

LE RÔLE DES THÉORIES INCARNÉES DE LA COGNITION DANS LE MACHINE LEARNING

Anthony PEDROSA DOS SANTOS | 21551547

Introduction

L'objectif de l'intelligence artificielle est de reproduire certains aspects de l'intelligence humaine dans des machines. Cependant, il faudrait pouvoir décrire avec précision ces aspects pour pouvoir les reproduire correctement (Lara, Astorga, Mendoza-Bock, Pardo, Escobar et Ciria, 2018). Même s'il y a eu de nombreuses recherches en intelligence artificielle, aucune n'a été capable de reproduire des capacités humaines basiques, tels que les mécanismes sensorimoteurs ou la perception. Lorsqu'un humain produit et apprend des comportements intelligents, son corps joue un rôle primordial en interagissant avec le monde qui l'entoure (Lara et al. 2018 ; Iani, 2019). Reproduire cette capacité dans une machine est donc nécessaire pour atteindre cet objectif.

Les théories récentes de la cognition incarnée

Les différentes théories de la cognition incarnée ont conceptualisé ce phénomène. Celles-ci soulignent la relation bilatérale entre l'interaction d'un humain avec son environnement et le développement de processus cognitifs plus élaborés (Iani, 2019). L'interaction des capacités motrices d'un humain (via son corps) avec le monde qui l'entoure (physiquement et socialement) constitue un élément indispensable de la cognition (Lara et al. 2018). Grâce à leur bagage génétique, les humains ont la capacité d'apprendre des schémas sensori-moteurs qui leur permettent d'interagir efficacement avec leur environnement. (Harvey, 2018). Ces schémas sensori-moteurs sont associés à différentes modalités sensorielles qui vont refléter la connaissance qu'un être humain a de son environnement et des conséquences de ses actions sur le monde qui l'entoure (Lara et al. 2018). Progressivement, suite à ses nombreuses interactions avec l'environnement, l'être humain sera capable de prédire les conséquences de

ses comportements et va intégrer les régularités de son environnement. C'est grâce à cette capacité de déduire certaines règles auxquelles l'environnement est soumis (comme la gravité) qu'un individu pourra modifier les schémas sensori-moteurs qu'il a déjà appris, tout en continuant d'en apprendre de nouveaux. Plus l'individu intégrera des schémas sensori-moteurs, plus il développera des capacités physiques et motrices complexes. Il va ainsi déployer des comportements plus élaborés qui lui permettront d'interagir avec son environnement d'une nouvelle manière (Harvey, 2018). Concrètement, un individu qui lâche un objet pour la première fois va intégrer cette régularité de notre monde appelée la gravité. Il va alors modifier ses schémas sensori-moteurs pour ne plus lâcher un objet s'il ne veut pas que celui-ci tombe. Reproduire cette capacité humaine dans une machine est primordial pour qu'elle soit capable d'intégrer les régularités d'un environnement complexe et dynamique. Une machine doit donc nécessairement disposer d'un corps physique pour être capable de reproduire certains aspects de la cognition humaine et ainsi, pouvoir produire des comportements intelligents (Lara et al., 2018).

