécanismes de transformation des entités virtuelles (automates cellulaires, systèmes multi-agents..), par le degré d'abstraction (représentation d'une membrane composée de cellules virtuelles, ayant une topologie ou simple organisation conceptuelle sans forme) ou par les phénomènes pris en compte; en particulier le rôle de l'environnement peut être parfois laissé de côté pour se focaliser sur les mécanismes de maintien d'une organisation (Mc Mullin 2004). La marche à franchir est alors celle de l'évolution au cours du temps de cette organisation. Quand celle-ci est topologique, on parle alors de morphogenèse qui peut être également simulée (Chavoya et Duthen 2008, Varenne et al. 2015). La complexité de la morphogenèse est telle que souvent sa simulation ne permet plus d'intégrer les principes de l'autopoïèse. Il existe alors deux orientations : celle des biologistes qui veulent partir de l'observation du vivant et de toutes ses dimensions et celle des informaticiens qui cherchent les mécanismes fondamentaux épurés de la complexité du vivant. Le fait est que les entités virtuelles simulées n'ont pas de but autre que celui de leur propre maintien. L'objectif est alors d'aller vers un comportement guidé ou imposé par l'environnement pour obtenir ce maintien. La complexité est telle que l'on fait alors des choix de simplification en ne remettant plus en cause le maintien de l'entité mais en lui fixant des buts à atteindre, c'est le cas par exemple en robotique évolutionnaire.

3.2.2 Robotique évolutionnaire

La robotique évolutionnaire reprend au compte d'entités artificielles, les principes du Darwinisme. Les modèles reposent sur des variables pouvant avoir différentes valeurs potentielles (tels les chromosomes du vivant). De nombreux individus sont générés en fixant ces variables aléatoirement. Selon qu'ils se rapprochent d'un comportement recherché ou non, ils seront sélectionnés ou éliminés. Les individus sélectionnés peuvent alors servir de base à une nouvelle génération d'individus qui héritera d'une partie des valeurs des variables de leurs aînés. L'équivalent de croisement et de mutation aléatoire étant simulé et donnant, génération après génération, naissance à des individus de plus en plus performants pour atteindre un comportement. Les variables peuvent représenter la morphologie d'un agent et donc ses capacités de mouvement (Bongard et Pfeifer 2001), elles peuvent également représenter certains paramètres de modèles cognitifs tels que des réseaux de neurones qui contrôlent des actionneurs en fonction de la valeur de capteurs. Ces modèles peuvent même aller jusqu'à reproduire des capacités cognitives plus abstraites que l'on constate chez l'humain (Wood et Di Paolo 2007) et peuvent également rendre compte de phénomènes de mémorisation alors qu'il n'existe pas de mémoire explicite dans les modèles (Ollion et al. 2009 - De Loor et al. 2014). Les individus peuvent être également mis en relation avec d'autres au sein d'un essaim afin de faire émerger des comportements collectifs. Les travaux adoptent le plus souvent une démarche itérative pour voir comment à partir de programmes simples contrôlant le comportement des robots au cours des premières générations d'individus il peut émerger des propriétés d'organisation de l'essaim non prévues par l'expérimentateur en fin d'évolution. Une partie de ces travaux cherche à contribuer à la recherche en biologie/éthologie en analysant comment le comportement d'un essaim d'animaux (ex: insecte) peut changer ou au contraire ne pas être perturbé par l'introduction d'un individu artificiel en son sein (Bredèche 2009, autres équipes, voir articles Living Machines 2014). Une autre des questions majeures dans ce domaine consiste à étudier comment l'introduction de critères de modularité des réseaux de neurones évolués, ou de diversité des comportements au

sein de l'essaim (même si cette diversité produit certains comportements individuellement inappropriés) peut contribuer à long terme à faire converger le processus évolutif simulé vers des comportements plus performants (Doncieux et Mouret 2014).

Loin de prétendre atteindre des capacités cognitives de haut-niveau, l'intérêt de ces approches réside dans l'intégration de principes phylogénétiques qui font apparaître (pour ne pas dire émerger) des représentations propres aux individus et à leur monde par le simple fait qu'elles les rendent viables au sens de l'objectif fixé par le programmeur. Ainsi, ce dernier ne définit pas une description pré-donnée du monde des entités artificielles et celles-ci sont construites au fil des générations. Bien évidemment, l'objectif est toujours fixé de l'extérieur et n'est pas intrinsèque à l'organisme comme cela serait le cas s'il était autopoïétique. Nous reviendrons sur la question que cela pose pour la notion d'intelligence artificielle.

3.2.3 Apprentissage par renforcement

Au départ, le champ de l'apprentissage par renforcement s'est développé en IA à partir d'une forte inspiration des phénomènes de conditionnement Pavlovien et instrumental étudiés en Psychologie (Sutton et Barto 1998). Il propose une modélisation formelle du processus d'apprentissage par essai-erreur qui peut se produire dans une boucle dynamique et itérative d'interaction entre un agent et son environnement de façon à maximiser une certaine fonction de récompense définie sur le long-terme. Mais les applications initiales sur des robots ayant donné des résultats décevants car trop lents à converger, trop instables et restreints à des situations simples (voir Kober et al. 2013; Khamassi 2014 pour des revues), ce champ s'est progressivement scindé en trois communautés : (1) une partie des recherches s'est de plus en plus dissociée des Sciences Cognitives pour se spécialiser sur les aspects apprentissage statistique et preuves mathématiques de convergence des algorithmes mis au point ; (2) une autre partie des recherches s'est rapprochée des Neurosciences Computationnelles à partir de la découverte que ce cadre formel permet de bien rendre compte des dynamiques de plasticité synaptique dans une partie du cerveau centrée autour des ganglions de la base, régulée par un neuromodulateur particulier, la dopamine, qui émet des signaux d'erreur de prédiction de la récompense similaires à des signaux de renforcement (Schultz et al. 1997). Cette partie des recherches s'est aussi progressivement déconnectée de la cognition incarnée pour se focaliser sur des simulations informatiques de plus en plus précises des processus neuraux étudiés ; (3) Enfin, une dernière communauté a récemment fait des progrès importants dans la mise en œuvre de capacités d'apprentissage par renforcement efficaces sur des robots. En prenant en compte l'aspect continu des perceptions, des actions et du temps, donc en prenant davantage acte de l'interaction corps-environnement et du caractère incarné de la résolution du problème d'apprentissage, elle permet à des robots l'apprentissage de primitives sensori-motrices robustes, dans des situations certes simples et requérant des démonstrations et des connaissances préalables par l'humain (Kober et al. 2013).

