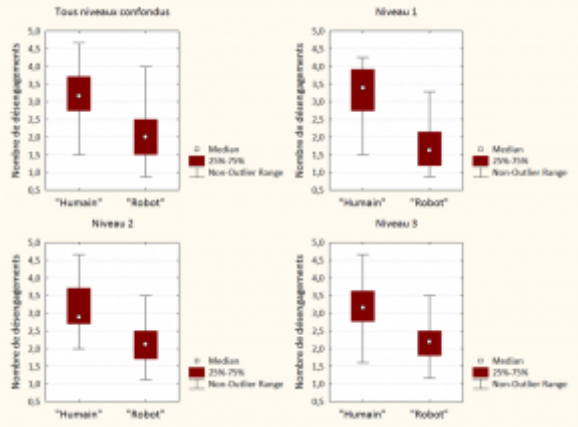
Cahiers
#Costech

> Loïc Deschamps
> Charles Lenay
> Katia Rovira
> Dominique Aubert

L'engagement mutuel comme condition de succès d'un test de Turing non-verbal

une expérience de croisement perceptif minimaliste



> #Numéro 1
> Suppléance perceptive et cognition sociale
> Rapports de recherche
> CRED - Cognitive Research and Enactive Design (Costech-UTC)
> Cognition sociale - > Enaction - > Sciences cognitives

Citer cet article

Deschamps, Loïc., Lenay, Charles., Rovira, Katia., Aubert, Dominique. "L'engagement mutuel comme condition de succès d'un test de Turing non-verbal. une expérience de croisement perceptif minimaliste", 25 juillet 2017, *Cahiers Costech*, numéro 1.
URL https://www.costech.utc.fr/CahiersCostech/spip.php?article44

Résumé

54 participants ont interagi dans un environnement virtuel minimaliste à l'aide d'un dispositif tactile. Chaque participant a rencontré successivement soit un autre participant, soit un robot programmé pour imiter le comportement humain selon différents degrés de complexité. Selon le principe du célèbre « jeu d'imitation » (Turing, 1950), les participants devaient identifier, tout au long de 16 essais d'une minute, la nature de leur partenaire (humain ou robot). Le caractère minimaliste du dispositif a permis de suivre la co-construction de la dynamique d'interaction, et de confronter cette dynamique avec la réponse fournie individuellement par les participants après chaque essai. En accord avec l'expérience princeps du croisement perceptif minimaliste (Auvray, Lenay, & Stewart, 2009), les résultats suggèrent que l'engagement mutuel des participants précède la reconnaissance individuelle des partenaires de l'interaction. Ces résultats sont discutés au regard d'une approche énactive de la cognition sociale.

Test de Turing non-verbal ; Engagement mutuel ; Croisement perceptif ; Minimalisme ; Cognition sociale

Auteur(s)

Loïc Deschamps, Chercheur en psychologie cognitive / développement ; post-doctorant au Costech-UTC, sa thèse de doctorat est intitulée « Suppléance perceptive et cognition sociale : étude des interactions tactiles minimalistes ».

Charles Lenay, Professeur de sciences cognitives et de philosophie des sciences, ex-directeur du

COSTECH . Habilité à diriger les recherches en Philosophie (17e) et en Histoire des sciences (72e), il consacre l'essentiel de ses recherches aux technologies cognitives : comment les outils participent à l'activité cognitive : raisonnement, mémorisation, perception, interaction,…

Katia Rovira, Professeur de psychologie à l'Université de Rouen, membre du Centre de Recherche sur les Fonctionnements et Dysfonctionnements Psychologiques (CRFDP, EA 7475), spécialisée sur le handicap visuel et la suppléance perceptive (stratégies perceptives et catégorielles) ainsi que l'attention conjointe, la discrimination interpersonnelle, la perception de l'espace dans le contexte du handicap visuel.

Dominique Aubert, Ingénieur d'étude spécialisé dans le développement de logiciels pour la recherche cognitive et le design d'interaction.

Plan

1 - Introduction

2 - Engagement mutuel dans les interactions précoces

3 - Le croisement perceptif minimaliste : un paradigme expérimental pour l'étude de la cognition sociale

4 - Méthode
- 4.1 - Population
- 4.2 - Matériel
- 4.3 - Procédure

5 - Résultats
- 5.1 - Analyse des réponses fournies par les participants
  - 5.1.1 - Condition Humain vs condition Robot
  - 5.1.2 - Programmes clos s Programmes d'interaction
  - 5.1.3 - Modalités collectives de reconnaissance dans la condition Humain
- 5.2 - Analyse des trajectoires perceptives
  - 5.2.1 - Durée de l'accroche mutuelle
  - 5.2.2 - Variabilité de l'accroche mutuelle
  - 5.2.3 - Désengagements de l'accroche mutuelle

6 - Discussion

7 - Conclusion

8 - Références

1 - Introduction

Dans la plupart des études de la cognition sociale, des participants observent passivement des situations sociales extérieures avant de produire un jugement après-coup (p. ex. Gallotti & Frith, 2013). Dans ces conditions, la performance des participants est nécessairement interprétée en termes de mentalisation implicite ou explicite (p. ex. Frith, 2008), et la cognition sociale est considérée comme une habileté strictement individuelle à traiter les informations sociales (voir Hutto, 2004). En conséquence, l'interaction sociale n'est qu'un simple input sur

lequel opèrent des modules internes (Gergely & Watson, 1996, 1999). Cette approche individualiste de la cognition sociale revient à négliger l'*embodiment* des partenaires sociaux (Gallagher, 2001), ainsi que le rôle du processus d'interaction dans la compréhension sociale (De Jaegher, Di Paolo, & Gallagher, 2010 ; Di Paolo & De Jaegher, 2016 ; Lenay, 2017).