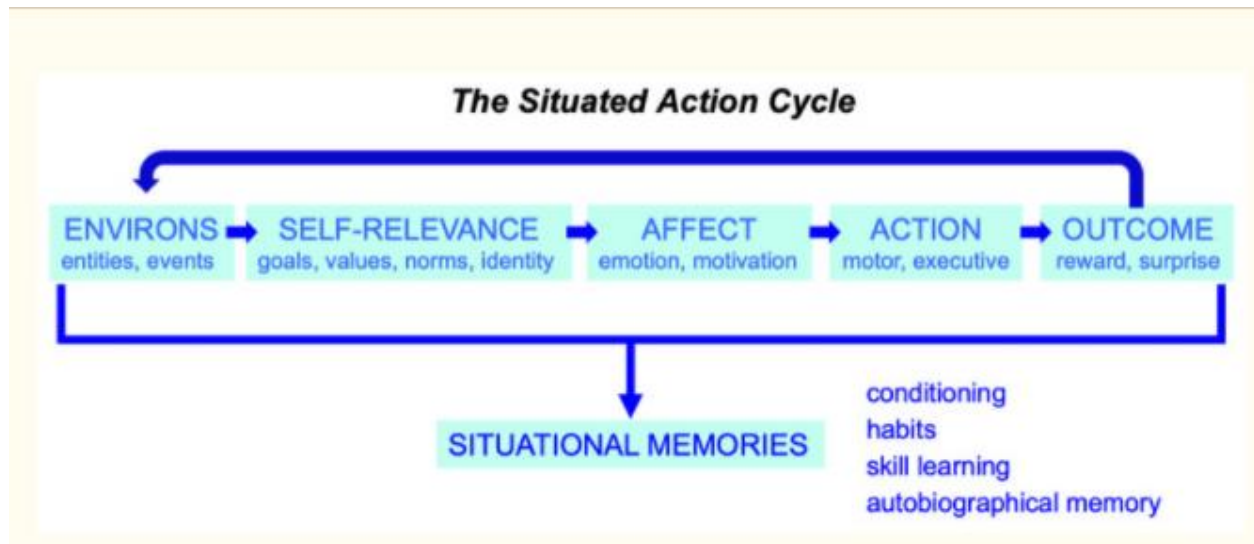
La cognition incarnée est non modulaire. Elle est souvent appelée en anglais « embodied cognition », à tort. En effet, c'est une description problématique car certains auteurs considèrent que le terme « embodied » implique uniquement le corps et que le corps doit être nécessairement engagé pendant la cognition, ce qui est faux selon Barsalou (2020). Le corps n'offrirait qu'une seule forme d'encrage particulière alors que les aspects sociaux sont aussi à prendre en compte (Barsalou, 2020). Il existe donc d'autres formes d'ancrage : l'environnement physique, l'environnement social, etc. Pour ces raisons, les chercheurs font références à la perspective non modulaire de la cognition en l'appelant « cognition incarnée » ou « cognition 4E » pour « embodied, embedded, enactive, extended ». Dans cette perspective incarnée de la cognition, la cognition, les affects et les comportements émergent du fait que le corps est intégré dans des environnements qui vont étendre la cognition. La cognition ne peut alors se comprendre que si on accepte cette proposition et donc que lorsqu'on la considère comme un cycle (Barsalou, 2020). Barsalou (2020) développe le concept de « Cycle de l'Action Située » (« Situated Action Cycle ») qui offre une description des relations entre la perception, la cognition, l'action et d'autres domaines pertinents (l'environnement, les affects, les buts, ...). La

cognition émerge des interactions entre les processus cognitifs, les modalités perceptives, le corps, l'environnement physique et l'environnement social. Par conséquent, pour étudier la cognition, il ne faut pas étudier les processus cognitifs isolément (Barsalou, 2020). Il faut donc étudier la cognition dans son ensemble. Dans ce point de vue, le Cycle de l'Action Située va permettre de comprendre la cognition d'une manière globale.

Selon ce concept développé par Barsalou (2020), il existe une première phase perceptive où l'individu perçoit une entité et/ou un événement dans son environnement (par exemple, la perception d'une voiture qui nous fonce dessus). Une fois qu'une entité ou un événement est perçu, il va y avoir une évaluation par l'agent¹ de sa pertinence. Celle-ci sera basée sur les buts, les valeurs et les normes sociales de l'agent (est-ce que cette voiture dangereuse est pertinente ?). Dans un troisième temps, cette entité ou cet événement va induire chez l'agent des affects qui vont souvent prendre la forme d'émotions et de motivations (cette voiture dangereuse peut induire de la peur et une motivation à fuir). Si cette motivation est suffisamment forte, des actions seront initiées (fuir la route). Ces actions vont produire des résultats à la fois sur le monde extérieur (en termes de récompenses et de punition) mais aussi à l'intérieur de l'agent (par exemple, lorsque l'agent réalise qu'il a commis une erreur de jugement, lorsque la voiture perçue comme dangereuse ne roulait pas très vite en réalité). Ces résultats produits vont, sous la forme d'une boucle itérative, modifier la perception qu'un agent a de son environnement. L'agent percevra alors différemment l'entité ou l'événement ayant déclenché ce cycle. Finalement, chaque cycle d'action située va superposer des informations sur son fonctionnement à travers les systèmes de mémoires, à travers des « souvenirs situationnels » (Barsalou, 2020). Dans la mesure où le Cycle de l'Action Située se déroule de façon similaire dans des situations qui se répètent et qui présentent les mêmes caractéristiques, un modèle de mise en œuvre du cycle va s'établir en mémoire, ce qui va faciliter son déploiement dans des situations futures. En d'autres termes, des schémas vont être mis en place et ceux-ci vont sous-tendre l'apprentissage et l'acquisition de nouvelles compétences.