Des travaux récents dans le champ de l'apprentissage par renforcement ont permis en parallèle de montrer comment le coût biomécanique du mouvement peut être approximé dans la fonction de récompense (dans ses valeurs négatives) pour tenir compte de l'effort du corps dans le choix des actions. Ceci permet à ces modèles, comme cela a été observé expérimentalement chez des sujets humains, de préférer parmi deux mouvements possibles celui qui a le plus faible coût biomécanique, même si les deux mouvements aboutissent à la même valeur de récompense externe donnée par l'environnement, ou même si ces mouvements permettent d'atteindre des cibles à la même distance du sujet (Cos et al. 2013). Ces travaux donnent des pistes pour mieux intégrer la contribution du corps et de ses contraintes physiques dans les choix d'action et d'adaptation à l'environnement.

Prendre en compte d'une part les développements les plus récents des algorithmes d'apprentissage par renforcement (i.e. hiérarchie des représentations, les différentes motivations et état homéostatiques de l'agent), les modèles développés en Neurosciences Computationnelles rendant compte d'interactions entre différentes parties du cerveau, et les développements robotiques les plus aboutis permettrait de renouer avec un cadre incarné proche des questions fondamentales en Sciences Cognitives. Cette prise en compte permettrait aussi de comprendre davantage comment l'interaction entre différents processus parallèles d'apprentissage et différents systèmes de mémoire dans le cerveau permettent aux individus de faire preuve de la grande flexibilité comportementale observée dans le vivant (Khamassi 2014).

Bien évidemment, comme pour la robotique évolutionnaire, ces approches nécessitent un programmeur externe au système qui lui fixe un but. Ici, il définit les conditions d'apparition d'un renforcement des robots ou agents artificiels. On ne peut pas parler de systèmes clos et auto-organisés au sens de la clôture opérationnelle de l'autopoïèse mais on peut parler d'une construction de représentations qui dépendra de leur capacité à faire en sorte que l'entité artificielle puisse résister aux perturbations de l'environnement dans l'atteinte de sa récompense.

Le fait néanmoins que les recherches actuelles au sein de ce champ étendent actuellement la notion de simple récompense unidimensionnelle à des récompenses intrinsèques (notamment en faisant le lien avec les recherches décrites dans le paragraphe « Approche développementale »), ces récompenses pouvant avoir d'autres dimensions, telles que l'information sur l'environnement ou les feedback sociaux, et pouvant être modulées en fonction du métabolisme de l'agent (i.e. son besoin relatif en nourriture, repos, information, interaction sociale à un moment donné). Ceci laisse penser que cette famille d'approches conduira bientôt vers la mise au point d'agents capables de s'autoréguler et de définir eux-mêmes leurs récompenses à court et moyen termes comme des buts intermédiaires à atteindre pour maintenir leur métabolisme sur le long-terme, un champ qui peut en partie s'identifier au nouveau domaine du « life-long learning ».

3.2.4 Neurorobotique

Dans ce champ de recherche, il s'agit de contrôler un robot autonome par un programme informatique qui s'inspire, mime, voire modélise certains principes anatomiques, physiologiques ou de communication et de traitement de l'information du système nerveux tels qu'ils sont compris à un moment donné en Neurosciences (Seth et al. 2005). L'exercice vise à la fois à contribuer à la quête de meilleures capacités d'apprentissage et de cognition chez les robots autonomes, et est en même temps prétexte à valider/réfuter ou simplement raffiner certaines hypothèses biologiques en les testant en condition réelle (Dario et al. 2014). Pour ce dernier objectif, le rôle du corps du robot (l'aspect incarné) et le fait que ce robot soit situé dans un environnement avec lequel il interagit sont centraux - par opposition aux simulations virtuelles sur ordinateur, souvent simplifiées et parfaitement contrôlées, qui sont faites en Neurosciences Computationnelles. L'hypothèse est en effet que les aspects incarnés et situés vont générer une dynamique d'interaction différente de celle obtenue en simulation (notamment du fait du bruit sensori-moteur et des aspects non prévus de l'interaction avec l'environnement), et que cela va permettre de mieux comprendre certaines des propriétés du mécanisme biologique étudié et reproduit sur le robot. Lorsque le modèle biologique programmé sur le robot échoue à reproduire le comportement attendu, des propositions de modifications du modèle ou de nouvelles solutions pour le faire fonctionner peuvent aider à raffiner l'hypothèse biologique en jeu, ou même à proposer une nouvelle hypothèse. Au contraire, lorsque le test robotique réussit, cela peut servir d'argument pour estimer qu'un mécanisme, tel que modélisé en biologie, est suffisant à générer tel comportement en interaction avec l'environnement.

