

Intelligence Artificielle Basée sur l'Enaction: et si l'homme était dans la boucle ?

Pierre De Loor, Kristen Manac'H, Jacques Tisseau

► **To cite this version:**

Pierre De Loor, Kristen Manac'H, Jacques Tisseau. Intelligence Artificielle Basée sur l'Enaction: et si l'homme était dans la boucle?. Version Française de l'article "Enaction-Based Artificial Intelligence: Toward Co-evolution with .. 2013, <10.1007/s11023-009-9165-3>. <hal-00654120>

HAL Id: hal-00654120

<https://hal.archives-ouvertes.fr/hal-00654120>

Submitted on 20 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intelligence Artificielle Basée sur l'Enaction: et si l'homme était dans la boucle ?

Version Française de: *Enaction Based Artificial Intelligence: What about Human in the Loop*, paru dans **Minds and Machines** **19** (2009) **319-343**.

Pierre De Loor, Kristen Manac'h, Jacques Tisseau
Université Européenne de Bretagne - ENIB
LISyC Laboratoire d'Informatique des Systèmes Complexes
CERV Centre Européen de Réalité Virtuelle
25, rue Claude Chappe
F-29280 Plouzané (France)
email : {deloor,manach,tisseau}@enib.fr

19 décembre 2011

Résumé

Cet article aborde l'articulation du paradigme de l'énaction avec l'intelligence artificielle. L'énaction y est considérée comme une métaphore pour l'intelligence artificielle car certaines des notions qu'elle aborde sont jugées incompatibles avec le domaine phénoménal de l'arfciel. Suite à l'explicitation de ce positionnement nous proposons une revue des principaux travaux relatifs à cette problématique en vie artificielle et en robotique. Nous mettons l'accent sur la faible prise en compte de la co-évolution au sein de ces approches. Nous faisons alors la proposition d'intégrer explicitement l'évolution de l'environnement pour raffiner l'ontogénèse du système artificiel et se rapprocher du paradigme de l'énaction. La complexité croissante des mécanismes ontogénétiques qu'il faut mettre en œuvre peut alors être compensée par un guidage interactif provenant de l'environnement. Cette proposition ne résoud pas celui de la pertinence du sens créé par la machine. Ces réflexions nous amènent à intégrer l'homme à cet environnement pour qu'une construction du sens pertinent dans une perspective d'intelligence artificielle participative, puisse advenir. Surgissent alors les questions liées à la mise en place d'une "interaction énaactive". L'article se termine par l'exploration de pistes permettant d'associer aux approches actuelles les principes de morphogenèse, de guidage, de phénoménologie des interactions ainsi que d'usage d'interfaces énaactives minimales pour mettre en place des expérimentations qui aborderont différemment le problème de l'intelligence artificielle inspirée de l'énaction.

1 Introduction

Depuis plusieurs années, les sciences cognitives vivent une évolution importante en prenant en compte la condition naturelle et engagée des organismes pour contribuer à l'explication de leurs capacités cognitives (Sharkey & Ziemke, 1998; Lakoff & Johnson, 1999). L'énaction est une proposition théorique s'inscrivant dans cette évolution (Varela, Thompson, & Rosch, 1993; Noë, 2004; Stewart, Gapenne, & E. Di Paolo, 2008). Même si le débat sur la pertinence des différents courants des sciences cognitives tend à s'estomper (Gershenson, 2004), l'énaction offre une alternative aux propositions cognitivistes (Pylyshyn, 1984) et connexionnistes (Rosenblatt, 1958) en poursuivant et dépassant les théories sensorimotrices initiées par (Gibson, 1966). Elle s'appuie sur des travaux en biologie (Maturana, Uribe, & Frenk, 1968; Maturana & Varela, 1980) et en neuroscience (Freeman & Sharkda, 1990; Freeman, 2001). Elle est en adéquation avec les thèses constructivistes (Piaget, 1970; Foerster, 1984; Shanon, 1993; Glasersfeld, 1995; Rosch, 1999) et celles de l'anthropologie culturelle (Hutchins, 2005, 2006). Enfin, son prolongement philosophique trouve échos en phénoménologie (Husserl, 1960; Merleau-Ponty, 1945; Varela et al., 1993; Lenay, 1996; Bickhard, 2003) et elle est à l'origine du programme de recherche en neurophénoménologie (Thompson & Varela, 2001; Lutz, Lachaux, Martinerie, & Varela, 2001).

L'énaction investit la construction de la cognition sur la base des interactions entre les organismes et leur environnement physique aussi bien que social (De Jaegher & Di Paolo, 2007). Elle s'inscrit donc sur les traces du constructivisme radical. La question qui nous anime ici est celle du tissage des liens entre éniacton et intelligence artificielle amorcés depuis quelques années.

Même si l'idée d'un système artificiel se construisant lui-même n'est pas neuve en IA (Turing, 1950; Von Neumann, 1966; Drescher, 1991; Hall, 2007), la Computational Theory of the Mind rencontre de nombreuses difficultés liées à la nature représentationnelle qu'elle propose (Dreyfus, 1979; Fodor, 2000)¹ : Frame problem (McCarthy, 1969; Korb, 2004), symbol grounding problem (Harnad, 1990, 1993), modélisation du sens commun (McCarthy, 1969), prise en compte du contexte (Minsky, 1982; McCarthy & Buva, 1998), créativité ou encore cognition sociale ou cognition en environnement ouvert. Pour dépasser ces difficultés, la *nouvelle AI* rejette l'idée de représentations et est à la source du courant de l'embodied-embedded AI (Brooks, 1991; Pfeifer & Gomez, 2005). Celle-ci intègre le rôle du corps et de la boucle sensorimotrice pour rendre compte des capacités cognitives d'un robot. Cependant, elle se trouve en difficulté concernant les questions d'agentivité, de téléologie et de construction du sens (Ziemke, 2001; Di Paolo, 2005; Di Paolo, Rohde, & De Jaegher, 2007) qui distinguent un système automatique, s'appuyant sur des valeurs fixées de l'extérieur, d'un sys-

1. Ces difficultés concernent également les approches connexionnistes qui, dans ce contexte, constituent un arrière plan du cognitivisme, maintenant la cognition au status de système de traitement entrée/sortie.

tème qui cré sa propre identité. L'origine biologique de ces notions, pressentie par I. KANT (Kant, 1790), J. VON UEXKÜLL (Uexküll, 1957) ou encore H. JONAS (Jonas, 1968) semble être une des clés possibles de ces problèmes. Ainsi, il faudrait procéder à une recopie minutieuse, dans l'artificiel, des mécanismes naturels (Dreyfus, 2007). La complexité d'une telle démarche semble incommensurable. Cependant (Di Paolo & Iizuka, 2008) insistent sur le fait que ce ne sont pas les détails de ces mécanismes qui importent, mais les principes sous-jacents qu'il faut identifier. Ce sont ces principes que tente d'explicitier l'énaction avec une vision radicale selon laquelle, à cause des contraintes de viabilité des organismes et de leurs capacités à agir, l'histoire de leurs interactions « cristallise » des invariants sensorimoteurs qui sont alors la source de « représentations incarnées » enactées, d'une agentivité et d'un "sense-making" (Di Paolo, 2005). Le paradigme insiste sur l'absence de représentations d'un monde pré-donné mais également sur les racines biologiques de l'autonomie : le principe d'autopoïèse. Elle prolonge celui-ci en y intégrant la boucle sensorimotrice, la co-évolution de l'organisme et de son environnement et finalement, l'énaction de son monde propre. La notion de monde propre (Uexküll, 1957), évoque le fait que la représentation du monde par un sujet lui est singulière et ne peut être détachée de son histoire et de ses capacités sensorimotrices.

Pour l'artificiel, la vision radicale de l'énaction permet de poser de nouvelles bases. Ceci a amené (Froese & Ziemke, 2008) à dresser les lignes directrices d'une « Enactive Artificial Intelligence » qui éclaircit le flou existant autour de la notion de cognition incarnée, souligné par (Clark, 1999; Nunez, 1999; Ziemke, 2004). Pour notre part, nous restons prudent sur les termes, en considérant l'énaction comme une *métaphore* pour l'intelligence artificielle. Nous parlerons plutôt d'« Enaction-Based Artificial Intelligence » (EBAI). En effet, le transfert direct d'un paradigme des sciences cognitives au domaine de l'artificiel peut emprunter des raccourcis, sources de malentendus et de confusions sur les notions initiales du paradigme. Par exemple, l'énaction emprunte à la phénoménologie la spécificité de l'expérience vécue « à la première personne » et la nécessité de passer par elle pour comprendre le 'mind'. Or, les notions d'expériences vecues, de conscience et de monde propre sont sans doute inaccessibles si ce n'est absurdes pour le cas d'une machine. Cet article ne cherche pas non plus à entrer dans le débat sur le fonctionnalisme de l'intentionnalité de l'autonomie ou de la conscience (Searle, 1997; Chalmer, 1995; Pylyshyn, 2003; Kosslyn, Thomson, & Ganis, 2006; Thompson, 2007). Nous reprendrons simplement à notre compte l'analyse de (Rohde & Stewart, 2008) qui proposent de remplacer la traditionnelle distinction entre autonomie "attributionnel" et "authentique" en avançant l'hypothèse qu'un jugement d'attribution fondé sur une connaissance du mécanisme sous-jacent générant le comportement aura une plus grande solidité qu'un jugement naïf fondé sur la seule observation du comportement. Ceci nous autorise à utiliser les idées et les avancées des sciences cognitives pour contribuer aux sciences de l'artificielle (Simon, 1969) qui en retour peuvent les consolider. En particulier, le problème de la construction du sens, crucial pour l'intelligence artificielle, peut être investie par une inspiration énaactive.