En considérant la cognition comme une action incarnée (Stewart, Gapenne, & Di Paolo, 2010 ; Varela, Thompson, & Rosch, 1991), l'énaction ouvre la voie à une approche active de la cognition sociale. Celle-ci renverrait alors à un processus dynamique de régulation des actions et des intentions qui se déploie à même la rencontre d'agents incarnés et situés (De Jaegher & Di Paolo, 2007 ; Fogel & Garvey, 2007). Il est donc question de prendre en compte l'engagement actif des partenaires d'interaction (Reddy & Morris, 2004), dont la rencontre fait émerger un ensemble de coordinations spatiotemporelles plus ou moins complexes (Fogel, 1993 ; Stern, 1985). Ces coordinations, qui concernent les mouvements, les attitudes et les vocalisations, constituent le cœur d'une construction collective de sens (*participatory sense-making*) : le sens est généré, transformé et toujours (re-)négocié à travers la dynamique d'interaction (De Jaegher et al., 2010). L'interaction sociale peut alors être définie comme « un couplage régulé entre au moins deux agents autonomes, où la régulation vise les aspects du couplage lui-même, de façon à ce qu'il constitue une organisation autonome émergente dans le domaine de la dynamique relationnelle, sans détruire dans le processus l'autonomie des agents impliqués (bien que la portée de celle-ci puisse être augmentée ou réduite) » (De Jaegher & Di Paolo, 2007, p. 8, *traduction personnelle*). Le concept d'autonomie est primordial ici : il renvoie à l'auto-organisation d'un réseau de processus qui se maintient malgré des conditions précaires (Di Paolo, 2009). Cette approche de la cognition sociale ne se focalise donc pas exclusivement sur les mécanismes individuels, mais aussi sur le processus d'interaction, de façon à prendre en compte une intrication fondamentale entre les facteurs individuels et les facteurs collectifs dans le processus de la compréhension sociale (Di Paolo & De Jaegher, 2016 ; Lenay & Stewart, 2012 ; Lenay, 2017).

2 - Engagement mutuel dans les interactions précoces

Cette approche énactive de la cognition sociale résonne avec certaines approches du développement de l'enfant, constructivistes et dynamiques, qui soulignent l'importance fondamentale des interactions précoces. En

effet, dès la naissance, l'enfant manifeste une forme d'*altéroception*, c'est-à-dire une capacité de perception émotionnelle immédiate d'autrui qui permet de participer à des échanges de nature dialogique avec des partenaires sociaux (Braten, 1998 ; Trevarthen, 1993). Ces échanges se caractérisent par une improvisation mutuelle des mouvements et des vocalisations qui engage émotionnellement les deux partenaires dans un rythme commun. Ces coordinations sensori-motrices précoces, qui ont lieu quotidiennement dans des situations contextualisées et pragmatiques, constitueraient la base des compétences ultérieures de l'enfant en matière de compréhension sociale (Gallagher, 2001 ; Gallagher & Hutto, 2008).

Bien entendu, d'un point de vue individuel, l'engagement dans une telle coordination requiert une sensibilité perceptive aux contingences sociales qui relient les comportements (Crown, Feldstein, Jasnow, Beebe, & Jaffe, 2002). La sensibilité aux contingences sociales est en effet mise en évidence dès les premiers mois de la vie de l'enfant, qui parvient à se synchroniser avec l'adulte pour prendre part à des proto-conversations et manifester des comportements spécifiques aux contextes de communication, comme le *turn-taking* (Bateson, 1979). Selon une approche classique, ces relations spatio-temporelles seraient traitées par un module inné, d'abord dédié aux stimulations parfaitement contingentes (Gergely & Watson, 1996, 1999), c'est-à-dire produites de façon immédiate par l'action qui en est à l'origine (par exemple, les stimulations visuelles, auditives, tactiles et proprioceptives produites par un hochet que l'enfant agiterait avec sa main). Progressivement, ce module se spécialiserait pour les contingences élevées mais imparfaites, qui se caractérisent par la présence d'un délai temporel, d'une durée brève mais variable (et accompagné ou non d'un décalage spatial), entre l'action produite et les stimulations reçues (ibid.). La préférence précoce pour les contingences imparfaites serait adaptative, dans la mesure où elles permettent à l'enfant de s'ouvrir au monde social (Rochat & Striano, 1999).