Par exemple, en connectant un réseau de neurones à un robot doté de senseurs auditifs pour reproduire des comportements d'approches entre femelles et mâles criquets s'attirant par des chants spécifiques, Webb et Scutt (2000) ont montré que deux neurones à spike interconnectés et répliquant les propriétés temporelles de neurones auditifs identifiés étaient suffisants pour reproduire un large spectre de comportements observés expérimentalement. Ces travaux confortent l'hypothèse selon laquelle la reconnaissance et la localisation d'un criquet en train de chanter pour attirer une partenaire ne nécessite pas la coordination de deux mécanismes neuronaux distincts séparés.

Un autre exemple concerne la reproduction d'activités de neurones miroirs sur des robots. Ces neurones miroirs, localisés dans le cortex prémoteur chez le singe, ont la caractéristique de répondre non seulement quand l'animal fait une action spécifique (e.g. saisir un objet) mais également lorsqu'il observe un autre individu faisant la même action, voire même lorsque cette action est faite à l'aide d'un outil tout en donnant le même résultat (e.g. l'objet est saisi) (Rizzolatti et al. 2001). Les implémentations robotiques de ce type d'activités neurales ont suggéré qu'elles permettaient d'encoder des primitives motrices utiles pour l'apprentissage par imitation (Billard et Mataric 2001).

Enfin, un dernier exemple concerne le domaine de la navigation, où des robots ont pu apprendre par essai-erreur à associer l'activité de cellules de lieux simulées (inspirées des neurones de l'hippocampe dont la réponse est spécifique d'un lieu donné dans l'environnement) à des directions de déplacement pour rejoindre efficacement une zone de récompense (Arleo et Gerstner 2000). A partir de ces premiers résultats neurorobotiques en navigation, des travaux plus récents ont permis de générer de nouvelles hypothèses biologiques sur la manière dont l'information de transition entre lieux peut être encodée efficacement dans le cerveau pour servir de support à une carte cognitive (Cuperlier et al. 2007), ou sur la manière dont le cerveau peut apprendre à coordonner différentes stratégies de navigation lorsque la carte cognitive estimée n'est pas suffisamment fiable (Caluwaerts et al. 2012).

Ces approches ne prétendent pas traiter l'ensemble du phénomène de la cognition et sont assez éloignées de considérations épistémologiques sur une construction de la connaissance ou sur le couplage individu-environnement car elles isolent des mécanismes un peu comme on recherche

une aiguille dans une botte de foin. Elles peuvent donc souffrir d'un réductionnisme qui ne s'accorde pas à la dimension circulaire de la cognition telle qu'elle est vue par le courant éactionniste. Cependant, leur intérêt réside dans le fait que contrairement à l'approche cognitiviste, elles s'appuient sur des données neurobiologiques et qu'elles ne font pas référence à une quelconque représentation prédonnée d'un monde.

3.2.5. Apprentissage développemental

L'apprentissage développemental plonge ses racines dans les travaux de Piaget (1937) qui, le premier, a théorisé l'apprentissage développemental précoce au stade sensori-moteur. Selon cette approche, l'expérience de son environnement permet à l'enfant de construire ses capacités cognitives progressivement par un mécanisme reposant « sur l'action du corps avec son milieu et les sensations qui en découlent ». Les capacités d'exploration de l'environnement sont déterminantes dans la construction de la cognition. Ces capacités sont directement liées aux capacités d'interaction du « corps » incarné avec l'environnement dans lequel il évolue. Ces capacités sont égocentrées car elles reposent exclusivement sur les capacités propres de l'enfant à construire sa cognition. Cette approche est interactionniste et constructiviste. L'interactionnisme considère qu'il n'est pas possible de séparer action et perception dans la construction cognitive comme l'a développé Merleau-Ponty. L'interactionnisme peut se décliner selon différentes facettes considérées dans l'environnement (environnement matériel, humain, social, affectif,...), mais dans le stade précoce du développement, il s'agit de l'environnement effectivement et directement interagissable par le corps incarné, avec ses capacités interactionnelles *primitives*. Le constructivisme emprunte également à Jean Piaget l'hypothèse que l'individu, confronté à une situation environnementale particulière, mobilise un schème opératoire sous la forme de structures cognitives construites pendant son développement. Le choix d'un schème particulier est guidé par une *motivation intrinsèque*, elle-même liée initialement aux capacités sensori-motrices de l'agent, mais évoluant pour s'attacher à des schèmes construits et se complexifiant avec l'expérience (e.g. Gottlieb et al. 2013).

L'apprentissage développemental constitue une branche relativement nouvelle de l'intelligence artificielle, et inspire des travaux souvent associés à la robotique interactionnelle. Citons le projet IDEAL (Implementation of DEvelopmentAI Learning) qui implémente l'approche interactionniste d'une manière radicale, en se basant sur des *primitives* qui représentent des schèmes sensorimoteurs (Georgeon et Aha 2013). Alors que les modèles d'IA classiques utilisent des primitives qui représentent *séparément* des percepts et des actions, pour éventuellement en dériver des *schèmes sensorimoteurs* (e.g., "the problem of AI is to describe and build agents that receive percepts from the environment and perform actions", Russel et Norvig 2005, p. iv), IDEAL implémente le mécanisme inverse : les primitives représentent *d'emblée* des schèmes sensorimoteurs (e.g., « sentir » un objet par le toucher—une expérience dans laquelle le mouvement relatif et la perception sont indissociables), puis l'agent construit des *actions intentionnelles* et des *perceptions signifiantes* sur la base de régularités constatées au fil de son expérience sensorimotrice. Au cours de ce processus d'apprentissage, les connaissances de l'agent restent ancrées dans son expérience sensorimotrice car elles sont encodées en termes de schèmes sensorimoteurs, et restent connectées aux préférences motivationnelles de l'agent (Georgeon et al. 2013).