Cet article se structure de la façon suivante : Le paragraphe 2 rappelle les notions relatives à l'enaction et les caractéristiques attendues d'un système artificiel se réclamant de ce paradigme. Nous résumons ensuite dans le paragraphe 3 les principes des approches, en vie artificielle et en robotique, qui s'inscrivent dans le cadre de l'enaction. Nous montrons la faible part de la prise en compte de l'évolution de l'environnement pour chacune d'elles ainsi que la difficulté d'obtention de mécanismes ontogénétiques. De même, la notion de construction du sens d'une machine pose problème pour un utilisateur humain si celui-ci est élaboré de façon autonome dans un monde purement virtuel. L'étude de ces problèmes nous amène à faire plusieurs propositions argumentées dans le paragraphe 4 : prise ne compte plus explicite de l'évolution irréversible de l'environnement et du couplage, guidage de l'entité artificielle pour aborder des ontogénèses plus complexes comme cela est le cas dans la nature et intégration de l'*homme dans la boucle* co-évolutive pour la co-création d'un sens compatible avec la constitution social du sens et les preceptes initiaux de l'intelligence artificielle, illustrés par le test de TURING. Ce paragraphe ce poursuit en présentant des pistes que nous explorerons dans l'avenir pour contribuer à ces objectifs avant d'enchaîner sur la conclusion (paragraphe 5) .

2 De l'enaction à l'intelligence artificielle

L'enaction propose d'aborder la cognition comme *l'histoire du couplage structurel entre un organisme et son environnement*. Nous allons ici faire un résumé très rapide de l'articulation des concepts qui y sont liés. Pour un état de l'art plus conséquent, nous conseillons au lecteur les articles de synthèse (McGee, 2005, 2006). L'enaction trouve ses origines au travers de la notion de système autopoïétique qui fut proposée par Maturana et Varela comme une modélisation du vivant, axée sur la capacité des organismes à préserver leur viabilité (Varela, Maturana, & Uribe, 1974). Pour les auteurs, cette préservation définit l'autonomie de l'organisme et constitue la racine biologique de ses capacités cognitives. Un système autopoïétique est une organisation qui se produit elle-même, grâce à son environnement. L'environnement perturbe l'organisation dont le fonctionnement va évoluer en conséquence. Si ce fonctionnement permet de préserver l'organisation malgré les perturbations, celle-ci est viable. Ce nouveau fonctionnement va en retour influencer sur l'environnement et le système organisme-environnement va co-évoluer. Le fait que l'environnement *ne soit qu'une perturbation* implique qu'il n'est pas représenté a priori *dans* l'organisme. De plus, les contraintes de viabilité, la nécessité de rester en vie, dotent l'organisme, par le biais de son métabolisme et de sa capacité à agir, d'une identité. Cette identité apparaît relativement aux contraintes de viabilité qui font que l'environnement *prend progressivement sens*.

Si l'on s'abstrait de la biologie, on parle de systèmes opérationnellement clos. Les systèmes opérationnellement clos forment une organisation de proces-

sus dépendant récursivement les uns des autres pour se régénérer et peuvent être identifiés comme une unité reconnaissable dans le domaine des processus. Rien ne s'oppose à ce que la notion de système opérationnellement clos soit compatible dans le domaine phénoménal de l'artificiel.

L'articulation scientifique consiste ensuite à généraliser ce mécanisme aux organismes multi-cellulaires (Varela, 1979), donc à l'homme, à l'esprit et à la cognition sociale (De Jaegher & Di Paolo, 2007). A chaque niveau, il y a une différence liée au point de vue associé aux notions de viabilité et d'unité (Stewart, 1996; Di Paolo, 2005). Sans entrer d'avantage dans les détails et argumentations de l'approche théorique, nous retenons trois caractéristiques importantes, destinés au développement de systèmes artificiels basés sur ce paradigme :

1. L'absence de représentations a priori : dans le domaine de l'IA, cette caractéristique se rapproche des considérations de R. BROOKS (Brooks, 1991) mais qui, pour être plus précis, se transforme en une absence de représentations *d'un monde prédonné*. L'organisme ne possède pas une représentation explicite et définitive qu'il pourrait alors manipuler à la façon d'un programme impératif pour, par exemple, faire une planification ou définir une intention sous la forme d'un calcul symbolique. Ce sont ses interactions qui lui permettent simplement de *survivre* en préservant des invariants sensorimoteurs.
2. La plasticité : l'organisme est viable parce qu'il est capable *d'absorber* les perturbations provoquées par son environnement et de s'y adapter. On retrouve cette plasticité au niveau du corps pour les interactions physiques mais également au niveau nerveux pour les interactions de plus haut niveau (plasticité cérébrale).
3. La co-évolution : elle impose de distinguer une cognition fondée sur le corps (physically grounded) d'une cognition qui se construit tout en construisant son environnement, enracinée dans son monde propre (rooted in their own world) (Sharkey & Ziemke, 1998). La modification du monde par l'organisme impose en retour une modification de celui-ci. Cette co-évolution peut être considérée aussi bien à une échelle phylogénétique qu'à une échelle ontogénétique. Elle donne lieu à un couplage structural caractérisé par une irréversibilité. Celle-ci est souvent illustrée par le fait que nous traçons un chemin en foulant le sol de nos pas.

Nous verrons que le système artificiel prend alors la forme d'un système complexe i.e hétérogène, dynamique ouvert et multi-échelle (Laughlin, 2005). Les propriétés d'émergence de ces systèmes sont gage d'ouverture et de multiplicité des possibilités d'évolution. La notion de « dérive naturelle », chère à l'énaction (Varela et al., 1993) est alors transformée en « dérive artificielle ». Elle est sous-jacente aux systèmes complexes et peut mener à la créativité et à l'engagement dans un « nouveau monde possible ». La créativité est considérée ici comme la possibilité de découvrir une fonctionnalité à un élément de

l'environnement qui n'a pas été défini pour cela ou qui n'a pas été défini du tout. Ces systèmes sont en mesure d'appréhender et d'énacter des parties du monde *non représentables* à l'aide de symboles ou qui résistent à l'abstraction. Ces caractéristiques sont fondamentales pour l'énaction qui considère que les savoir-faire précèdent les savoirs et qui porte une attention à la singularité de chaque expérience.

La co-évolution implique une transformation récursive du système et de son environnement. L'environnement est un acteur au même titre que l'entité qui l'occupe. Or généralement, les approches d'IA incarnée délaissent l'évolution de l'environnement au profit de la mise au point du système autonome. Cette préoccupation est illustrée par les premiers « Enactive AI design principle » de (Froese & Ziemke, 2008) :

- **principe EAI-1a** : an artificial agent must be capable of generating its own systemic identity at some level of description.
- **principe EAI-1b** : an artificial agent must be capable of changing its own systemic identity at some level of description.

L'identité systémique reprend la notion d'auto-maintenance d'une organisation telle qu'elle est considérée par la théorie de l'autopoïèse. Le principe 1b est une concession faite à la complexité de la mise en œuvre du principe 1a. Les seconds principes introduisent l'interaction entre l'organisme et l'environnement par le biais de sa boucle sensorimotrice :

- **principe EAI-2a** : an artificial agent must be capable of generating its own sensorimotor identity at some level of description.
- **principe EAI-2b** : an artificial agent must be capable of changing its own sensorimotor identity at some level of description.

Le comportement actif de l'agent est ici considéré explicitement. Il permet d'aborder la construction du sens en tant que préservation de boucles sensorimotrices mais ignore la co-évolution de l'environnement et de l'agent. Pour terminer, le rôle de l'environnement, relatif à sa capacité à mettre en danger la viabilité de l'agent, est introduit par le troisième principe :

- **principe EAI-3** : an artificial agent must have the capacity to actively regulate its structural coupling in relation to a viability constraint.

Cependant, la nature irréversible de l'évolution conjointe de l'entité et de l'environnement ne nous semble pas explicitée. Le challenge est pour l'instant d'introduire des mécanismes de régulation pour maintenir l'existence de l'entité, sachant que les contraintes exercées sur celle-ci vont évoluer. Il faut que le système régule sa régulation, qu'il accède donc à une méta-régulation (Morin, 1980). Cette mise en œuvre relève de ce que (Froese & Ziemke, 2008) nomment le « verrou (hard problem) de l'intelligence artificielle éactive ». Celui-ci consiste à réifier les règles de fonctionnement de l'organisation dans le but de définir les modifications permettant de la préserver. Cela implique une réconciliation entre le domaine de l'explicitation des connaissances (explicit design) et celui des approches évolutionnaires, seul moyen actuel pour tenter d'établir des systèmes artificiels auto-adaptatifs s'appuyant sur une approche dynamique et non représentationnelle. Avant de proposer nos pistes pour aborder ce problème, nous

allons identifier en quoi, les approches actuelles s'inscrivant dans le paradigme de l'énaction artificielle ne considèrent pas assez le rôle de l'environnement et donc la co-évolution.

3 Co-evolution et environnement dans une perspective d'« Intelligence Artificielle Enactive » ?

Très logiquement, les travaux en correspondance avec une approche enactive de l'intelligence artificielle se sont développés dans les domaines de la vie artificielle avec l'étude des principes de l'autopoïèse (McMullin, 2004; Beer, 2004; Bourguin & Stewart, 2004; Beurrier, Michel, & Ferber, 2006; Ruiz-Mirazo & Mavelli, 2008). Ces études portent sur les principes EAI-1a et EAI-1b. D'autres travaux en robotique ont suivi une orientation similaire avec le développement de la cognition dynamique artificielle que l'on peut associer à l'étude des principes EAI-2b et EAI-3 (Beer, 2000; Di Paolo, 2000; Nolfi & Floreano, 2000; Harvey, Di Paolo, Wood, Quinn, & Tuci, 2005; Wood & Di Paolo, 2007; Iizuka & Di Paolo, 2007).

Nous allons résumer ceux-ci en prêtant une attention particulière à l'intégration de l'environnement et de la co-évolution.

3.1 Simuler l'autopoïèse : Racines biologiques de l'autonomie

3.1.1 Principes

L'autopoïèse est le mécanisme biologique duquel s'inspire la théorie de l'énaction. L'autonomie d'un système autopoïétique constitue sa cognition minimale. Rappelons qu'un système autopoïétique est une unité composite, comme un réseau de production d'éléments qui 1) à travers leurs interactions, régénèrent de façon récursive le réseau de production qui les produit et 2) construisent un réseau dans lequel ils existent en constituant une frontière avec le milieu extérieur via leurs interactions préférentielles dans le réseau (Dempster, 2000). Un système autopoïétique possède les propriétés des systèmes émergents puisqu'il permet de faire apparaître un phénomène de nature indépendante de celle des éléments qui lui ont donné naissance (Laughlin, 2005). La figure 1 résume les principes des modèles de systèmes autopoïétiques minimums.