D'un point de vue expérimental, il a été effectivement démontré que le bébé est sensible non seulement à la qualité expressive et émotionnelle de la mère (paradigme du *still face*, Tronick, Als, & Adamson, 1979), mais également au timing et aux contingences des comportements expressifs. En effet, lorsque la dyade mère-enfant interagit via des écrans interposés, et que l'enfant est subitement confronté à un replay du comportement de sa mère (paradigme de double-vidéo), celui-ci manifeste un ensemble de

comportements caractéristiques de détresse, voire de retrait (Murray & Trevarthen, 1985 ; Nadel, Carchon, Kervella, Marcelli, & Réserbat-Plantey, 1999). En outre, il est intéressant de noter que dans une variante de cette expérience, il a été montré que la mère est également perturbée lorsqu'elle est confrontée à un replay du comportement de son enfant (Murray & Trevarthen, 1986). Ces réactions semblent indiquer que les comportements des deux protagonistes ne sont pas simplement corrélés de façon accidentelle, mais réellement coordonnés, et que la perturbation des contingences affecte à la fois l'enfant et la mère (McGann & De Jaegher, 2009). Selon une approche énactive, pour rendre compte pleinement de ce phénomène, il s'agit de ne pas se focaliser uniquement sur des mécanismes internes (comme la détection des contingences), mais de tenir compte également de la dynamique d'interaction elle-même (De Jaegher et al., 2010 ; Di Paolo & De Jaegher, 2016 ; Lenay, 2017). En effet, la focalisation sur les mécanismes individuels masque un ensemble de paramètres qui pourraient également intervenir dans l'interprétation des réactions de l'enfant et de la mère. Par exemple, il convient de ne pas négliger l'impact de la participation active des deux protagonistes dans l'établissement et le maintien de l'interaction (par exemple, les ajustements comportementaux, ou encore les tentatives de régulation du comportement du partenaire). Sans cette dimension active et dynamique, l'interaction se trouve au final réduite à une simple information, déconnectée du flux continu de l'activité sociale.

Comme nous l'avons déjà mentionné, les interactions précoces sont fondamentales dans certaines approches du développement. Par exemple, l'engagement de l'enfant dans les situations sociales serait nécessaire pour qu'il puisse avoir conscience qu'autrui est un être intentionnel (Hobson, 2002). À partir de sa propre intentionnalité, l'enfant peut alors éprouver autrui, non pas comme un observateur passif à la troisième personne, mais comme un être à la seconde personne, qui participe activement à l'interaction sociale (Reddy & Morris, 2004). De ce point de vue, l'interaction serait un processus dynamique qui engage les deux partenaires sur un versant émotionnel plutôt que représentationnel (Reddy, 2003). Les actions expressives peuvent alors être considérées comme des moyens d'investir l'espace commun et de susciter l'engagement social d'autrui (Krueger, 2011 ; Stern, 1985 ; Trevarthen, 2001). L'espace commun est un espace objectif, vécu à travers le corps et ses possibilités d'action. Les interactions reconfigurent cet espace en un espace social, où l'engagement mutuel des partenaires donne lieu à des coordinations émotionnelles et intersubjectives. La contribution

active des deux partenaires offre à chacun des affordances de nature sociale, qui structurent activement l'espace de l'interaction : les comportements déployés dans cet espace sont autant de contraintes et d'opportunités d'action disponibles pour les individus impliqués dans le processus d'interaction (Krueger, 2011). La ressaisie de ces affordances par les partenaires d'interaction permet l'émergence d'un couplage co-régulé qui donne du sens aux échanges (De Jaegher & Di Paolo, 2007 ; Lenay, 2017 ; Lenay & Stewart, 2012). Suivant cette voie, toute étude de la cognition sociale doit se donner les moyens de rendre compte de l'articulation entre les dimensions individuelles et la dynamique collective dans les situations sociales qu'elle prétend expliquer (Di Paolo & De Jaegher, 2016 ; Lenay, 2017).

3 - Le croisement perceptif minimaliste : un paradigme expérimental pour l'étude de la cognition sociale

D'un point de vue méthodologique, cette articulation est clairement mise en évidence dans le cadre du paradigme expérimental du croisement perceptif minimaliste (voir p. ex. Lenay, 2017 ; Lenay & Stewart, 2012). Ce paradigme a pour objectif de proposer des schèmes explicatifs de la cognition sociale (Lenay, 2017), à partir de l'analyse des croisements perceptifs rendus possibles par l'usage de médiations techniques minimalistes (en l'occurrence, des dispositifs de suppléance perceptive, voir Lenay, 2006 ; Lenay, Auvray, Sebbah, & Stewart, 2006). Ces médiations particulières permettent de simplifier à l'extrême le répertoire des actions et des retours sensoriels, de façon à rendre possible une analyse précise du déploiement spatiotemporel des activités individuelles et de la dynamique collective (Lenay & Stewart, 2012).

Dans l'étude princeps, un dispositif tactile permet à deux participants d'interagir dans un environnement numérique unidimensionnel (Auvray et al., 2009 ; Lenay et al., 2006). Les participants y déplacent latéralement un curseur au moyen d'une souris d'ordinateur. Ce curseur est « sensible » aux objets présents dans l'espace numérique : lorsqu'il croise un objet, cela provoque une stimulation tactile en tout ou rien sous le doigt du participant. Chaque participant peut rencontrer un objet statique, un objet mobile et le curseur déplacé par l'autre participant. Il est important de noter que l'objet mobile est en réalité un leurre attaché au curseur de chaque participant par un lien virtuel rigide (voir Figure 1). De cette façon, le leurre mobile se déplace exactement comme se déplace le curseur auquel il est attaché, mais n'est pas sensible aux objets rencontrés.