De son côté, l'équipe INRIA-ENSTA FLOWERS, animée par Pierre-Yves Oudeyer considère l'apprentissage sensorimoteur et linguistique en robotique développemental et social (Oudeyer 2009). Ces travaux s'inscrivent dans l'approche de l'apprentissage développemental, mais en s'intéressant plus particulièrement aux conditions d'émergence d'un langage : quels sont les apprentissages possibles d'éléments de langage par un robot en interaction avec des humains non-ingénieurs, c'est-à-dire ne connaissant pas les mécanismes de méta-apprentissage intégrés

dans l'agent robot ? Une autre série d'expérience a mis en évidence comment des mécanismes de maximisation du progrès à prédire le flux sensori-moteur du robot (i.e. considérés comme une motivation intrinsèque de l'agent) permettent à ce dernier de passer par des phases développementales où le robot montre une "curiosité artificielle" vers des tâches de complexité croissante, comme on pourrait l'observer chez l'enfant (Oudeyer et al. 2005). Ces travaux mobilisent des disciplines très variées autour de projets partagés : psychologie développementale, robotique, mécanique, neurosciences cognitives et intégratives, linguistique, ergonomie et facteurs humains, apprentissage statistique. Cette palette constitue le socle d'une Intelligence Artificielle intégrative.

Introduire les théories constructivistes en IA est le principal intérêt de l'approche développementale. Un avantage est également de démontrer que des mécanismes simples (non complexes au sens fort du terme) permettent d'implanter la notion d'apprentissage développemental de manière efficace ? Son inconvénient est de ne pas faire, comme en neurorobotique, de liens avec les avancées en neurosciences et de considérer une certaine 'méta-connaissance' initiale avec la notion de schèmes sensorimoteurs. Un autre problème est celui de l'implémentation informatique de la notion de motivation intrinsèque qui ici ne repose pas sur l'organisation circulaire d'un système qui serait autonome mais dépendant d'un environnement, mais qui est fixé a priori par l'expérimentateur.

3.2.6. Perception active

Les programmes d'Intelligence Artificielle classiques destinés à faire interagir un agent simulé ou un robot avec son environnement considèrent l'espace géométrique comme une donnée préexistante et indépendante de l'agent. Ceci pose d'une part le problème des connaissances a priori que l'humain doit injecter dans un tel programme pour lui permettre de fonctionner (e.g. dimension, métrique, isotropie). Cela pose d'autre part des problèmes d'efficacité et de robustesse : les contraintes des systèmes robotiques sont telles que les méthodes conventionnelles de perception trouvent vite leurs limites. En particulier le problème de la fusion multimodale visant à trouver des espaces communs de représentation pour arriver à une unité perceptive est loin d'être résolu en robotique.

C'est à partir de ce constat que certains groupes de recherche travaillent au développement de méthodes de perception dite active, dans lesquelles les actions effectuées par le robot vont non seulement servir à désambiguïser la perception, mais vont jouer un rôle clé dans l'apprentissage automatique des propriétés de l'espace géométrique dans lequel le robot est situé. Ces travaux ont pour la plupart un positionnement explicite et revendiqué en Sciences Cognitives, et se réfèrent souvent à des théories sensori-motrices qui ont été développées dans différents champs disciplinaires scientifiques comme les neurosciences, la psychologie et la psychophysique, l'informatique et les mathématiques. La robotique apporte alors un terrain expérimental particulièrement intéressant pour l'évaluation de ces théories - du fait de son caractère intégratif, de son agentivité et de sa mobilité -, en plus de constituer une cible applicative naturelle.

Un des exemples que l'on peut citer concerne l'étude de la perception active multimodale chez le robot, abordée par le biais de l'apprentissage automatique des propriétés et des régularités des flux sensoriels et moteurs du robot, sans connaissances a priori, et permettant notamment d'extraire des caractéristiques de l'espace géométrique environnant comme par exemple la dimensionnalité du monde dans lequel le robot évolue (Laflaquiere et al. 2010). Ces travaux se fondent sur la théorie des contingences sensorimotrices (O'Regan et Noë 2001), et partent d'une intuition de Poincaré sur la perception de l'espace qui a été récemment formalisée mathématiquement (Philippon et al. 2003). L'espace n'est alors plus vu comme une donnée innée mais comme purement subjective et directement liée aux capacités d'action de l'agent qui va

apprendre par le biais d'une exploration active. Si celui-ci est incapable de compenser par ses actions des variations sensorielles qui ont lieu dans une dimension particulière de l'espace, alors cette dimension ne sera jamais perçue comme faisant partie de l'espace géométrique. L'action du robot n'est alors pas envisagée comme une simple mobilité donnée aux capteurs pour la recherche d'informations sur l'environnement mais bien comme une composante indispensable de la perception. Suivant cette idée, (Suzuki et Floreano 2008) ont utilisé une approche évolutionnaire pour développer une 'vision robotique éactive' : ce que perçoit l'œil guide ses mouvements qui définissent ce qu'il perçoit.