Depuis les travaux pionniers de (Von Neumann, 1966; Gardner, 1970; Langton, 1984), les chercheurs sont passés à des modélisations plus riches introduisant des mécanismes biochimiques, physiques ou des éléments de génomique (Dittrich, Ziegler, & Banzhaf, 2001; Madina, Ono, & Ikegami, 2003; Watanabe, Koizumi, Kishi, Nakamura, Kobayashi, Kazuno, Suzuki, Asada, & Tominaga, 2007; Hutton, 2007). Ceci diversifie énormément les pistes de recherches et les résultats présentés. Par conséquent, nous tentons dans cette partie, de n'aborder

FIGURE 1 – Illustration du principe d'autopoièse : Une membrane cellulaire enferme un catalyseur qui ne peut la traverser. Un substrat peut passer à travers la membrane. En présence du catalyseur, ce substrat se transforme en éléments permettant de reconstituer la membrane lorsque celle-ci se troue. Ainsi parce que le catalyseur est emprisonné il permet à la cellule de se régénérer et parce qu'elle se régénère, le catalyseur reste prisonnier de celle-ci.

que des travaux qui référencent explicitement l'étude de l'autopoièse.

Poursuivant l'analyse proposée par BARRY MCMULLIN dans (McMullin, 2004), nous identifions trois catégories d'approches :

1. L'étude de la dynamique des principes de base sur des modèles minimalistes ayant pour objectif une analyse mathématique de la viabilité du système (Bourgine & Stewart, 2004; Ruiz-Mirazo & Mavelli, 2008). Cette analyse est effectuée à l'aide d'équations différentielles stochastiques. Ces équations modélisent l'évolution des concentrations d'éléments formant le système et établissent des critères de stabilité de ces éléments. Pour ces approches, la viabilité du système est sa capacité à préserver un invariant de concentration sous contraintes dynamiques. Ces modèles ne permettent pas de manipuler la topologie du système. Par exemple, la position de la membrane de l'automate de tessellation dans l'espace est prédéfinie dans (Bourgine & Stewart, 2004). Il en découle que les notions d'intérieur et d'extérieur sont elles-mêmes implicites. Or, cette distribution topologique joue un rôle important sur le principe d'autopoièse et sur des phénomènes évolutifs tels que la déformation qui n'est donc pas modélisable.

2. L'étude de la plasticité de configurations qui peuvent être préservées en présence de perturbations ou qui permettent une action minimale d'une entité artificielle (Beer, 2004; Moreno, Etxeberria, & Umerez, 2008). Ces études portent sur des configurations de différents automates cellulaires. Cette fois, les éléments topologiques sont simulables grâce à ce type d'automate. La viabilité est la préservation ou l'évolution d'une forme inscrite sur la grille. Alors que (Beer, 2004) s'intéresse à des configurations du jeu de la vie, (Moreno et al., 2008) enrichissent l'automate initial de (Varela et al., 1974) pour le doter de capacités de déplacement influencé par la présence d'un flux de substrat sur la grille. Ils montrent, également l'influence des paramètres de cet automate sur la capacité de la cellule à se déplacer.
3. L'étude de l'émergence d'un comportement autopoïétique (Beurier, Simoin, & Ferber, 2002). Les auteurs s'appuient sur la notion d'émergence multiple générée à l'aide d'un système multi-agents situés discrets. La viabilité se résume au maintien du processus émergent. Différents agents situés sur une grille s'attirent ou se repoussent mutuellement selon des règles prédéfinis et à l'aide de phéromones virtuelles qu'ils diffusent sur la grille. Ils peuvent également changer de *nature* (cette nature étant représentée par une variable) en fonction de l'état de leur voisinage. Ce modèle exhibe des propriétés de systèmes autopoïétiques : une auto-organisation membranaire du système, une interaction préférentielle entre les éléments de cette auto-organisation et une robustesse aux perturbations, et enfin, une régénération du système en cas de dégradation.

3.1.2 Problème de co-évolution

La possibilité d'une co-évolution pour chacune de ces approches est liée à la variation qu'elles proposent autour de la notion de viabilité. Ceci illustre la diversité des façons dont le principe d'autopoïèse peut être interprété. Ceci pose au passage le problème du statut de la « nature située et physique » des principes autopoïétiques. Par exemple, la notion de frontière est intuitivement topologique mais elle peut devenir complètement abstraite dans un domaine phénoménal numérique. Néanmoins, la catégorie d'approche (1) ne modélise pas la causalité des mécanismes internes à l'entité. Ces modèles n'apportent donc pas la granularité nécessaire à l'introduction de l'équivalent d'une déformation membranaire et donc d'une interaction avec un environnement dont les caractéristiques évolueraient. Pour cela, il faut passer à la simulation, en intégrant des contraintes physiques de collision et de déplacement. Dans (Manac'h & De Loo, 2007), nous présentons une simulation d'un tel modèle à base d'agents situés dans un univers tridimensionnel continu (voir figure 2). Ces simulations nous ont montré à quel point il était difficile de retrouver les résultats théoriques de stabilisation pourtant démontrés dans les cas simplifiés de l'analyse mathématique. Des travaux similaires introduisant des paramètres physiques tels que la pression ou l'hydrophobie sont proposés par (Madina et al., 2003). Ils constituent des pistes pour intégrer la déformation et donc l'évolution de la cellule.

FIGURE 2 – Version tridimensionnelle souple d'un automate de tessellation (à droite, sa désagrégation par effondrement). Les cellules de la membrane (en vert) sont reliées par des ressorts (en gris) qui se désagrègent avec le temps. Cependant le substrat (en bleu), qui traverse la cellule, peut régénérer les liens en présence d'un catalyseur enfermé dans la cellule. Au bout d'un certain temps, les contraintes de collisions déforment la cellule et celle-ci finit par se désagréger (Manac'h & De Looz, 2007).

La catégorie (2) introduit explicitement l'évolution de la forme. Cependant, la nature discrète des automates cellulaires de (Beer, 2004) fait que les changements de forme sont brutaux. Le système est donc très fragile car sensible à un environnement qui évolue. De plus, c'est la préservation de la forme dans le temps qui est considérée comme gage de viabilité. Dans un tel contexte, il est impossible d'accéder à l'irréversibilité. Notons que ce problème n'existe pas dans l'approche de (Moreno et al., 2008) qui pourrait plus facilement tendre vers une co-évolution. La catégorie (3) investit explicitement l'émergence favorisée par des règles et des variables internes. Il reste à faire en sorte que ces règles puissent évoluer en fonction de la présence d'un environnement.

De façon plus générale, pour accéder à la co-évolution, il faut que ces approches abordent la possibilité d'agir et de transformer l'environnement qui en retour pourra transformer l'entité autopoïétique. Pour cela, il faut intégrer explicitement le rôle de l'environnement et de transformation. Néanmoins, la question de fond qui reste en arrière plan de cette démarche dans une perspective d'intelligence artificielle basée sur l'enaction reste celui de la pertinence de l'abord de ce niveau de détail et de la nature phénoménal du principe même d'autopoïèse. Rien ne semble exiger l'usage de considérations biologiques précises si les principes énumérés par (Froese & Ziemke, 2008) peuvent exister au sein d'un modèle artificiel. C'est sur de telles considérations que s'est développée la cognition dynamique artificielle.

3.2 Autonomie par l'action : la cognition dynamique artificielle

Liée à la robotique évolutionnaire (Pfeifer & Scheier, 1999; Nolfi & Floreano, 2000), la cognition dynamique artificielle s'intéresse explicitement à la capacité des boucles sensorimotrices dans la préservation de la viabilité d'un agent (Beer, 2000; Daucé, 2002; Harvey et al., 2005). Le lien avec l'énaction est parfois affirmé même si, prudemment, le terme d'« Enactive Artificial Intelligence » n'est pas explicitement mentionné. Par exemple, (Rohde & Di Paolo, 2006) proposent, dans un premier temps, que la robotique évolutionnaire puisse simplement servir à étudier les hypothèses des sciences cognitives. Pour cela, ils proposent de se concentrer sur des aspects bien particuliers du comportement naturel afin de diminuer la complexité du problème global. Ceci nécessite cependant de prendre des précautions dans la réduction opérée, car complexité, dynamique et plasticité doivent subsister. C'est un des défauts de la démarche. En particulier, le domaine phénoménal physicochimique abordé par les approches simulant l'autopoïèse est ici mis à l'écart pour s'orienter vers le domaine neuronal dynamique d'un agent pleinement constitué. Il y a donc un changement de perspective de la notion de viabilité. Le modèle de référence est le réseau de neurones dynamiques récurrents (CTRNN²) (Beer & Gallagher, 1992) qui s'inscrit dans la théorie des systèmes dynamiques (Strogatz, 1994). Il présente l'avantage de pouvoir approximer la majorité des familles de systèmes dynamiques (Funahashi & Nakamura, 1993). Un tel réseau possède un comportement dynamique chaotique doté d'attracteurs (Bersini & Sener, 2002). Concrètement tous les noeuds sont connectés entre-eux et la valeur de sortie de chacun d'eux est définie par une équation différentielle. Les paramètres de ces équations sont définis à l'aide d'algorithmes génétiques qui permettent de sélectionner et d'améliorer les bonnes solutions selon une métaphore Darwinienne. Pour la suite de l'article, nous parlerons d'approches phylogénétiques. Les sorties des noeuds possèdent un régime oscillatoire. Certains arcs - mais dans une proportion très faible - sont reliés à des capteurs ou des actionneurs de l'agent.