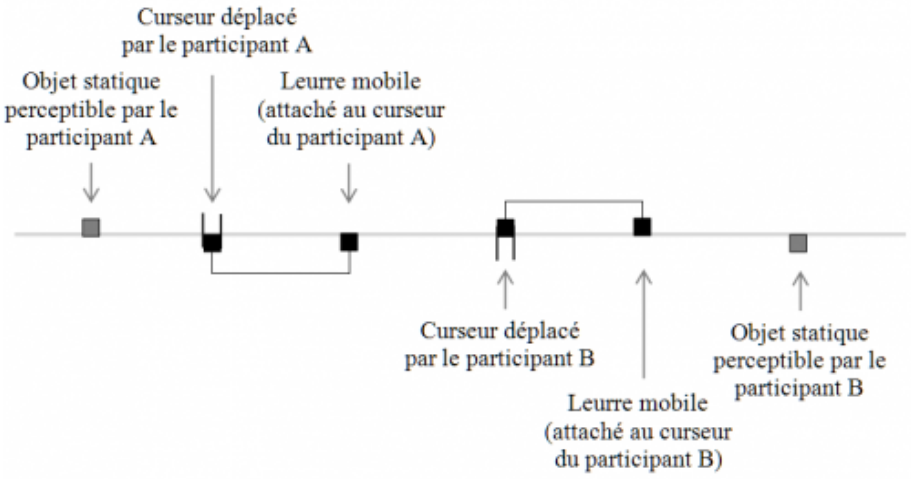


Figure 1

Dispositif expérimental utilisé dans l'expérience princeps du croisement perceptif minimaliste (Auvray et al., 2009 ; Lenay et al., 2006). Dans un espace unidimensionnel, chaque participant déplace un curseur pouvant détecter trois types d'objets : un objet statique, un leurre mobile, et le curseur de l'autre participant.

À première vue, les résultats suggèrent que les participants réussissent la tâche. La majorité des clics est effectuée quand les curseurs des participants se croisent dans l'espace numérique. Une analyse détaillée révèle cependant que la fréquence des stimulations reçues par les participants en fonction du type d'objet rencontré est plus importante pour le curseur de l'autre participant (0.52) que pour l'objet statique (0.32) et pour le leurre mobile (0.15), et que les ratios entre le nombre de clics et de stimulations pour le curseur de l'autre participant et pour le leurre mobile ne présentent pas de différence significative (respectivement 1.26 et 1.51). Autrement dit, les participants ne cliquent pas *plus fréquemment* quand la stimulation est due au curseur de l'autre participant que quand elle est due au leurre mobile. Cette observation tend à montrer que les participants ne font pas la différence entre le curseur de l'autre participant et le leurre mobile qui s'y trouve attaché.

Comment expliquer dans ce cas la réussite de la tâche ? L'explication la plus plausible est que cette réussite présente un caractère foncièrement *collectif* : autrement dit, elle s'explique par une dynamique qui échappe largement aux participants considérés individuellement. En effet, une analyse des trajectoires perceptives montre une corrélation négative globale (-.72) entre la vitesse du mouvement avant la stimulation et l'accélération des participants après la stimulation. Ceci peut s'expliquer par la nature minimaliste du dispositif : il n'y a pas de perception périphérique, et par conséquent, la rencontre d'une source de stimulation

est un évènement local qui échappe le plus souvent au participant (à peine rencontrée, la source de stimulation « disparaît » car elle est dépassée par le curseur). Pour reproduire cette stimulation, le participant doit alors inverser sa propre trajectoire. Ce mouvement est satisfait, c'est-à-dire produit une nouvelle stimulation, si l'objet rencontré est statique, ou s'il incarne une activité perceptive engagée dans la même recherche intentionnelle. La fréquence des stimulations dues à la rencontre de l'autre participant est donc plus élevée parce que l'activité perceptive de chaque participant est organisée selon l'activité perceptive de leur partenaire. Tout comme la perception d'un objet statique est réalisée par la constitution d'un invariant sensori-moteur qui attire l'activité d'un observateur (O'Regan & Noë, 2001), la rencontre de deux activités perceptives intentionnellement dirigées l'une vers l'autre fait émerger un attracteur de la dynamique collective (Auvray et al., 2009). Les différences constatées en ce qui concerne les fréquences de clics peuvent être attribuées à la stabilité du couplage, qui est rendue possible par la rencontre des stratégies déployées individuellement (Di Paolo, Rohde, & Iizuka, 2008). De cette façon, la reconnaissance de l'autre, dans les conditions du croisement perceptif minimaliste, est intrinsèque à l'activité perceptive partagée et à la dynamique collective qu'elle engendre (Auvray et al., 2009). Il s'agit là d'une démonstration expérimentale de l'autonomie du processus d'interaction : l'interaction ne joue pas seulement un rôle contextuel ou déclenchant, mais elle *constitue* la cognition sociale (De Jaegher et al., 2010).

Une seconde expérience a été réalisée pour étayer cette explication (Lenay & Stewart, 2012) et répondre à des critiques faisant valoir que la réussite de la tâche dépend en premier lieu de la formation individuelle de jugements sociaux, qui conduisent les participants à cliquer (Michael & Overgaard, 2012). Dans celle-ci, le dispositif et les conditions sont strictement identiques, hormis le fait qu'un évènement sonore distinct est attribué à chacune des trois sources de stimulations. À la fin de chaque essai de 2 minutes, les participants sont invités à associer chacun des trois évènements sonores avec la source de stimulation correspondante. Les résultats montrent des scores élevés d'association. Il semble donc que les évènements sonores attribués à chaque source de stimulation suffisent pour lever l'ambigüité dans les différentes situations d'interaction, et que les dynamiques d'interaction spécifiques aux différents objets rencontrés peuvent être reconnues par les participants. En d'autres termes, les résultats sont expliqués par une ressaisie individuelle de la dynamique collective (ibid.), ce qui démontre la

nécessité de prendre en compte l'intrication profonde des mécanismes individuels et de la dynamique collective dans les études de la cognition sociale (Di Paolo & De Jaegher, 2016 ; Lenay, 2017).