Un autre exemple de travaux récents sur la perception active se fonde sur une approche développementale, dans le but de permettre au robot d'interpréter progressivement son environnement et de le segmenter en objets individuels (Lyubova et al. 2013a). Il faut mettre ce type d'approche en regard des très nombreux algorithmes de traitement d'image qui existent aujourd'hui et qui sont tous développés spécifiquement pour une tâche et qui requièrent des bases de données adaptées et préparées a priori pour leur apprentissage. A l'inverse, les travaux décrits ici s'inspirent des capacités plus génériques de l'homme à extraire les caractéristiques et à apprendre tous les éléments souhaités de son environnement. Pour cela le programme qui contrôle le robot utilise un système unique, obtient les exemples d'apprentissage de manière autonome ou en interaction avec un humain, et se fonde sur le concept de proto-objets - issu de la psychologie cognitive (Pylyshyn 2001) - qui sert d'intermédiaire entre l'information visuelle primaire perçue et les objets reconnus qui sont utilisés par les processus cognitifs plus complexes. Les modèles de proto-objets permettent ainsi à des robots dans différentes applications utilisant différentes approches de traitement d'images d'apprendre à reconnaître progressivement les éléments de leur environnement (Orabona et al. 2007; Lyubova et al. 2013b).

3.2.7. Robotique souple

La recherche d'une cognition incarnée en Robotique atteint son paroxysme dans un ensemble de travaux regroupés sous l'étiquette de "soft robotics" (Pfeifer et al. 2012) cherchant explicitement à exploiter un maximum de propriétés du corps du robot et de l'interaction entre le corps et l'environnement pour simplifier ou pour résoudre une tâche (e.g. marcher jusqu'à un endroit) ou atteindre un objectif (e.g. saisir un objet). Certains chercheurs de ce courant utilisent à ce sujet le terme d' "intelligence encorporée", plutôt qu'incarnée, pour souligner l'idée que le corps peut contribuer à façonner le raisonnement et que l'intelligence peut en partie émerger du mouvement, de la locomotion et de la manipulation d'objets de l'environnement (Pfeifer et Pitti, 2012).

Un des plus anciens exemples de travaux relevant de cette approche fut la mise au point d'un robot exploitant la dynamique des mouvements de son corps pour produire une marche bipède la plus naturelle possible dite "marche dynamique passive" (Collins et al. 2001). En effet, la plupart des robots "classiques" effectuant une marche bipède produisent des mouvements qui ne ressemblent pas à la marche naturelle chez l'humain mais qui consistent plutôt en une série de mouvements stéréotypiques du corps rigide du robot, contrôlés par un programme informatique utilisant des paramètres résultant d'une optimisation mathématique du mouvement. A l'inverse, les travaux de Collins et collègues (2001) à Cornell University ont montré qu'un robot simple, sans capteurs, sans moteurs et sans programme informatique pour le contrôler, pouvait avoir un mouvement rythmique très naturel en exploitant la morphologie de son corps (i.e. ses deux jambes munies d'une articulation en guise de genou) et ainsi marcher sur un plan incliné si une première impulsion lui était donnée. D'autres travaux ont ensuite voulu étendre cette exploitation des propriétés du corps du robot en ajoutant un programme informatique minimal contrôlant un moteur de très faible puissance pour permettre de reproduire le même type de marche sur un sol plat.

Parmi les développements les plus récents liés à cette approche, on peut noter que de nombreuses recherches visent à intégrer des matériaux mous dans le corps du robot pour fluidifier

l'interaction de ce corps avec son environnement et exploiter certaines propriétés de cette interaction rendue fluide pour faciliter la résolution d'une tâche. On peut citer le préhenseur sous forme de ballon rempli de café moulu permettant d'initier un contact mou avec un objet, puis en faisant le vide dans le ballon, le préhenseur est rigidifié tout en conservant la forme prise en contact avec l'objet qui peut du coup être saisi et déplacé de façon stable (Brown et al. 2010).

La recherche d'une souplesse et d'une flexibilité mécanique, appelées la compliance, est de plus en plus appliquée à la robotique humanoïde. On peut citer le robot ECCE dont le squelette est constitué d'une structure flexible et de matériau déformable (Pfeifer et al. 2012), ou encore le robot ACROBAN dont les articulations sont contrôlées de manière souple et dont les actionneurs sont équipés d'élastique, permettant une stabilisation souple et une robustesse aux interactions physiques avec l'environnement (Ly et al. 2011) : contrairement aux robots dont le corps est contrôlé de manière rigide, le robot peut être tiré par la main pour être entraîné dans un mouvement ; le robot peut également se restabiliser sans perdre l'équilibre lorsqu'il est poussé par une force modérée.

La principale limite de la robotique souple (soft robotics) est de ne s'intéresser qu'au corps et de ne pas se préoccuper (ou pas encore) d'aller vers une cognition de plus haut-niveau. Cependant, son intérêt a été de montrer à la communauté que le corps permet une "économie de moyens" en termes de computation. Il faut garder à l'esprit que faire marcher un robot requière des programmes complexes de régulation et de contrôle qui font appel à des calculs complexes et souvent peu robustes, alors que les neurophysiologistes estiment que le corps associé à quelques mécanismes nerveux relativement simples (e.g. les "central pattern generators" qui peuvent générer un mouvement rythmique régulier propice à la marche), semble régler la question de l'équilibre à lui seul. Par ailleurs, cette approche illustre bien la difficulté à distinguer le corps de l'environnement et du système cognitif et indiquent aux roboticiens qu'il n'est plus possible de faire des boîtes traitant des informations entrée/sortie mais qu'il faut considérer le tout avant les parties.