Le problème est de mettre au point des réseaux qui vont permettre une auto-adaptation de ces boucles sensorimotrices au fil de l'expérience. Les travaux de (Beer & Gallagher, 1992) montrent une telle adaptation avec l'exemple d'une réseau dont la dynamique compense une modification du corps du robot. Pour cela, l'algorithme génétique a sélectionné à l'avance des individus qui fonctionnent avec et sans la modification. De plus, le réseaux est pré-structuré et non complètement récurrent. Il n'est plus un approximateur aussi universel que le modèle non structuré des CTRNN. Pour dépasser ces limites, E. DI PAOLO propose de rendre le réseau de neurones plastique, c'est à dire de permettre une modification des caractéristiques des connexions au fil de l'expérience du robot. Différentes plasticités peuvent être utilisées. La plasticité homéostatique, la plus proche d'un rapproche avec l'énaction, est basée sur la notion d'ultra-stabilité de ASHBY (Ashby, 1960). Elle consiste en une mise en place d'une boucle de

2. Continuous Time Recurrent Neural Networks

stabilisation qui va modifier les arcs du réseau impliqués dans des sur-activités ou des sous-activités de neurones lorsque celles-ci ont lieu. Si l'on fait le parallèle avec les conditions biologiques des organismes, le maintien de ces valeurs à l'intérieur d'un interval correspond à une condition de viabilité du réseau telle que le maintien d'une température ou d'un niveau de sucre dans le sang. La plasticité Hebbienne (hebb) consiste quand à elle à modifier les poids des arcs du réseau en fonction de la corrélation ou de la non corrélation des activités des noeuds qu'ils relient. Dans les deux cas, les règles de plasticité sont définies par un algorithme génétique. (Wood & Di Paolo, 2007) comparent ces techniques en complexifiant le comportement homéostatique grâce à la définition de zones de fonctionnements homéostatiques stables destinées à des activités précises de l'agent (Iizuka & Di Paolo, 2007).

La pertinence générale de ces démarches a été montré en reproduisant différentes expériences, souvent inspirées de la psychologie. Par exemple (Di Paolo, 2000) explique l'architecture utilisée pour donner à un robot la capacité de compenser une inversion visuelle lors d'une poursuite de cible (inversion des capteurs du robot). Ce qui est remarquable, c'est qu'au moment où les capteurs sont inversés, les règles de plasticité sont activées et vont permettre de retrouver un comportement correct alors que celui-ci n'a pas été appris phylogénétiquement dans de telles conditions. Ce qui a préservé la phylogénèse, c'est la possibilité de préserver un comportement dynamique interne satisfaisant pour la viabilité du système, même si cela nécessite de modifier les boucles sensorimotrices. Un autre élément remarquable est que plus la période de fonctionnement dans un mode particulier est longue, plus la ré-adaptation le sera, ce qui corrobore la théorie de ASHBY et la psychologie. A l'aide d'autres expériences (Harvey et al., 2005) montrent que ces réseaux possèdent des capacités à mémoriser et (Wood & Di Paolo, 2007) mettent en évidence des comportements inspirés d'expérimentations psychologiques chez l'enfant.

3.2.1 Problems for co-evolution

D'autres travaux impliquant l'évolution plastique d'un réseau de neurones sont proposés en robotique évolutionnaire (Floreano & Urzelai, 2000). Néanmoins, nous avons présenté les travaux de l'équipe de E.A. DI PAOLO car ils sont particulièrement représentatifs d'une inspiration énaactive et ils insistent sur l'agentivité du système. La plasticité permet une auto-adaptation à un environnement en revisant le principe d'ultra-stabilité qui est considéré fondamental dans ce domaine (Ikegami & Suzuki, 2008).

Cependant, même si l'action du robot est modélisée, l'environnement n'est pas modifié au sens irréversible du terme mentionné section 2. Le robot se déplace, mais n'inscrit pas une modification irréversible dans l'environnement. Par exemple, si les capteurs du robot phototaxique sont inversés, la plasticité du réseau de neurones va permettre d'aboutir à un comportement correct. A priori, si l'on remet les capteurs dans le sens initial, la configuration du réseau de neurones devrait redevenir celle du départ et l'expérience pourrait être ré-itéré autant de

FIGURE 3 – Le réseau de neurones est récurrent et génère des oscillations qui sont perturbées par l’environnement. La plasticité consiste à modifier les paramètres des équations différentielles selon différents critères (ultra-stabilité, règles de hebb ...).

fois que l’on veut, sans qu’il n’y ait de changements majeurs entre chacune de ces expériences mis à part peut-être le temps de réadaptation. En d’autres termes, l’expérience de l’inversion visuelle n’aura pas fait évoluer irréversiblement le robot phototaxique. Son histoire n’aura pas permis de lui apprendre quelque chose, de le modifier. Ainsi, l’adage selon lequel « on n’oublie rien », semble difficilement supportable à l’aide de ce type de modèle. La connaissance est inscrite dans la dynamique du réseau, mais il manque l’étape suivante qui consisterait au fait qu’elle se sédimente, qu’elle se mémorise et que le système se repose dessus pour de futures expériences. La difficulté est de trouver des *variables essentielles* associées à des règles qui puissent admettre un évolution plus radicale de celui-ci. Le principe d’ultra-stabilité ne permet pas à lui seul, d’accéder à celui d’irréversibilité, du moins à un certain degré de simplicité des modèles. Il faudrait, comme le suggère (Ikegami & Suzuki, 2008), qu’il soit lui aussi sujet à évolution. En fait, les approches évolutionnaires sont confrontées à la difficulté de l’articulation phylogénétique/ontogénétique. Ce problème semble extrêmement difficile à résoudre.

4 Propositions : Toward co-evolution with Human in the loop

4.1 Positionnement

Nous allons maintenant présenter une proposition ayant pour objectif de repousser les limites identifiées précédemment pour faire accéder une EBAI à

une constitution plus fine de son agentivité par le biais d'une co-évolution plus élaborée.

Cette proposition repose sur des argumentations suivantes autour de l'irréversibilité, de l'ontogenèse et la création du sens.

4.1.1 Le problème de l'irréversibilité

L'irréversibilité de la co-évolution n'est pas mise assez en avant car l'évolution de l'environnement, consécutive aux actions de l'agent, est négligée au profit de la mise en place d'une adaptivité à des changements externes, c'est à dire non consécutifs à l'action de l'agent lui-même. Nous proposons d'imposer que l'agent modifie activement un environnement qui soit également évolutif. Un tel principe se retrouve dans certains travaux proposant que l'environnement d'une entité soit constitué d'autres entités semblables (Nolfi & Floreano, 1998; Floreano, Mitri, Magnenat, & Keller, 2007). C'est un tel mécanisme qu'il faut mettre en place pour les entités précédentes. Néanmoins, nous argumenterons dans les sections qui suivent sur le fait que ceci n'est pas suffisant pour maîtriser cette co-évolution et la faire accéder à un *sense-making* pertinent pour l'homme. Auparavant, nous proposons de compléter les principes initiés par (Froese & Ziemke, 2008) pour la constitution d'un agent selon une perspective énaïve, par un *principe d'irréversibilité* :

- **EBAI irreversibility design principle** : an artificial agent must have the capacity to actively regulate its structural coupling, in relation to a viability constraint, avec un environnement qu'il modifie et pour lequel, certaines modifications sont irréversibles.

Ce principe implique qu'il soit possible que la perception que possède l'agent de son environnement soit modifiée suite à ses actions de telle sorte qu'il soit possible qu'il ne reprenne plus jamais cet environnement de la même façon. Le fait que cela ne concerne que certaines modifications permet tout de même à l'agent de stabiliser son couplage, chose qu'il ne peut faire avec un environnement trop plastique. Une des difficultés étant de trouver un équilibre entre une résistance qui lui permette une inscription mémorielle des interactions, un *dépot en habitus* (Husserl, 1938) et une plasticité qui lui permette d'évoluer.

4.1.2 Le problème de l'ontogenèse

Même si les agents modélisés sont complexes dans le sens où l'on fait appel à la notion d'émergence pour caractériser leur comportement d'ensemble, leur ontogenèse peut être considérée relativement simple. Soit les principes d'autopoïèse et de viabilité font l'objet de toutes les attentions au détriment de l'évolution de ces principes. Soit l'ontogenèse de l'agent est définie à l'aide d'une approche évolutionnaire. Néanmoins, l'inspiration Darwinienne de l'approche évolutionnaire a tendance à s'opposer à un effort d'explicitation de l'ontogenèse puisque l'on

évalue un agent complet. Il est prêt à fonctionner et à remplir la tâche pour laquelle il a été sélectionné. Pourtant, si l'on veut progresser en termes de capacité et élargir le domaine cognitif d'agents artificiels, il faut bien prendre en compte le fait que plus les organismes sont complexes, plus la part ontogénétique de leur comportement devient importante relativement à la part phylogénétique. De même, durant leur développement, l'influence de l'environnement devient prépondérante par rapport à celle de la prédétermination génétique (Piaget, 1975; Vaario, 1994). Dans une perspective éactive, l'évolution est considérée davantage comme un processus d'auto-organisation que comme un processus d'adaptation. Par conséquent, il faut distinguer un système auto-adaptif d'un système qui apprend (Floreano & Urzelai, 2000). En robotique par exemple, il est nécessaire d'articuler différemment la recherche évolutionnaire pour qu'elle ne porte pas sur la sélection d'agents capables de faire une tâche ou de s'adapter à un environnement changeant mais sur la sélection d'agents capables *d'adapter leur adaptation* à celle de l'autre et donc d'aborder des environnements non encore rencontrés. Ceci peut être débattu car on peut par exemple argumenter sur le fait que la créativité comportementale des organismes naturels est héritée de caractéristiques d'adaptations sélectionnées tout au long de la phylogenèse. Il n'en reste pas moins vrai que l'histoire de chaque organisme conditionne également son identité et son devenir et que cela est d'autant plus vrai que ses capacités cognitives sont évoluées (Piaget, 1975). Même si les travaux évoqués précédemment montrent que le principe d'ultra-stabilité va dans ce sens, il reste un grand pas à franchir : celui de la généralisation de principes de développements ontogénétiques. Ce problème est si difficile que nous proposons d'associer aux approches évolutionnaires un apprentissage *guidé en-ligne*, durant l'ontogénèse. Here, we fall under a Vygotskian prospect according to which training constitutes a systematic enterprise which fundamentally restructures all of the behavioral functions; it can be defined as the artificial control of the natural development process (Vygotsky1986). C'est l'occasion de revenir sur un rapprochement avec le monde biologique pour lequel, globalement, on peut constater que plus les capacités cognitives d'un organisme sont élevées, plus celui-ci a besoin d'un guide durant le début de sa vie. Ceci nécessite peut-être de passer par des modèles de plasticité d'un autre ordre, comme par exemple une plasticité morphologique de la configuration du système lui-même, qui pourrait s'accroître et spécialiser certains de ses composants au fur et à mesure de l'expérience. Ceci pose alors un problème d'explicitation de tels principes et d'imagination de modèles, de techniques et de processus capables d'en rendre compte.