Dans la continuité de ces expériences, il nous a semblé intéressant de proposer une tâche de croisement perceptif minimaliste qui permettrait d'étudier des interactions strictement dyadiques, en l'absence de tout distracteur. Pour cela, nous avons développé un protocole expérimental qui reprend le principe du célèbre « jeu d'imitation » (Turing, 1950). Dans notre étude, chaque participant interagit à l'aide d'un dispositif tactile soit avec un autre participant, soit avec un robot présentant différents degrés de complexité, et a pour tâche d'identifier la nature de son partenaire (humain ou robot). Les degrés de complexité des robots sont élaborés pour induire des dynamiques d'interaction plus ou moins complexes, notamment en présentant des comportements plus ou moins contingents. L'objectif de cette étude est de tenter de caractériser les paramètres objectifs qui conduisent les participants à reconnaître comme humaine l'activité qu'ils rencontrent dans l'environnement virtuel minimaliste. Dans une visée plus générale, il est question d'évaluer l'impact de la dynamique collective sur les comportements individuels (Lenay, 2017 ; Lenay & Stewart, 2012), et de démontrer expérimentalement que l'engagement mutuel est au moins partiellement constitutif de la cognition sociale (De Jaegher et al., 2010).

Nous posons l'hypothèse que malgré la relative complexité de certains robots, les participants seront capables de reconnaître autrui grâce à l'émergence d'une dynamique d'interaction caractéristique qui résulte de l'engagement mutuel des deux partenaires. Par ailleurs, si certains robots, en particulier ceux qui pourront produire des degrés variés de contingence, pourront occasionnellement induire les participants en erreur, nous nous attendons à ce que de telles erreurs diminuent avec la répétition des essais. Cette répétition entraînera en effet une stabilisation de la dynamique interpersonnelle, qui pourra devenir un support de plus en plus pertinent pour la reconnaissance mutuelle des partenaires.

4 - Méthode

4.1 - Population

54 participants ont pris part à cette expérience. Chaque participant a été confronté à deux autres participants et à deux robots. Les participants

sont des étudiants de l'Université de Rouen et de l'Université de Technologie de Compiègne âgés entre 18 et 25 ans, qui ont été affectés aléatoirement à un des trois groupes expérimentaux définis par le niveau de complexité des robots. Ils n'avaient aucune connaissance préalable du dispositif, des objectifs ou du contexte théorique de l'étude.

Le protocole utilisé est conforme aux recommandations de la Déclaration d'Helsinki de Juin 1964 (modifiée au terme de la 64e Assemblée Générale de l'Organisation Mondiale de la Santé d'octobre 2013). Tous les participants ont signé un document présentant le contexte général de l'étude et attestant de leur consentement éclairé pour l'utilisation des données à des fins scientifiques.

4.2 - Matériel

Le dispositif expérimental est composé de 7 ordinateurs portables, dont un serveur pour gérer les connexions en réseau et enregistrer les données, 3 ordinateurs dédiés aux participants humains et 3 ordinateurs dédiés au contrôle des robots. Tous ces ordinateurs sont interconnectés via un réseau local filaire (Local Area Network).

Les participants prennent place devant un ordinateur portable équipé d'une souris optique et d'un boîtier de 16 stimulateurs tactiles (deux cellules Braille piézoélectriques adjacentes). La souris permet au participant de déplacer un curseur à l'écran au moyen de sa main dominante, tandis que l'index de sa main non-dominante est placé sur les stimulateurs tactiles. Quand le curseur rencontre un autre curseur, qu'il soit déplacé par une autre participant ou contrôlé par un robot, cela active la montée simultanée des 16 stimulateurs tactiles, en tout-ou-rien.

L'espace dans lequel les participants se déplacent est un espace unidimensionnel horizontal de 200 pixels de long et bouclant sur lui-même, de sorte que les participants ne sont pas conscients de limites ou de frontières. Du point de vue d'un observateur extérieur, le franchissement du bord gauche de l'écran par un curseur se traduit par sa réapparition à droite et vice et versa.

Le curseur est un champ récepteur de 2 pixels de long (le « corps-percevant », qui permet la détection des autres curseurs), auquel correspond topologiquement un « corps-image » de 2 pixels de long (que le « corps-percevant » d'un autre participant peut détecter). Les curseurs

se déplaçant dans un espace unidimensionnel horizontal, les déplacements verticaux de la souris ne sont pas pris en compte.

Les robots contrôlent également le déplacement d'un corps-image de 2 pixels de long, via des scripts LUA qui produisent des comportements déterminés selon trois degrés de complexité. Ceux-ci sont définis par la manipulation de trois paramètres : la vitesse et l'amplitude du mouvement, et la capacité à s'arrêter. Ces paramètres ont été choisis pour influencer les contingences qui relieront les comportements des robots avec ceux des participants, de façon à induire une variété de dynamiques d'interaction. Hormis ces paramètres variables d'un niveau de complexité à l'autre, tous les robots partagent les mêmes règles de base. Leur position de départ dans l'espace est déterminée aléatoirement à chaque essai. Chaque changement de direction décrit une courbe sinusoïdale, c'est-à-dire une décélération suivie d'une ré-accélération inverse : l'oscillation ainsi produite forme ce que nous appelons un cycle. Les robots du niveau 1 (robots simples) décrivent un cycle régulier d'une amplitude de 35 pixels et d'une vitesse maximale de 120 pixels par seconde, ce qui se traduit par une oscillation latérale régulière. Les robots du niveau 2 (robots intermédiaires) ont une amplitude et une vitesse maximale qui varie périodiquement (± 40% des valeurs du niveau 1 appliquées à chaque demi-cycle). Les robots du niveau 3 (robots complexes) sont identiques aux robots intermédiaires, avec l'ajout d'une certaine probabilité d'arrêt (5% de chances de s'arrêter à chaque demi-cycle pour une durée comprise entre 1 et 5 secondes).