4. Conclusion

Cet article avait pour objectif de dresser un panorama et des perspectives sur le paysage actuel des recherches se réclamant du courant « incarné » dans le domaine de l'Intelligence Artificielle. Nous avons pour cela distingué 4 grandes orientations de la communauté IA positionnée par rapport aux Sciences Cognitives : (1) une ayant abandonné la référence à la cognition naturelle et l'inspiration biologique pour mettre au point des algorithmes efficaces et robustes appliqués au traitement de données, à l'analyse ou au calcul scientifique ; (2) une autre s'intéressant à une approche intégrative de simulation des connaissances sur le fonctionnement neurobiologique du cerveau ; (3) une troisième concernant les chercheurs qui se sont rapprochés des paradigmes *incarnés* selon lesquels l'interaction avec l'environnement est à la base et ne peut pas être détachée de la cognition ; (4) enfin une orientation qui tente d'améliorer les capacités d'interaction de la machine au service de la cognition humaine. Nous avons également développé une discussion sur les questions fondamentales liées aux sciences cognitives et à la compréhension de phénomènes spécifiques au vivant. L'hypothèse physicaliste et objectiviste propre à la science reste confrontée à des problèmes qui ne permettent pas de trancher sur la possibilité ou non pour une machine à éprouver ou à comprendre même si l'on sait que ce ne seront pas les approches cognitivistes qui le permettront. La question de la conscience phénoménale par exemple, est souvent considérée émergente ou résultante de l'« Immurgence » d'un organisme dans son environnement. Or une machine peut être immergée. De même, que dire de la simulation du vivant dans un environnement artificiel ? Peut-elle engendrer cette émergence qui bloque l'explication physicaliste ? Toutes ces questions peuvent être abordées par l'IA et en particulier par l'IA incarnée qui s'inspire des courants incarnés et constructivistes de la cognition, les seuls qui adressent

réellement les questions difficiles mentionnées ci-dessus. Par ailleurs, la notion même « d'incarné », permettant de considérer le corps d'un agent comme lieu de la construction cognitive, ne pose pas explicitement la question de la construction de ce corps. En effet, comme le met en évidence Patricia Violi dans son article « How our Bodies Become Us: Embodiment, Semiosis and Intersubjectivity » (Violi 2009), le corps est une construction indissociable de l'agent. Cette question doit éclairer l'agenda de la recherche dans le domaine de l'IA incarnée.

Dans cet article nous avons rappelé l'historique de développement du courant incarné en IA, avant de détailler une série de familles de travaux représentatifs de cette orientation, en donnant à chaque fois des exemples de réalisations, des éléments de méthodologie et de positionnement par rapport aux autres disciplines des Sciences Cognitives (ex : Neurosciences Computationnelles, Psychologie Développementale), et en essayant de dégager des forces et des faiblesses à chaque fois. Les grandes familles de travaux que nous avons présentées sont : l'autopoïèse artificielle, la robotique évolutionnaire, l'apprentissage par renforcement, la neurorobotique, l'approche développementale et la robotique souple (soft robotics). Nous avons à la fois tenté de montrer : (i) comment ce champ de recherche s'est scindé en communautés distinctes, abordant des questions spécifiques tout en continuant à interagir avec les autres communautés. Ceci fait notamment que les objectifs de reproduire une "intelligence artificielle" ne sont pas partagés par toutes ces communautés, et que différentes finalités sont envisagées comme le fait de pouvoir extraire de l'information de grandes quantités de données, le fait de pouvoir reproduire pour mieux le comprendre un processus biologique, le fait de mettre en place des systèmes artificiels qui interagissent au mieux avec l'homme, et le fait de mettre au point des robots les plus adaptatifs et autonomes possibles lorsqu'ils sont confrontés à des situations non prévues par l'homme. Nous avons également tenté de montrer (ii) comment les évolutions les plus récentes des travaux de ce champ et la perspective de leur intégration sur des robots physiques en interaction avec le monde réel pourraient permettre sur le long terme de contribuer à mieux comprendre le rôle du corps dans la cognition, en permettant notamment de comparer les mêmes modèles computationnels lorsqu'ils sont testés en simulation ou dans le monde physique ; (iii) comment les résultats de ces travaux pourraient en parallèle nous aider à formaliser des modèles computationnelles et des hypothèses plus précis(es) et plus riches des capacités d'adaptation et des processus neuraux sous-jacents tels qu'ils sont observés expérimentalement chez les vertébrés. Nous avons enfin abordé la question plus philosophique et fondamentale de savoir s'il sera un jour possible de mettre au point ce qu'on appelle une véritable "intelligence artificielle", et si même cette dénomination a un sens. Nous pensons que les avancées des différents travaux mentionnés dans cette article pourront tout au moins progressivement nous aider à mieux caractériser et mieux comprendre les mécanismes sous-jacents aux grandes fonctions cognitives chez l'homme, et ainsi peut-être nous aider à mieux définir ce qu'on appelle "l'intelligence".

References

M.A. Arbib. *The handbook of brain theory and neural networks*. MIT press, 2003.