4.1.3 Le problème du « sense-making »

Imaginons que l'étape précédente soit franchie et que nous sachions obtenir un système artificiel capable de co-évolution. Imaginons également que nous modélisions l'environnement d'un tel système de la même façon que le système lui-même. Il y aura une co-évolution de ces deux entités. Les deux systèmes pourraient s'engager le long de *dérives naturelles non maîtrisables*. Or, l'énac-

tion considère que le monde d'un sujet n'est rien d'autre que le résultat de ses actions sur ses sens. Ainsi, la présence d'invariants sensorimoteurs évoluant au sein d'un système artificiel est l'équivalent du monde propre virtuel de la machine, d'un *sens-making virtuel*. Que serait cette création de sens d'un système artificiel co-évoluant avec un autre système artificiel ? L'anthropomorphisme est tentant là où il n'y a aucune raison pour que ce sens en ait un pour l'homme. Nous argumentons donc sur le fait qu'un sens, cohérent avec une perspective d'usage de la machine par l'homme et d'une coopération entre la machine et l'homme - ce qui est le point de vue que nous adoptons - n'émergera que par le biais d'interactions avec un observateur humain. Sinon, nous serions face à ces machines un peu comme devant des motifs créés par des algorithmes fractaux évolutionnaires, très complexes et semblant forts organisés, mais incapables de permettre une constitution sociale et partageable du sens. Ceci ne remet pas en question l'intérêt des expérimentations de robotique évolutionnaire pour la compréhension de principes cognitifs fondamentaux. Il s'agit de tenter d'aborder le problème de la création du sens. Il faut toutefois prendre de grandes précautions en gardant à l'esprit la potentielle impossibilité d'aboutir à une telle connaissance, comme (Rohde & Stewart, 2008) l'argumentent pour la notion d'autonomie. Nous souhaitons simplement explorer les pistes permettant de nous rapprocher d'un des objectifs de l'intelligence artificielle : La confrontation entre un utilisateur humain et une machine (Turing, 1950). Nous faisons l'hypothèse qu'une piste pertinente dans une perspective énaïve, consiste à explorer la confrontation sensorimotrice homme/machine. Dans ce contexte, il nous semble que l'homme doit ressentir une *présence* de la machine qui s'exprime par une résistance sensorimotrice pour constituer un sens à son sujet. Le sentiment de présence, tout comme le test de TURING, s'évalue subjectivement. Il est particulièrement étudié dans le domaine de la réalité virtuelle (Auvray, Hanneton, Lenay, & O Regan, 2005; Sanchez & Slater, 2005; Brogni, Vinayagamoorthy, Steed, & Slater, 2007). Il permet d'établir un pont entre la phénoménologie et l'« Enaction-Based Artificial Intelligence ». Nous faisons donc l'hypothèse qu'un test de présence pourrait être à l'« Enaction-Based Artificial Intelligence » ce que le test de TURING est à l'approche computationnelle de l'IA. Une EBAI compatible avec ce test de présence doit être en interaction sensorimotrice avec l'homme de sorte qu'il puisse coordonner ses actions avec celles de la machine, qu'elle puisse en retour le guider et apprendre; qu'ils puissent constituer ensemble des *signifiants d'interactions*.

(De Jaegher & Di Paolo, 2007) argumentent sur l'aspect participatif et sur la coordination comme socle dans la constitution du sens dans une perspective énaïve. L'action de l'autre a autant d'importance que l'action d'un sujet pour contribuer à l'énaction de sa connaissance. Nous argumentons donc sur le fait que la participation de l'homme à cette co-évolution, permettra à lui et à la machine de *donner du sens*. Sans homme dans la boucle, il n'y a pas de système intelligent du point de vue de l'homme. A l'inverse, avec sa participation, le couplage fait émerger un monde propre chez l'utilisateur. Ceci pose alors le problème du choix du mode d'interaction homme/machine que nous abordons

au paragraphe 4.2.

4.1.4 Résumé de nos propositions

Pour clarifier notre propos, nous résumons nos propositions ci-dessous :

- **Proposition 1** : *Pour pallier le problème de l'irréversibilité, nous proposons d'ajouter un principe qui impose que l'agent modifie activement un environnement qui soit également évolutif.*
- **Proposition 2** : *Pour pallier le problème de la complexification de l'ontogénèse, nous proposons que, comme c'est le cas avec la complexification de la cognition dans le règne animal, nous passions à un guidage interactif de l'agent durant son ontogénèse afin d'y laisser une inscription mémorielle de ses interactions.*
- **Proposition 3** : *Pour pallier le problème de la création d'un sens pertinent, au sens du test de présence, nous proposons d'intégrer l'homme dans la boucle pour qu'une co-crédation de sens pertinent pour l'homme puisse également advenir au niveau du système artificiel.*

Ces trois propositions ne sont sans doute pas à aborder de front. Il nous semble tout à fait pertinent d'aborder l'évolution de l'environnement sans considérer la présence de l'homme dans la boucle ou de mettre en place des techniques de guidages interactifs sans aborder l'environnement. Cependant, il faut, pour chacune de ces étapes, ne pas perdre de vue la nécessité ultime des deux autres afin de guider les choix théoriques ou techniques à faire lors de leur mise en place. L'objectif final est de mettre au point des mécanismes ontogénétiques guidés par l'homme, sur des systèmes complexes dynamiques. Cet objectif est illustré par la figure 4.

L'entité artificielle est un système complexe doté de mécanismes ontogénétiques permettant un guidage de son évolution associé à un *dépot en habitus* de ses interactions. Ce guidage peut être effectué par un environnement simulé mais à terme, il doit intégrer l'homme. Nous verrons qu'il devra se faire au travers d'interfaces énaactives. La complexité d'un tel guidage « en-ligne » doit nous amener à imaginer des exercices progressifs articulant approches évolutionnaire et ontogénétique.

Nous allons donc présenter les pistes qui nous semble pertinentes pour mettre en oeuvre ce programme de recherche. La section 4.2 aborde le problème de l'interface entre l'homme et la machine, la section 4.3 celui du guidage de l'ontogénèse.

FIGURE 4 – Entité artificielle reposant sur une métaphore de l'énaction

4.2 Interface requirement

L'interface entre le système et son environnement est un des points délicats de notre proposition. En effet, pour l'énaction, la notion de corps propre, inspirée de MERLEAU-PONTY, en tant que capacité à sentir et à agir, est essentielle. Elle est à revisiter pour un système artificiel. Le corps propre conditionne la création d'un monde propre. Que deviennent monde et corps propres pour une entité artificielle? Nous devons admettre, qu'en l'état actuel de notre technique, il existe une différence de continuité corps/cognition sans commune mesure entre une machine et un organisme vivant. Nous sommes contraints de rétablir, à cause de l'implémentation technique, une séparation entre une partie cognitive et son enveloppe qui, ensemble, constituent l'équivalent d'un corps propre. L'enveloppe correspond à un clavier, une souris, un écran, un haut parleur ou tout autre périphérique plus ou moins complexe comme les interfaces comportementales de réalité virtuelle. Elle transforme un signal mécanique et énergétique en signaux électriques et vis et versa. L'enveloppe conditionne les combinaisons des signaux physiques et donc des signaux électriques. Ces signaux électriques constituent des entrées/sorties (dont nous rétablissons le statut temporairement) du système cognitif. L'enveloppe ne fait que limiter les combinaisons possibles entre entrées et sorties du système cognitif. Dans notre contexte, ses entrées/sorties ne sont pas à considérer comme des représentations d'un monde prédonné mais comme moyen de couplage entre le système cognitif et l'environnement. Ce qui techniquement s'appelle une entrée ou une sortie du système ne doit pas être en rapport avec la notion d'information mais de dynamique. Ces entrées/sorties sont des éléments de boucles sensorimotrices. La capacité à agir du système artificiel est alors corrélée à sa capacité à moduler les liens entre les entrées et les sorties du système cognitif déjà contraint par son enveloppe. La complexité du monde propre artificiel sera associée à la richesse des événements possibles de successions entrées/sorties du système cognitif. Plus il y aura de multiplicité de successions possibles, plus le système pourra exhiber de la variabilité et donc de la créativité. Certes, la complexité de l'enveloppe peut compenser la simplicité du système cognitif (McGeer, 1990), mais l'inverse est également vrai. La nature du système sensorimoteur d'une machine, aussi riche soit-il, est incomparable à celle de l'homme. Ceci rend problématique l'affirmation selon laquelle la machine doit être dotée d'un corps physique proche du notre (Brooks, 1991). Quelle que soit l'interface physique qui permet une interaction entre la machine et son environnement, celle-ci se résume à une perturbation du signal sensorimoteur numérique. Ce n'est pas le cas de l'homme biologique, totalement incarné. Il doit, lui, être doté d'interfaces énactives (Luciani & Cadoz, 2007). Ces interfaces consistent à remplacer la communication symbolique (mots, icônes) entre homme et machine par une interaction par l'action à l'aide de gestes et de forces qui forment alors des *phycones*. Nous estimons que de nombreux types d'interfaces énactives entre le système et l'utilisateur sont possibles car la perception est un processus morphogénétique (Gapenne, 2008). Une fois qu'une imbrication dynamique entre perception et action virtuelle ou numérique est en œuvre au niveau de la machine, l'interface technique peut être simple et variée. L'impor-

tant est la présence d'une dynamique non interruptible, l'absence de symboles pré-donnés et la présence de processus évolutifs des deux côtés de l'interface. Un exemple simple de telles interfaces est celui utilisé lors d'expérimentations minimalistes de la prise de conscience de l'espace chez les non voyants (Auvray et al., 2005). Ce qui importe pour que le non-voyant perçoive, c'est qu'il puisse agir et trouver des invariants sensorimoteurs. Cette expérimentation est d'autant plus intéressante que (Stewart & Gapenne, 2004) ont montré que ces interactions pouvaient être simulées par une machine à partir de descriptions qualitatives de l'expérience. De telles expériences amènent à revisiter la notion de réalité virtuelle pour la rapprocher de celle de résistance (Tisseau, 2001) et de présence (Sanchez & Slater, 2005; Brogni et al., 2007; Rohde & Stewart, 2008) que nous avons mentionné tout à l'heure : quel que soit le moyen d'interaction choisie, l'essence d'une réalité virtuelle se trouve dans sa capacité à résister aux actions, à permettre à l'utilisateur de se construire un sens. Réciproquement, une *virtualité du réel* pourra être constituée par un système artificiel s'il peut négocier sa propre résistance avec celle de l'utilisateur et établir ses propres invariants sensorimoteurs. Nous aurons alors affaire à une construction de sens artificiel partageable avec l'homme.