Pour chaque niveau de complexité, deux types de robots sont implémentés : des robots élaborés en termes de programme clos (les cycles de mouvements sont déployés indépendamment des mouvements du participant humain), et des robots élaborés en termes de programmes d'interaction (les cycles de mouvements sont déclenchés par la rencontre du participant humain) (voir Figure 2). Dans le cas des programmes clos, une légère translation est appliquée au début de chaque cycle (5 pixels dans une direction aléatoire), de façon à ce que ces robots ne soient pas identifiés par une zone d'oscillation persistant tout au long de l'essai.

Au final, les robots sont capables, en fonction de leur type et de leur niveau de complexité, d'influencer la dynamique d'interaction, en déployant des comportements non-contingents (programmes clos), des comportements parfaitement contingents (Programmes d'interaction - Robots simples), et des comportements imparfaitement contingents

(Programmes d'interaction - Robots intermédiaires), qui peuvent être associés à une capacité de désengagement et de réengagement de la dynamique d'interaction (Programmes d'interaction - Robots complexes).

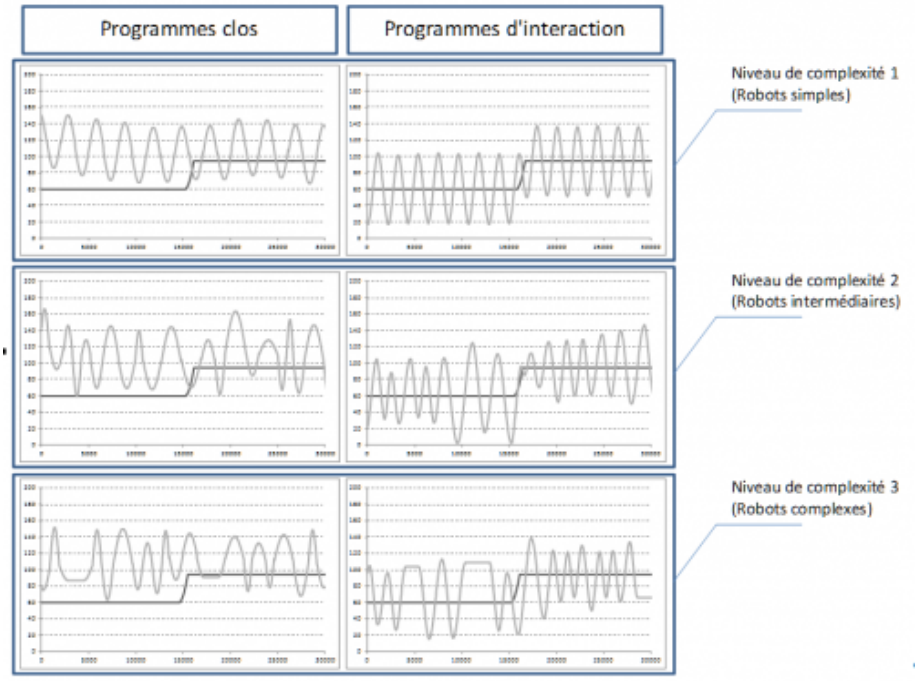


Figure 2

Trajectoires déployées par les programmes clos et les programmes d'interaction selon les trois niveaux de complexité. Pour chaque graphique, l'axe des abscisses représente le temps (en ms), et l'axe des ordonnées représente la position en x (en pixels). La courbe claire représente le mouvement des programmes. La courbe foncée représente la trajectoire hypothétique d'un participant humain, qui occupe initialement une position fixe, avant de se déplacer légèrement et de s'arrêter à une nouvelle position.

4.3 - Procédure

Trois groupes indépendants de 18 participants ont été constitués. Chaque groupe est confronté à deux robots d'un même niveau de complexité (un robot en programme clos et un robot en programme d'interaction). Dans chaque groupe, les participants sont divisés en trinômes, et chaque trinôme donne lieu à une passation expérimentale. Chaque passation est constituée de 17 essais : un essai de familiarisation et 16 essais expérimentaux d'une minute, dans lesquels chaque participant rencontre quatre fois chacun des deux autres participants (condition Humain) et quatre fois chacun des deux robots (condition Robot) dans un ordre aléatoire. Notons que lorsque deux participants sont confrontés l'un à l'autre dans un essai donné, le troisième participant est nécessairement confronté à l'un des deux

robots. Toutefois, dans certains essais, les trois participants sont confrontés respectivement et simultanément à un des deux robots.

Les participants ne sont pas aveuglés pendant la tâche, mais aucune information n'est visible sur leur écran. Les passations se déroulent dans deux pièces adjacentes de l'université, ou dans une seule et même pièce. Dans tous les cas, des précautions sont prises pour que les participants ne puissent ni se voir ni s'entendre (casques insonorisés) pendant les essais : les seules interactions possibles sont celles qui sont permises par le dispositif tactile.