¹ J'utilise ici le terme « agent » au lieu d' « individu » car ce cycle est censé pouvoir être modélisé par des systèmes computationnels.

Figure 1. Le Cycle de l'Action Située (Barsalou, 2020).



Intégrer les modèles incarnés de la cognition dans des systèmes computationnels

Dans l'approche incarnée de la cognition de Lara et al. (2018), pour qu'une machine puisse intégrer des capacités cognitives de haut niveau (comme la résolution de problèmes complexes) et reproduire des comportements intelligents, elle doit posséder des habiletés sensibles et motrices (Lara et al. 2018). En effet, la capacité de se mouvoir dans l'espace permet l'acquisition de nombreuses expériences perceptives qui permettent à un agent de développer certaines habiletés cognitives (Harvey, 2018 ; Lara et al., 2018).

Cependant, l'approche de Lara et al. (2018) pose problème. Posséder un corps pouvant interagir avec l'environnement physique et social ne suffit pas pour pouvoir intégrer les habiletés cognitives humaines. Cette limite est soulignée par les travaux de Barsalou (2020) qui affirme que la cognition doit être étudiée dans son ensemble. Selon cet auteur, beaucoup d'autres aspects de la cognition sont à prendre en compte dans le processus d'élaboration et de modification de patterns comportementaux. Les valeurs, les normes sociales et les affects de l'agent jouent notamment un rôle important dans ce processus. Un agent ne peut pas reproduire uniquement certains aspects spécifiques de la cognition humaine. Il reproduit nécessairement la cognition humaine dans son ensemble. Un tel agent possède des valeurs, des buts et des normes sociales singulières. Il est aussi capable de produire certaines émotions et

motivations. Ces aspects de la cognition ne sont pas pris en compte dans les travaux de Lara et al. (2018).

Selon Barsalou (2020), un robot intelligent est capable d'évaluer la pertinence d'un événement selon ses propres buts, valeurs et normes sociales. Cette évaluation devra induire chez la machine certains affects. Celle-ci doit donc pouvoir simuler des émotions et des motivations qui lui permettront d'initier ou de ne pas initier des actions. La machine devra être capable d'intégrer les conséquences de ses actions sur l'environnement externe en termes de récompense et de punition mais aussi sur l'environnement interne en termes de raisonnement sur ses actions. En résumé, la machine devra intégrer le modèle incarné de la cognition conceptualisé par l'auteur : le Cycle de l'Action Située. Cette approche permettra à la machine de modifier en permanence sa perception du monde environnant pour pouvoir déployer des comportements plus adaptés aux situations rencontrées. Cette capacité permettra à la machine d'apprendre et d'acquérir de nouvelles compétences.

D'autres auteurs soulignent aussi la limite des capacités motrices pour produire des comportements humains élaborés. Mingon et Sutton (2021) ont questionné la capacité des robots de reproduire le « Māori haka », une danse rituelle pratiquée par certains peuples indigènes de Nouvelle Zélande. Les auteurs admettent qu'une telle performance ne peut pas être envisagée comme un simple enchaînement de mouvements moteurs. Cette danse implique aussi des processus affectifs et cognitifs. Lorsqu'un groupe d'individu pratique cette danse, ils partagent des émotions et des pensées. Plusieurs chercheurs ont tenté de développer un robot capable de reproduire le Māori haka (Ser et al. 2016; Sandoval et al. 2016; Rudhru et al. 2016 ; cités par Mingon et Sutton, 2021). Le but était de conserver cette pratique et de la transmettre aux générations futures dans le cas où aucun humain ne pourrait la reproduire. Un autre objectif était de permettre à de futurs chercheurs d'étudier cette pratique traditionnelle même si cette culture venait à disparaître. Pour Mingon et Sutton (2021), la reproduction d'une telle pratique culturelle est complexe. Au-delà de simples performances motrices, le Māori haka implique des expériences affectives et cognitives transmises par les peuples indigènes de Nouvelle Zélande. De plus, les robots récents n'ont pas la capacité d'intégrer les connaissances historiques qui influencent les décisions et les actions des individus (Prescott et al., 2019 ; cités