4.3 Guider et expliquer l'ontogenèse

Nous avons argumenté la nécessité d'utiliser des modèles dont l'ontogenèse transforme les caractéristiques de façon irréversible au cours des interactions qui servent alors de guide. Ceci nécessite d'associer des techniques d'apprentissage par renforcement (Sutton & Barto, 1998) ou par imitation (Mataric, 2001) avec des principes de transformation et d'évolution. Différentes pistes peuvent être utilisées et croisées.

L'apprentissage par renforcement, qui permet à une entité de tirer partie d'expériences pour modifier son comportement est utilisé aussi bien pour des modèles à connotation symbolique (Holland & Reitman, 1978; Wilson, 1987; Butz, Goldberg, & Stolzmann, 2000; Gerard, Stolzmann, & Sigaud, 2002) qu'avec des approches neurocomputationnelles (Daucé, Quoy, Cessac, Doyon, & Samuelides, 1998; Henry, Daucé, & Soula, 2007). Nous parlons d'approches à *connotation symbolique* pour la première catégorie car elles s'appuient sur des variables discrètes et sur une sélection d'actions atomiques. Cependant, elles ne se résument pas à utiliser une représentation prédonnée d'un environnement mais bien à construire un modèle de couplage possible avec cet environnement. Elles peuvent donc être considérées dans notre contexte. Récemment, (Chandrasekharan & Stewart, 2007) ont montré qu'il était possible d'associer un réseau de neurones bouclant sur lui-même à un algorithme de type Q-learning. Le fonctionnement de ce réseau fait office de proto-représentation. L'idée est que ces proto-représentations jouent le rôle de structures épistémiques internes reflétant des invariants sensorimoteurs appris par expérience. Il reste que le processus d'apprentissage s'effectue sur des centaines de pas de simulations pour un exemple

très simple (fourragement de fourmis virtuelles). C'est un frein important pour une application 'en-ligne' de telles approches. Du coté de la neurocomputation, (Henry et al., 2007) proposent un apprentissage par renforcement d'un réseau de neurones récurrent. La plasticité Hebbienne du réseau n'est activée que lors de la présence d'un signal de récompense ou de punition. Cette approche s'inscrit bien dans notre contexte. Il existe toutefois une différence avec la cognition dynamique artificielle puisque les expériences ne portent pas sur un apprentissage sensorimoteur et les réseaux de neurones utilisés ne sont pas des CTRNN.

Pour aborder pleinement la notion de transformation, l'introduction de principes morphogénétiques présente l'intérêt d'accéder à l'irréversibilité. Nous effectuons alors un retour sur les racines biologiques de la cognition avec les travaux sur la modélisation de la croissance d'un organisme multicellulaire (Federici & Downing, 2006; Stockholm, Benchaouir, Picot, Rameau, Neildeiz, & Paldi, 2007; Neildeiz, Parisot, Vignal, Rameau, Stockholm, Picot, Allo, Le Bec, Laplace, & Paldi, 2008). Certains chercheurs introduisent même le rôle de l'environnement lors de cette transformation (Eggenberger, 2004; Beurier et al., 2006). En robotique, on s'intéresse à l'évolution du corps de la machine (Dellaert & Beer, 1994; Hara & Pfeifer, 2003). A terme, ces approches peuvent accéder à une co-évolution au sens fort du terme. Dans le cadre d'une EBAl, l'articulation à mener consiste à intégrer les principes de l'autopoièse à ceux de la morphogénèse pour préserver l'essence biologique d'une identité construite sous contrainte de viabilité (Miller, 2003). Les travaux afférents en neurocomputing se retrouvent chez (Gruau, 1994; Nolfi & Parisi, 1995; G.vHornby & J.Pollack, 2002). Enfin, pour étudier ces principes, il est nécessaire de s'appuyer sur des outils formels adaptés à des modèles présentant des propriétés de dérives comportementales multiples. Ces outils sont rares et nous voyons dans la théorie de la viabilité de (Aubin, 1991), qui tente de définir l'ensemble des paramètres d'un modèle permettant de maintenir son comportement dans une zone définie, une perspective intéressante pour associer l'approche bottom-up de la simulation à l'analyse de propriétés globales. Cette théorie est à notre sens trop peu exploitée alors qu'elle propose un renversement de point de vue propice à l'étude des systèmes complexes.

Il reste que dans une perspective d'interaction homme/machine, c'est bien une association entre les principes de renforcement et de transformation qui doit être développée.

5 Conclusion

L'objectif de ce papier était d'analyser et de définir de nouvelles pistes pour aborder les verrous de la constitution de systèmes autonomes artificiels reposant sur une métaphore éactive. Dans un premier temps, nous avons ramené la notion d'Enactive Artificial Intelligence à celle d'Enaction-Based Artificial

Intelligence. Il s'agissait d'éviter d'aborder certains aspects phénoménologiques tels que la notion d'expérience vécue pour éviter tout amalgame avec la perspective humaine du paradigme. Nous avons ensuite montré que les principales approches actuelles étaient confrontées aux trois problèmes suivants :

1. L'absence d'une mise en oeuvre d'une réelle co-évolution caractérisée par son irréversibilité. Pour pallier ce problème, nous proposons que l'agent modifie plus activement un environnement qui soit également évolutif et qui présente un certain degré d'irréversibilité.
2. La difficulté d'établir un processus ontogénétique complexe *en devenir* qui nécessite de modifier le ratio phylogenèse/ontogenèse qui en découle pour que l'auto-organisation puisse prendre le pas sur l'auto-adaptation. A ce problème nous proposons que, comme c'est le cas avec la complexification de la cognition dans le règne animal, nous passions à un guidage interactif de l'agent durant son ontogenèse.
3. L'incommensurable distance entre la création de sens dans la machine et chez l'homme. Ceci rendant alors très hypothétique l'usage de machines capables d'un échange et d'une collaboration sociale avec l'homme. A ce problème, nous avons proposé d'intégrer l'homme dans la boucle pour qu'une création de sens pertinent pour l'homme puisse advenir au niveau du système artificiel et qu'un test de présence, équivalent éenactif du test de TURING de l'approche computationnelle, puisse avoir la possibilité d'être passé par la machine.

Il en découle une proposition intégrant une interaction homme/machine durant le processus ontogénétique d'une entité artificielle par le biais d'interfaces éenactives. Une des difficultés est alors de mettre en place des mécanismes évolutifs irréversibles s'exécutant en temps réel au sein du système. C'est pourquoi nous avons énuméré des pistes permettant d'explicitier des transformations ontogénétiques et de les façonner. Nos perspectives s'orientent vers une intégration de ces pistes par le biais d'expériences minimalistes associant robotique évolutionnaire et guidage interactif. Malgré la complexité de la tâche à accomplir, il nous semble que l'intégration de l'homme dans la boucle, en plus d'être indispensable *in-fine*, peut nous aider à établir des stratégies d'évolution et de guidage et à repousser les limites du verrou soulevé par (Froese & Ziemke, 2008). En effet, la considération de la phénoménologie des interactions homme/machine dans la constitution de représentations sensorimotrices chez l'homme peut être une base pertinente pour établir des principes analogues au niveau d'une machine.

Les tentatives de modélisation et de simulation de telles interactions par (Stewart & Gapenne, 2004) nous semble une piste pertinente. Elles peuvent nous aider à imaginer des expérimentations minimales mixant phylogenèse et ontogenèse dans l'établissement de mécanismes d'*apprentissage de l'apprentissage* au travers de principes inspirés de la morphogenèse. Ces interactions minimales passent par des interfaces simples mais éenactives, i.e. basées sur une dynamique sensorimotrice. L'important est d'établir un couplage sensorimoteur homme/machine et de garder à l'esprit l'aspect praxéologique et non ontolo-

gique du système. La complexité pourra subvenir après.

Ceci n'enlève en rien l'intérêt des approches sans l'homme dans la boucle qui permettent de progresser plus rapidement sur la compréhension des mécanismes internes grâce à la vie artificielle, la robotique évolutionnaire ou co-évolutionnaire. Nous nous fixons cependant un autre cap et avons présenté les pistes qui nous semblent les plus pertinentes. Ces réflexions font suite à nos travaux sur le couplage nécessaire entre homme et machine pour la co-construction de la connaissance (Parent, Desmeulles, Querrec, Redou, Kerdélo, Misery, Rodin, & Tisseau, 2006; Favier & De Loor, 2006; De Loor, Bénard, & Bossard, 2008). Ceci est un déficit pour le génie-logiciel qui doit s'intéresser à *l'expérience de la machine* et à ses interactions tout autant qu'à celles de l'utilisateur. C'est également un déficit pour l'intelligence artificielle théorique qui doit intégrer l'interaction au cœur de ses modèles comme le proposent (Goldin & Wegner, 2008).