Pour tous les groupes, un essai de familiarisation de deux minutes est d'abord effectué. Cet essai consiste à explorer l'espace au moyen de la souris d'ordinateur et à rencontrer un objet statique de 2 pixels de long. Cette étape est nécessaire pour prendre connaissance du dispositif. Les participants sont alors informés que l'activation des stimulateurs tactiles résulte de la rencontre de leur curseur avec l'objet présent dans l'espace numérique. Ensuite, ils sont invités à estimer subjectivement la longueur de l'espace, en constatant la distance qui sépare deux rencontres successives avec l'objet statique quand ils déplacent la souris dans une même direction (le curseur franchit alors le bord de l'espace, réapparaît de l'autre côté et rencontre à nouveau le même objet). À la fin de la phase de familiarisation, l'expérimentateur s'assure que le principe du dispositif est bien compris, et délivre la consigne pour les essais expérimentaux.

Les participants sont informés qu'ils pourront interagir soit avec un robot, soit avec un humain pendant 16 essais d'une minute. À la fin de chaque essai, ils seront invités à remplir une feuille de réponse sur laquelle ils devront indiquer s'ils pensent avoir interagi avec un robot ou avec un humain. Les participants ont aussi la possibilité de ne pas répondre. Aucune autre information sur la nature des robots, leur nombre et la façon dont ils se déplacent n'est délivrée. De plus, aucun feedback n'est donné aux participants concernant leurs réponses pendant la durée de l'expérience. Une pause de 5 minutes est proposée à la fin de l'essai 8, ce qui permet d'obtenir deux blocs de 8 essais. Pendant cette pause, les participants ne sont pas autorisés à communiquer les uns avec les autres.

5 - Résultats

5.1 - Analyse des réponses fournies par les participants

Dans un premier temps, nous nous intéressons aux réponses fournies par les participants. La distribution de ces données suivant une loi normale, nous avons effectué une analyse de variance (ANOVA et tests de Student), pour comparer les fréquences de réponse « Humain » en fonction du niveau de complexité des robots. Nous définirons d'abord la condition expérimentale (condition Humain vs condition Robot), puis le type de programme de la condition « Robot » (Programme clos vs Programme d'interaction) comme variables intragroupes. Enfin, nous nous focaliserons sur la condition Humain en comparant les modalités collectives de reconnaissance (reconnaissance mutuelle, reconnaissance unilatérale et non-reconnaissance) en fonction des deux blocs d'essais.

5.1.1 - Condition Humain vs condition Robot

Tous groupes confondus, la fréquence de réponses « Humain » dans la condition Humain est statistiquement supérieure à la fréquence de réponses « Humain » dans la condition Robot ($d = .28$; $t(51) = 8.2230$; $p < .01$) (voir Figure 3).

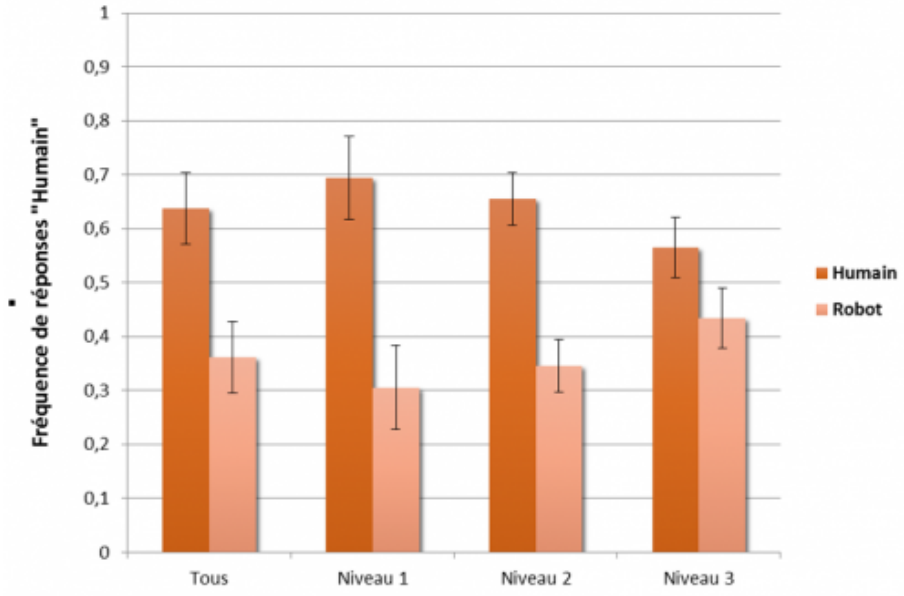


Figure 3

Fréquence de réponses « Humain » en fonction du niveau de complexité des robots et de la condition expérimentale (condition Humain vs condition Robot)

En ce qui concerne les différences entre les niveaux de complexité, nous

mettons en évidence un effet d'interaction significatif entre le niveau de complexité des robots et la condition expérimentale ($F(2,51) = 5.1495$; $p < .01$). En effet, la différence entre la fréquence de réponse « Humain » dans la condition Humain et la fréquence de réponse « Humain » dans la condition Robot diminue à mesure que le niveau de complexité des robots augmente. Les robots complexes sont plus souvent identifiés comme des humains ($d = .13$) que les robots intermédiaires ($d = .31$) et les robots simples ($d = .39$). Toutefois, une analyse post-hoc montre que la différence de fréquence de réponses « Humain » entre les conditions expérimentales est statistiquement significative pour tous les niveaux de complexité des robots (voir Figure 3 et Tableau 1).