par Mingon et Sutton, 2021). Un robot capable de reproduire une danse traditionnelle complexe n'est pas seulement capable de reproduire un enchaînement d'actions motrices. D'autres aspects de la cognition doivent être intégrés pour pouvoir reproduire un comportement aussi élaboré. Selon Mingon et Sutton (2021), le caractère multimodal d'une pratique culturelle comme le Māori haka permet de mieux comprendre les différences existantes entre humains et robots dans le cadre des performances qui requièrent des compétences particulières, mais aussi de faire émerger un champ de questions dans le domaine des sciences cognitives.

Ainsi, considérer que les capacités motrices d'un robot lui permettent d'élaborer des comportements intelligents est une erreur. La cognition humaine doit être appréhendée dans son ensemble pour pouvoir l'intégrer dans un robot et lui permettre de produire des comportements humains complexes. Les affects (émotions et motivations), les valeurs, les normes sociales jouent un rôle important dans la cognition humaine. Les aspects culturels aussi doivent être pris en compte : les connaissances historiques et l'héritage culturel apparaissent comme des aspects importants de la cognition.

Implémentation des modèles incarnés de la cognition dans un robot

L'une des premières implémentations des théories de la cognition 4E dans un robot a été effectuée par Escobar, Hermosillo et Lara (2012) (cité par Lara et al., 2018). Deux caméras ont été disposées sur un robot capable de se mouvoir dans l'espace. Les auteurs ont implémenté dans ce robot une modélisation lui permettant de percevoir son environnement et de situer son corps physique dans ce dernier. La modélisation intégrée dans ce robot lui permettait aussi de se mouvoir de façon autonome. Dans un premier temps, le robot avançait jusqu'à rencontrer un obstacle. Grâce au modèle implémenté, le robot était capable d'associer ses perceptions de l'environnement avec les conséquences de sa trajectoire en ligne droite via des capteurs. Lorsque le robot avançait droit devant lui et entrait en collision avec un élément, il associait ses perceptions visuelles capturées par ses caméras avec les perceptions « tactiles » qu'il percevait grâce à ses capteurs. Le modèle implémenté détectait qu'une collision correspondait à une erreur de prédiction. Ainsi, au fur et à mesure que le robot se déplaçait dans son environnement

et interagissait avec ce dernier en entrant en collision avec des obstacles, le robot pouvait déduire les régularités de son environnement et produire des prédictions sur ses futures actions motrices (prédire s'il allait, ou non, rencontrer un obstacle en s'engageant dans une certaine trajectoire). Les sorties de ce modèle sont des prédictions visuelles et tactiles sur la présence d'obstacle et produites lorsque le robot se déplace. Ces prédictions sont binaires, où 1 correspond à la présence d'un obstacle. Grâce à ses capteurs, le modèle est capable de détecter si une prédiction est erronée pour ensuite modifier les poids des connections entre les neurones afin de produire des prédictions plus fiables à l'avenir. Néanmoins, ce simple robot ne conçoit que le rôle joué par le corps dans la cognition, alors qu'il existe d'autres formes d'ancrage à prendre en compte (Barsalou, 2020).

Les théories de l'intelligence humaine

Depuis le début, j'évoque le terme « comportement intelligent » sans le définir précisément. Il est néanmoins important de préciser ce qu'est un comportement intelligent. Dans les travaux de Lara et al. (2018), « l'intelligence peut être conçue comme un comportement qui amène un agent à exécuter correctement une tâche située dans un certain contexte ». Dans la publication de Barsalou (2020), l'auteur ne définit pas le terme « intelligence » malgré son utilisation à plusieurs reprises. Enfin, Mingon et Sutton (2021) n'évoquent jamais ce terme dans leurs travaux et préfèrent parler de « comportements habiles ». Plusieurs chercheurs en psychologie ont pourtant tenté de définir ce qu'est l'intelligence et ont développé plusieurs théories à ce sujet.