- A. Arleo, W. Gerstner. Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biological Cybernetics*, vol. 83(3), pp. 287-299, 2000.
- E. Bertin, G. Beslon, O. Gandrillon, S. Grauwin, P. Jensen, N. Schabanel, Les complexités : point de vue d'un institut des systèmes complexes. *Hermès*, 60, pp. 145-150, 2011.
- A. Billard, M.J. Mataric, Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture. *Robotics and Autonomous Systems*, vol. 37, pp. 145-160, 2001.
- J.C. Bongard, R. Pfeifer, Repeated structure and dissociation of genotypic and phenotypic complexity in artificial ontogeny, *Proceedings of the Genetic and Evolutionary Computation Conference*, San Francisco: Kaufmann, pp. 829–836, 2001.
- N. Bredèche, Contributions to Evolutionary Design of Embodied Agents: from autonomous artificial creatures to self-organizing, self-adaptive swarm of embodied agents, HDR Univ. Paris-Sud 11, 2009.
- R. A. Brooks, Intelligence without representation, *Artificial Intelligence*, vol. 47, pp. 139-159, 1991.
- E. Brown, N. Rodenberg, J. Amend, A. Mozeika, E. Steltz, M. Zakin, H. Lipson, H. Jaeger, Universal robotic gripper based on the jamming of granular material, *PNAS*, vol. 107, pp. 18809–18814, 2010.
- K. Caluwaerts, M. Staffa, S. N'Guyen, C. Grand, L. Dollé, A. Favre-Félix, B. Girard, M. Khamassi. A biologically inspired meta-control navigation system for the psikharpax rat robot. *Bioinspiration and Biomimetics*, vol. 7(2), 2012.
- A. Chavoya, Y. Duhten, *A Cell Pattern Generation Model Based on an Extended Artificial Regulatory Network*. Dans : *Bio Systems*, Elsevier, Vol. 94 N. 1-2, (en ligne), juin 2008.
- P.S. Churchland, and T.J. Sejnowski. *The computational brain*. The MIT press, 1992.
- S.H. Collins, M. Wisse, A. Ruina, A Three-Dimensional Passive-Dynamic Walking Robot with Two Legs and Knees. *The International Journal of Robotics Research*, vol. 20(7), pp. 607-615, 2001.
- I. Cos, M. Khamassi, B. Girard, Modeling the Learning of Biomechanics and Visual Planning for Decision-Making of Motor Actions. *Journal of Physiology - Paris*, vol 107(5), pp. 399-408, 2013.
- N. Cuperlier, M. Quoy, P. Gaussier, Neurobiologically inspired mobile robot navigation and planning. *Frontiers in Neurobotics*, vol. 1, pp. 1–15, 2007.
- A.R. DAMASIO, B.J. EVERITT et D. BISHOP. The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 351, no 1346, p. 1413-1420, 1996.
- P. Dayan and L.F. Abbott. *Theoretical neuroscience*. Vol. 806. Cambridge, MA: MIT Press, 2001.
- P. De Loor, K. Manac'h, and P. Chevaillier, "The memorization of in-line sensorimotor invariants: toward behavioral ontogeny and enactive agents," *Artif. Life Robot.*, vol. 19, no. 2, pp. 127–135, Jun. 2014.
- S. Doncieux, J.-B. Mouret, Beyond black-box optimization: a review of selective pressures for evolutionary robotics. *Evolutionary Intelligence*, vol. 71-93, pp. 1-23, Springer Berlin Heidelberg, 2014.
- H. Dreyfus, *Intelligence artificielle : mythes et limites*. Flammarion, 1984.
- G. Dulac Arnold, L. Denoyer, Ph. Preux, P. Gallinari, Sequential approaches for learning datum-wise sparse representations. *Machine Learning*, vol. 89 (1-2), pp. 87-122, 2012.
- C. ELIASMITH, T.C. STEWART, X. CHOO *et al.* A large-scale model of the functioning brain. *science*, 2012, vol. 338, no 6111, p. 1202-1205.
- J. Fix, N. Rougier, and F. Alexandre. From physiological principles to computational models of the cortex. *Journal of Physiology Paris*, 101(1–3), pp. 32–39, 2007.
- D. Floreano, A.J. Ijspeert, S. Schaal, *Robotics and Neuroscience*, *Current Biology*, vol. 24(18), pp. R910-R920, 2014.

- H. von Foerster, *Understanding Understanding, Essays on Cybernetics and Cognition*, 2003.
- K. FRISTON. A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, vol. 360, no 1456, p. 815-836, 2005.
- T. Froese, K. Suzuki, S. Wakisaka, Y. Ogai, and T. Ikegami, "From Artificial Life to Artificial Embodiment : Using human-computer interfaces to investigate the embodied mind 'as-it-could-be' from the first-person perspective", in *Proceedings of AISB'11: Computing & Philosophy*. York, UK: Society for the Study of Artificial Intelligence and the Simulation of Behavior, 2011, pp. 43–50.
- J.-G. Ganascia, *Les sciences cognitives*, Flammarion. 1996.
- J.-G. Ganascia, J. Velcin, Clustering of conceptual graphs with sparse data, *Conceptual Structures at Work*, pp. 156-169, 2004.
- M.S. Gazzaniga. *The social brain: Discovering the networks of the mind*. Basic Books (AZ), 1985.
- O. Georgeon, D. Aha, The Radical Interactionism Conceptual Commitment, *Journal of Artificial General Intelligence*, vol. 4(2), pp. 31-36, 2013.
- O. Georgeon, J. Marshall, R. Manzotti, ECA: An enactivist cognitive architecture based on sensorimotor modeling, *Biologically Inspired Cognitive Architectures*, vol. 6, pp. 46-57, 2013.
- J.J. Gibson. "The theory of affordances." *Hilldale, USA* (1977).
- J. Gottlieb, P.-Y. Oudeyer, M. Lopes, A. Baranes, Information-seeking, curiosity, and attention: computational and neural mechanisms, *Trends in Cognitive Sciences*, 2013.
- E. Hollnagel, D.D. Woods (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. London, UK: Taylor & Francis.
- M. Khamassi, Coordination of parallel learning processes in animals and robots, HDR Univ. Paris 6, 2014.
- J. Kober, J. A. Bagnell, J. Peters, Reinforcement learning in robotics: A survey, *The International Journal of Robotics Research*, 2013.
- A. Laflaquiere, S. Argentieri, B. Gas, E. Castillo-Castaneda, Space Dimension Perception from the Multimodal Sensorimotor Flow of a Naive Robotic Agent, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1520 - 1525, 2010.
- J. Lassègue, Article dans *Intellectica*, 1993.
- O. Ly, M. Lapeyre, P.-Y. Oudeyer, Bio-inspired vertebral column, compliance and semi-passive dynamics in a lightweight robot, *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1465 –1472, 2011.
- N. Lyubova, S. Ivaldi, D. Filliat, Developmental object learning through manipulation and human demonstration, *Proceedings of the ICRA Mobile Manipulation Workshop on Interactive Perception*, 2013a.
- N. Lyubova, D. Filliat, S. Ivaldi, Improving object learning through manipulation and robot self-identification, *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, 2013b.
- H. R. Maturana, F. J. Varela, *Autopoiesis and Cognition: The realization of the living*. Reidel: Boston, 1980.
- B. McMullin, Thirty Years of Computational Autopoiesis: A Review. *Artif. Life*, vol. 10, pp. 277–295, 2004.
- J.-A. Meyer, A. Guillot, *From SAB90 to SAB94: Four years of Animat research*. SAB94, MIT Press, 1994.
- G. Millet, S. Haliyo, S. Régnier, V. Hayward, The Ultimate Haptic Device: First Step. *Third Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems IEEE*, IEEE, publisher, pp. 273-278, 2009..
- K. O'Regan, A. Noë, A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, vol. 24, pp. 939-973, 2001.