6 References

- Ashby, W. (1960). *Design for a brain : The origin of adaptive behavior (2nd ed.)*. Chapman and Hall.
- Aubin, J. (1991). *Viability theory*. Birkhäuser.
- Auvray, M., Hammeton, S., Lenay, C., & O'Regan, J. (2005). There is something out there : distal attribution in sensory substitution, twenty years later. *Journal of Integrative Neuroscience*, 4, 505–521.
- Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3), 91–99.
- Beer, R. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, 10(3), 309–326.
- Beer, R., & Gallagher, J. (1992). Evolving dynamical neural networks for adaptive behavior. *Adaptive Behavior*, 1(1), 91–122.
- Bersini, H., & Sener, P. (2002). *Approche dynamique de la cognition artificielle*, Chap. Le chaos dans les réseaux de neurones, pp. 45–58. Hermes, Traité des sciences cognitives.
- Beurier, G., Michel, F., & Ferber, J. (2006). A morphogenesis model for multi-agent embryogeny. *Artificial Life X , Proceedings of the Tenth International Conference* (pp. 84–90).
- Beurier, G., Simonin, O., & Ferber, J. (2002). Model and simulation of multi-model emergence. *Proceedings of IEEE ISSPIT* (pp. 231–236).
- Bickhard, M. H. (2003). The biological emergence of representation. *Emergence and Reduction : Proceedings of the 29th Annual Symposium of the Jean Piaget Society* (pp. 105–131). T. Brown, L. Smith.
- Bourgine, P., & Stewart, J. (2004). Autopoiesis and cognition. *Artificial Life*, 10, 327–345.

- Brogni, A., Vinayagamoorthy, V., Steed, A., & Slater, M. (2007). Responses of participants during an immersive virtual environment experience. *The International Journal of Virtual Reality*, 6(2), 1–10.
- Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Butz, M. V., Goldberg, D. E., & Stolzmann, W. (2000). Investigating Generalization in the Anticipatory Classifier System. *Proceedings of Parallel Problem Solving from Nature (PPSN VI)*. Also technical report 2000014 of the Illinois Genetic Algorithms Laboratory.
- Chalmer, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, 200–219.
- Chandrasekharan, S., & Stewart, T. (2007). The origin of epistemic structures and proto-representations. *Adaptive Behavior*, 15(3), 329–353.
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Science*, 9, 345–351.
- Daucé, E. (2002). *Approche dynamique de la cognition artificielle*, Chap. Systèmes dynamiques pour les sciences cognitives, pp. 33–44. Hermes, *Traité des sciences cognitives*.
- Daucé, E., Quoy, M., Cessac, B., Doyon, B., & Samuelides, M. (1998). Self-organization and dynamics reduction in recurrent networks : stimulus presentation and learning. *Neural Networks*, 11, 521–533.
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making, an enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6(4), 485–507.
- De Loor, P., Bénard, R., & Bossard, C. (2008). Interactive co-construction to study dynamical collaborative situations. *To appear in the proceedings in the International Conference on Virtual Reality, Laval Virtual*.
- Dellaert, F., & Beer, R. (1994). Toward an evolvable model of development for autonomous agent synthesis. *Artificial Life IV, Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, 246–257.
- Dempster, B. (2000). Sympoietic and autopoietic systems : A new distinction for self-organizing systems. *Proceedings of the World Congress of the Systems Sciences and ISSS 2000* (pp. 1–18).
- Desmeulles, G., Querrec, G., Redou, P., Kerdélo, S., Misery, L., Rodin, V., & Tisseau, J. (2006). The virtual reality applied to the biology understanding : the *in virtuo* experimentation. *Expert Systems with Applications*, 30(1), 82–92.
- Di Paolo, E. (2000). Homeostatic adaptation to inversion in the visual field and other sensorimotor disruptions. In J. Meyer, A. Berthoz, D. Floreano, H. Roitblat, & S. Wilson (Eds.), *From Animals to Animats 6. Proceedings of the VI International Conference on Simulation of Adaptive Behavior* (pp. 440–449).

- Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4, 429–452.
- Di Paolo, E., & Iizuka, H. (2008). How (not) to model autonomous behavior. *BioSystems*, 91(2), 409–423.
- Di Paolo, E., Rohde, M., & De Jaegher, H. (2007). Horizons for the enactive minds : Value, social interaction, and play. *Cognitive Science Research Paper*, 587.
- Dittrich, P., Ziegler, J., & Banzhaf, W. (2001). Artificial chemistries - a review. *Artificial Life*, 7(3), 225–275.
- Drescher, G. (1991). *Made-up minds. a constructivist approach to artificial intelligence*. MIT Press.
- Dreyfus, H. (1979). *What computers can't do. The limits of artificial intelligence*. Harper&Row, Publisher, Inc.
- Dreyfus, H. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology*, 20(2), 247–268.
- EGgenberger, P. (2004). Genome-physics interaction as a new concept to reduce the number of genetic parameters in artificial evolution. In R. Sarke, R. Reynolds, H. Abbass, K.-C. Tan, R. McKay, D. Essam, & T. Gedeon (Eds.), *Proceedings of the IEEE 2003 Congress on Evolutionary Computation* (pp. 191–198). Congress of Evolutionary Computation.
- Favier, P., & De Loor, P. (2006). From decision to action : intentionality, a guide for the specification of intelligent agents' behaviour. *International Journal of Image and Graphics*, 6(1), 87–99.
- Federici, D., & Downing, K. (2006). Evolution and development of a multicellular organism : scalability, resilience, and neutral complexification. *Artif. Life*, 12(3), 381–409.
- Floreano, D., Mitri, S., Magnenat, S., & Keller, L. (2007). Evolutionary conditions for the emergence of communication in robots. *Current Biology*, 17(6), 514–519.
- Floreano, D., & Urzelai, J. (2000). Evolutionary robots with on-line self-organization and behavioral fitness. *Neural Networks*, 13, 431–443.
- Fodor, J. (2000). *The mind doesn't work that way*. MIT Press.
- Foerster, H. (1984). *The invented reality : How do we know what we believe we now ? contributions to constructivism.*, Chap. On constructing a reality, pp. 41–61. Northon and Company.
- Freeman, W. (2001). *How brains make up their minds*. Columbia University Press : New York.
- Freeman, W., & Sharkda, C. (1990). *Brain organization and memory cells, systems & circuits*, Chap. Representations : Who needs them ?, pp. 375–380. Oxford University Press : New-York.
- Froese, T., & Ziemke, T. (2008). Enactive artificial intelligence. *This is a draft paper submitted for journal publication (August 2007)*. Available on http://www.eucognition.org/white_papers.htm.

- Funahashi, K., & Nakamura, N. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6, 801–806.
- Gapenne, O. (2008). *Enaction : A new paradigm for cognitive science (to appear)*, Chap. Kinaesthetics and the construction of perceptual objects. MIT Press.
- Gardner, M. (1970). The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223(4), 120–123.
- Gerard, P., Stolzmann, W., & Sigaud, O. (2002). YACS, a new LCS using anticipation. *Journal of Soft Computing*, 6(3–4), 216–228.
- Gershenson, C. (2004). Cognitive paradigms : wich one is the best ? *Cognitive Systems Research*, 5(2), 135–156.
- Gibson, J. (1966). *The senses considered as perceptual systems*. Boston :Houghton Mifflin.
- Glaserfeld, E. v. (1995). *Radical constructivism : a way of knowing and learning*. Falmer Press : London.
- Goldin, D., & Wegner, P. (2008). The interactive nature of computing : Refuting the strong Church-Turing thesis. *Minds and Machines*, 18(1), 17–38.
- Gruau, F. (1994). Automatic definition of modular neural networks. *Adaptive Behavior*, 3, 151–183.
- G.vHornby, & J.Pollack (2002). Creating high-level components with a generative representation for body-brain evolution. *Artificial Life*, 8, 223–246.
- Hall, J. S. (2007). Self-improving ai : an analysis. *Minds and Machines*, 17(3), 249–259.
- Hara, D. F., & Pfeifer, R. (2003). *Morpho-functional machines the new species : Designing embodied intelligence*. Springer.
- Harnad, S. (1990). The symbol groundong problem. *Physica*, 335–346.
- Harnad, S. (1993). Grounding symbols in the analog world with neural nets. *NETS.THINK 2*, 2, 1.
- Harvey, I., Di Paolo, E., Wood, R., Quinn, M., & Tuci, E. (2005). Evolutionary robotics : A new scientific tool for studying cognition. *Artificial Life*, Vol. 11 (pp. 79–98).
- Henry, F., Daucé, E., & Soula, H. (2007). Temporal pattern identification using spike-timing dependent plasticity. *Neurocomputing*, 70, 2009–2016.
- Holland, J. H., & Reitman, J. S. (1978). Cognitive systems based on adaptive algorithms. In D. A. Waterman, & F. Hayes-Roth (Eds.), *Pattern-directed inference systems*. New York : Academic Press. Reprinted in : *Evolutionary Computation. The Fossil Record*. David B. Fogel (Ed.) IEEE Press, 1998. ISBN : 0-7803-3481-7.
- Husserl, E. (1938). *Expérience et jugement*. PUF (1991).
- Husserl, E. (1960). *Cartesian meditations : An introduction to phenomenology*. Trans. Dorian Cairns. The Hagues : Martinus Nijhoff.
- Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of pragmatics*, 37, 1555–1577.