La théorie des intelligences multiples, conceptualisée par Gardner en 1983, suggère qu'il existe 8 formes d'intelligences. Une des différentes formes d'intelligence, l'intelligence interpersonnelle, est définie comme l'aptitude d'un individu à comprendre autrui et à communiquer avec lui afin de prédire ses intentions (Eysenck, 2000). Selon cette théorie, un robot qui fait preuve d'empathie aurait une intelligence interpersonnelle élevée. Ainsi, un robot intelligent serait capable de détecter les émotions éprouvées par autrui et de les comprendre. Par conséquent, un robot intelligent serait un robot ayant intégré ce que sont les émotions et pouvant les simuler. Cette forme d'intelligence concorde avec l'approche incarnée de la

cognition développée par Barsalou (2020) et celle conceptualisée par Mingon et Sutton (2021). Les autres formes d'intelligences théorisées par Gardner confortent aussi une vision incarnée de la cognition, où la cognition humaine étant capable de produire des comportements intelligents doit être appréhendée dans son ensemble. En effet, cette approche intégrative de l'intelligence définit un large panel de comportements pouvant être définis comme intelligents. Celle-ci ne va pas à l'encontre d'une vision de la cognition qui émerge des interactions entre les processus cognitifs, les modalités perceptives, le corps, l'environnement physique et l'environnement social. Cependant, cette théorie de l'intelligence est souvent critiquée car elle n'a pas pu être soutenue par des preuves empiriques.

Une autre conception de l'intelligence est celle proposée par Binet et Simon en 1905. Selon eux, seul un test peut mesurer l'intelligence d'un individu (Eysenck, 2000). Dans cette approche métrique de l'intelligence, un comportement intelligent est un comportement permettant l'acquisition d'un bon score à ce test. Cette définition de l'intelligence est en désaccord avec la conception de l'intelligence portée par Lara et al. (2018). En effet, selon Lara et al. (2018), fuir une voiture qui nous fonce dessus est un comportement intelligent, pourtant ce comportement ne permet pas d'obtenir un bon score au test de Binet et Simon.

Les différentes théories sur l'intelligence humaine sont anciennes. Les récentes recherches sur les théories incarnées de la cognition ont tendance à admettre qu'un comportement intelligent est un comportement inhérent à des capacités cognitives de haut-niveau et permettant à l'individu d'accomplir une tâche adaptée à la situation rencontrée. Que ce soit le fait de fuir une voiture qui nous fonce dessus ou la pratique d'une danse traditionnelle, il existe de nombreux comportements pouvant être définis comme intelligents. Ce qui est certain, c'est qu'un comportement intelligent implique l'engagement de nombreux aspects de la cognition humaines, difficiles voire impossibles à intégrer dans un robot de nos jours. Pour intégrer une vision incarnée de la cognition humaine dans un robot, il faudrait que ce dernier soit capable d'appréhender des représentations modales qui sous-tendent l'apprentissage de connaissances. Il serait aussi nécessaire que le robot intègre des variables sociales ainsi que la dynamique du Cycle de l'Action Située développé par Barsalou (2020). Envisager une architecture aussi

complexe est nécessaire pour qu'un robot puisse, dans le futur, reproduire l'intelligence humaine dans son ensemble et puisse déployer des comportements élaborés.

Bibliographie

Lara, B., Astorga, D., Mendoza-Bock, E., Pardo, M., Escobar, E., & Ciria, A. (2018). Embodied cognitive robotics and the learning of sensorimotor schemes. *Adaptive Behavior*, 26(5), 225-238.

Harvey, K. (2018). An open-ended approach to Piagetian development of adaptive behavior. *Open Access Library Journal*, 5(03), 1.

Mingon, M., & Sutton, J. (2021). Why robots can't haka: skilled performance and embodied knowledge in the Māori haka. *Synthese*, 1-29.

Barsalou, L. W. (2020). Challenges and Opportunities for Grounding Cognition. *Journal of Cognition*, 3(1).

Eysenck, M. W. (2000). *Psychology: A student's handbook*. Taylor & Francis.