- F. Orabona, G. Metta, G. Sandini, A proto-object based visual attention model, *Lecture Notes in Artificial Intelligence*, vol. 4840, pp. 198–215, Springer, 2007.
- P.-Y. Oudeyer, Sur les interactions entre la robotique et les sciences de l'esprit et du comportement, *Informatique et Sciences Cognitives : influences ou confluences ?* Presses Universitaires de France, 2009.
- P.-Y. Oudeyer, F. Kaplan, V. Hafner, A. Whyte, The playground experiment: Task-independent development of a curious robot, *Proceedings of the AAAI Spring Symposium on Developmental Robotics*, pp. 42–47, 2005.
- J. Panksepp. *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press, 1998.
- R. Pfeifer, M. Lungarella, F. Iida, The challenges ahead for bio-inspired 'soft' robotics. *Communications of the ACM*, vol. 55(11), pp. 76–87, 2012.
- R. Pfeifer, A. Pitti, La révolution de l'intelligence du corps, Manuella Editions, 2012.
- D. Philipona, K. O'Regan, J-P. Nadal, Is there something out there ? inferring space from sensorimotor dependencies. *Neural computation*, vol. 15(9), pp. 2029-2049, 2003.
- Z. Pylyshyn, Visual indexes, preconceptual object, and situated vision, *Cognition*, vol. 80, pp. 127–158, 2001.
- G. Rizzolatti, L. Fogassi, V. Gallese, Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, vol. 2, pp. 661–670, 2001.
- M. Roche, *Fouille de Textes : De l'extraction des descripteurs linguistiques à leur induction*, HDR Univ. Montpellier 2, 2011.
- S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach* (3rd edition), New Jersey, USA: Prentice Hall Press, 2009.
- W. Schultz, P. Dayan, and P.R. Montague. A neural substrate of prediction and reward, *Science*, vol. 275(5306), pp. 1593-1599, 1997.
- J. Searle, "Minds, Brains, and Programs," *Behav. Brain Sci.*, pp. 417-424, 1980.
- A.K. Seth, O. Sporns, J.L. Krichmar, Neurobotic models in neuroscience and neuroinformatics, *Neuroinformatics*, vol. 3, pp. 167-170, 2005.
- P. Steiner, "Survenance, émergence et immersion : Le problème de la conscience d'un point de vue externaliste," *Rev. Philos. Louv.*, vol. 111, no. 1, pp. 69–108, 2013.
- J. Stewart, Cognition = Life : Implications for higher-level cognition. *Behavioural Processes* 35: 311-326., pp. 311–326, 1996.
- M. Suzuki and D. Floreano, "Enactive Robot Vision," *Adapt. Behav.*, vol. 16, pp. 122–128, 2008.
- A. Turing, Computing machinery and intelligence, *Minds*, vol. 59, pp. 433-460, 1950.
- F. J. Varela, *Autonomie et connaissance*. Seuil, traduction de "Principles of Biological Autonomy" par P. Bourguin et P. Dumouchel, 1989.
- F. J. Varela, E. Thompson, E. Rosch, *The Embodied Mind*. MIT Press, 1993.
- F. Varenne, P. Chaigneau, P. Petitot, & R. Doursat, Programming the emergence in morphogenetically architected complex systems. *Acta Biotheoretica*, 2015.
- P. Violi, « How our bodies become us: embodiment, semiosis and intersubjectivity », (The intersubjectivity of embodiment) in *Cognitive Semiotics*, 4(1), 1–5, pp 57-75, 2009.
- B. Webb, T. Scutt, A simple latency-dependent spiking-neuron model of cricket phonotaxis. *Biological Cybernetics*, vol. 82, pp. 247–269, 2000.
- S. W. Wilson, *The Animat path to AI*. SAB90, MIT Press, 1990.
- T. Winograd, Heidegger and the Design of Computer Systems, in *Conference on Applied Heidegger*, Berkeley, September 1989.
- T. Winograd, *Shifting Viewpoints: Artificial Intelligence and Human-Computer Interaction*. *Artificial Intelligence*, 170(18), 1256-1258, 2006.
- R. Wood and E. A. Di Paolo, "New models for old questions: Evolutionary robotics and the 'A not B' error," in *Proceedings of the 9th European Conference on Artificial life ECAL 2007*.

V. Zykov, E. Mytilinaios, B. Adams and H. Lipson. "Robotics: Self-reproducing machines" Nature, 435, 163-164, 2005. Voir le commentaire court sur l'article avec un lien vers une vidéo du robot ici : <http://www.nature.com/news/2005/050509/full/news050509-6.html>