- Hutchins, E. (2006). Imagining the cognitive life of things. Presented at the symposium : "The Cognitive Lifer of Things : Recasting the boundaries of Mind", organized at the McDonald institute for Archeological Research, Cambridge University.
- Hutton, T. (2007). Evolvable self-reproducing cells in a two-dimensional artificial chemistry. *Artificial Life*, 13(1), 11–30.
- Iizuka, H., & Di Paolo, E. (2007). Toward spinozist robotics : Exploring the minimal dynamics of behavioural preference. *Adaptive Behavior.*, 15, 359–376.
- Ikegami, T., & Suzuki, K. (2008). From a homeostatic to a homeodynamic self. *BioSystems*, 91, 388–400.
- Jonas, H. (1968). Biological foundations of individuality. *International Philosophical Quarterly*, 8, 231–251.
- Kant, I. (1790). *Kritik der urteilskraft*. Hackett Publishing Compagny, 1987.
- Korb, K. B. (2004). The frame problem : An AI fairy tale. *Minds and Machines*, 8(3), 317–351.
- Kosslyn, S., Thomson, W., & Ganis, G. (2006). *The case for mental imagery*. Oxford University Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh : The embodied mind and its challenge to western thought*. Basics Books : New York.
- Langton, C. (1984). Self-reproduction in cellular automata. *Physica D*, 10, 135–144.
- Laughlin, R. (2005). *A different universe : Reinventing physics from the bottom down*. Basic Book.
- Lenay, C. (1996). Mental symbols and genetic symbols : analogies between theoretical perspectives in biology and cognitive science. *Behavioural Processes*, 35, 251–262.
- Luciani, A., & Cadoz, C. (Eds.). (2007). *Enactive/07. enaction in arts. 4th international conference on enactive interfaces.*, Grenoble, France.
- Lutz, A., Lachaux, J., Martinerie, J., & Varela, J. (2001). Guiding the study of brain dynamics by using first-person data : Synchrony patterns correlate with ongoing conscious states during a simple visual task. *Proceedings of the National Academy of Sciences*, 1–6.
- Madina, D., Ono, N., & Ikegami, T. (2003). Cellular evolution in a 3d lattice artificial chemistry. *ECAL* (pp. 59–68).
- Manac'h, K., & De Loor, P. (2007). Passage du discret au continu : identification des verrous pour la simulation de systèmes autopoïétiques. *ARCo'07, Colloque de l'Association pour la Recherche Cognitive. Cognition-Complexité-Collectif* (pp. 29–30).
- Mataric, M. J. (2001). Learning in behavior-based multi-robot systems : policies, models, and other agents. *Cognitive Systems Research*, 2(1), 81–93.
- Maturana, H., Uribe, G., & Frenk, S. (1968). A biological theory of relativistic colour coding in the primat retina. *Arch. Biologia y Med. Exp.*, 1, 1–30.

- Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition : The realization of the living*. Reidel : Boston.
- McCarthy, J. (1969). *Semantic information processing*, Chap. Programs with common sense, pp. 403–418. Cambridge : MIT.
- McCarthy, J., & Buva, S. (1998). Formalizing context (expanded notes). In A. Aliseda, R. van Glabbeek, & C. Westersth (Eds.), *Computing natural language, in : Csl lecture notes*, Vol. 8 (pp. 13–50).
- McGee, K. (2005). Enactive cognitive science. part 1 : Background and research themes. *Constructivist Foundations*, 1(1), 19–34.
- McGee, K. (2006). Enactive cognitive science. part 2 : Methods, insights, and potential. *Constructivist Foundations*, 1(2), 73–82.
- McGeer, T. (1990). Passive dynamic walking. *International Journal of Robotics Research*, 9(2), 62–82.
- McMullin, B. (2004). Thirty years of computational autopoiesis : A review. *Artificial Life*, 10, 277–295.
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Collection " Tel ", 1990. Éditions Gallimard edition.
- Miller, J. (2003). Evolving developmental programs for adaptation, morphogenesis, and self-repair. *Advances in Artificial Life : 7th European Conference, LNAI*, Vol. 2801 (pp. 256–265).
- Minsky, M. (1982). Why people think computers can't. *AI Magazine*, 3(4), 224–236.
- Moreno, A., Etxeberria, A., & Umerez, J. (2008). The autonomy of biological individuals and artificial models. *BioSystems*, 91, 308–319.
- Morin, E. (1980). *La vie de la vie (t. 2)*. Le Seuil, Nouvelle édition, coll. Points.
- Neildeiz, T. M. A., Parisot, A., Vignal, C., Rameau, P., Stockholm, D., Picot, J., Allo, V., Le Bec, C., Laplace, C., & Paldi, A. (2008). Epigenetic gene expression noise and phenotypic diversification of clonal cell populations. *Differentiation*, 76, 33–40.
- Noë, A. (2004). *Action in perception*. Cambridge, MA : MIT Press.
- Nolfi, S., & Floreano, D. (1998). How co-evolution can enhance the adaptive power of artificial evolution : Implications for evolutionary robotics. In *Lecture notes in computer science*, Vol. 1468 (pp. 22–38). Springer Berlin / Heidelberg.
- Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics : The biology, intelligence, and technology of self-organizing machines*. MIT Press/Bradford Books.
- Nolfi, S., & Parisi, D. (1995). *The handbook of brain theory and neural networks*, Chap. Genotypes for neural networks. MIT Press.
- Nunez, R. (1999). Could the future taste purple? reclaiming mind, body and cognition. *Journal of Consciousness Studies*, 6, 41–60.
- Pfeifer, R., & Gomez, G. (2005). Interacting with the real world : design principles for intelligent systems. *Artificial life and Robotics*, 9(1), 1–6.

- Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. MIT Press.
- Piaget, J. (1970). *L'épistémologie génétique*. Paris : Que sais-je, PUF.
- Piaget, J. (1975). *The equilibration of cognitive structures*. The University of Chicago Press. The original French edition : L'équilibration des structures cognitives : Problème central du développement, reprint 1985.
- Pylyshyn, Z. (1984). *Computation and cognition : Toward a foundation for cognitive science*. Cambridge, Massachusetts, The MIT Press.
- Pylyshyn, Z. (2003). *Seeing and visualizing : It's not what you think*. MIT/Bradford, Cambridge.
- Rohde, M., & Di Paolo, E. (2006). An evolutionary robotics simulation of human minimal social interaction (long abstract). *SAB'06 Workshop on Behaviour and Mind as a Complex Adaptive System*. Rome, Italy.
- Rohde, M., & Stewart, J. (2008). Ascriptional and 'genuine' autonomy. *BioSystems Special issue on Modeling Autonomy*, 91(2), 424–433.
- Rosch, E. (1999). *Reclaiming cognition : The primacy of action, intention, and emotion.*, Chap. Reclaiming concepts, pp. 61–67. Imprint Academic : Thorverton, UK.
- Rosenblatt, F. (1958). The perceptron : a probabilistic model for information storage and organisation in the brain. *Psychological Review*, 65, 386–408.
- Ruiz-Mirazo, K., & Mavelli, F. (2008). On the way toward 'basic autonomous agents' : Stochastic simulations of minimal lipid-peptide cells. *Biosystems*, 91, 374–387.
- Sanchez, M., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature*, 6, 332–339.
- Searle, J. (1997). *The mystery of consciousness*. review collection.
- Shanon, B. (1993). *The representational and the presentational : An essay on cognition and the study of mind*. London and New-York Harvester-Wheatsheaf and Prentice Hall.
- Sharkey, N., & Ziemke, T. (1998). Biological and physiological foundations. *Connection Science*, 10, 361–391.
- Simon, H. (1969). *The science of the artificial*. MIT Press.
- Stewart, J. (1996). Cognition=life : Implication for higher-level cognition. *Behavioural Process*, 35, 311–326.
- Stewart, J., & Gapenne, O. (2004). Reciprocal modelling of active perception of 2-d forms in a simple tactile-vision substitution system. *Minds and Machines*, 14, 309–330.
- Stewart, J., Gapenne, O., & E. Di Paolo (2008). *Enaction : A new paradigm for cognitive science*. MIT Press.
- Stockholm, D., Benchaouir, R., Picot, J., Rameau, P., Neildeiz, T. M. A., & Paldi, A. (2007). The origin of phenotypic heterogeneity in a clonal cell population in vitro. *PLoS ONE*, 2(4).
- Strogatz, S. (1994). *Nonlinear dynamics and chaos*. Addison-Wesley.
- Sutton, R., & Barto, A. (1998). *Reinforcement learning*. Cambridge, MA : MIT Press.

- Thompson, E. (2007). Look again : Phenomenology and mental imagery. *Phenomenology and the Cognitive Science*, 6, 137–170.
- Thompson, E., & Varela, F. J. (2001). Radical embodiment : Neural dynamics and consciousness. *Trends in Cognitive Sciences*, 5(10), 418–425.
- Tisseau, J. (2001). *Virtual reality - in virtuo autonomy*. Accreditation to direct research, field : Computer science, University of Brest.
- Turing, A. (1950). Computing machinery and intelligence. *Minds*, 59, 433–460.
- Uexküll, J. v. (1957). A stroll through the worlds of animals and men. *Instinctive behavior : The development of a modern concept*, 5–80.
- Vaario, J. (1994). Artificial life as constructivist AI. *Journal of SICE (Japanese Society of Instrument and Control Engineers)*, 33(1), 65–71.
- Varela, F. (1979). *The principles of biological autonomy*. New York : North Holland.
- Varela, F., Maturana, H., & Uribe, H. (1974). Autopoiesis : the organization of living systems, its characterization, and a model. *Biosystems*, 5, 187–196.
- Varela, F., Thompson, E., & Rosch, E. (1993). *The embodied mind*. MIT Press.
- Von Neumann, J. (1966). *Theory of self-reproducing automata*. Arthur Burks.
- Watanabe, T., Koizumi, K., Kishi, K., Nakamura, M., Kobayashi, K., Kazuno, M., Suzuki, Y., Asada, Y., & Tominaga, K. (2007). A uniform framework of molecular interaction for an artificial chemistry with compartments. *Proceedings of the 2007 IEEE Symposium on Artificial Life (CI-ALife 2007)* (pp. 54–60).
- Wilson, S. W. (1987). Classifier Systems and the Animat Problem. *Machine Learning*, 2, 199–228. Also Research Memo RIS-36r, the Rowland Institute for Science, Cambridge, MA, 1986.
- Wood, R., & Di Paolo, E. A. (2007). New models for old questions : Evolutionary robotics and the 'a not b' error. In Springer-Verlag. (Ed.), *Proceedings of the 9th European Conference on Artificial life ECAL 2007*.
- Ziemke, T. (2001). The construction of 'reality' in the robot. *Foundations of Science*, 6(1), 163–233.
- Ziemke, T. (2004). Embodied AI as science : Models of embodied cognition, embodied models of cognition, or both? *LNAI*, 3139, 27–36.