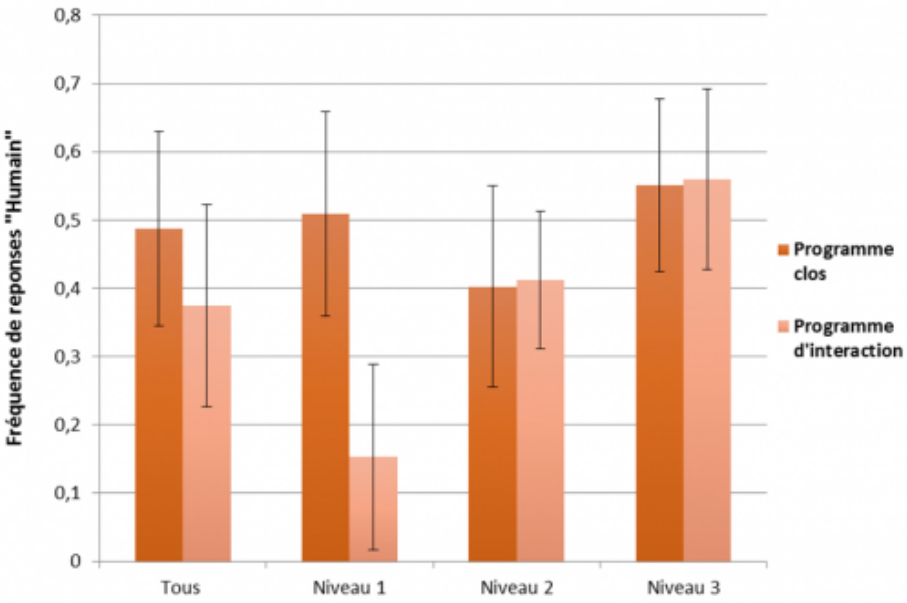


Figure 4

Fréquences de reponses « Humain » dans la condition Robot en fonction du niveau de complexité et du type de robot (Programme clos vs Programme d'interaction)

5.1.2 - Programmes clos s Programmes d'interaction

Nous nous focalisons maintenant sur la condition Robot, pour évaluer l'effet du type de robot (Programme clos vs Programme d'interaction) sur la fréquence de réponses « Humain » en fonction du niveau de complexité des robots. Nous observons que tous niveaux de complexité confondus, la fréquence de réponses « Humain » est plus élevée lorsque les participants sont confrontés aux programmes clos que lorsqu'ils sont confrontés aux programmes d'interaction ($d = .11$) (voir Figure 4).

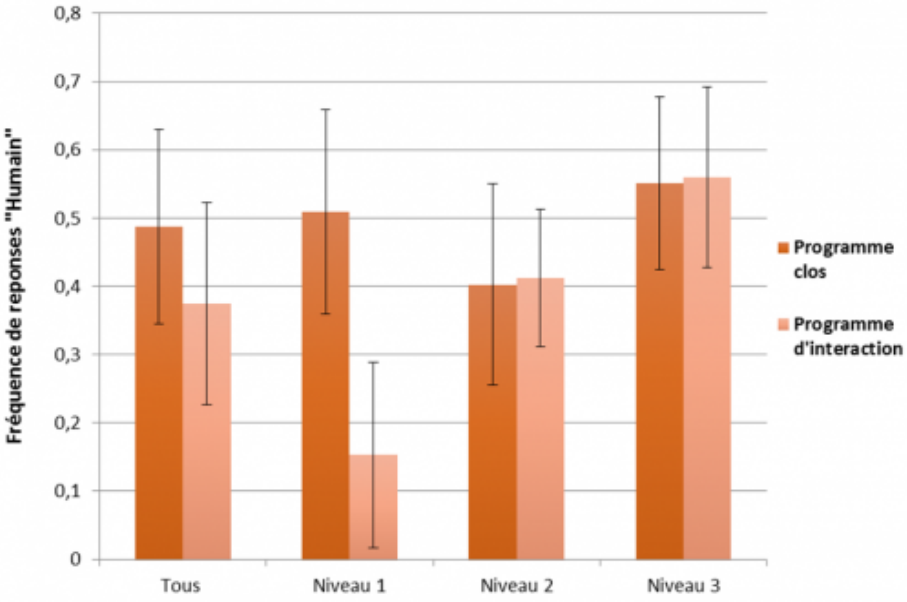


Figure 4

Fréquences de reponses « Humain » dans la condition Robot en fonction du niveau de complexité et du type de robot (Programme clos vs Programme d'interaction)

Mais nous ne pouvons pas conclure à un effet du type de robot, malgré une différence statistique marginale ($t(51) = 2.0040$; $p = .0504$). Toutefois nous constatons un effet d'interaction significatif entre le type de robot et le niveau de complexité ($F(2 ; 51) = 4.7035$; $p = .01$). Cette interaction se traduit par un effet significatif du type de robot pour les participants soumis au niveau 1 de complexité ($d = .36$; $t(17) = 3.2578$; $p < .01$), alors que nous ne pouvons pas conclure pour les niveaux 2 et 3 ($p > .05$).

5.1.3 - Modalités collectives de reconnaissance dans la condition Humain

Une dernière analyse des réponses concerne la condition Humain. Nous nous intéressons ici aux modalités collectives de la reconnaissance, définies selon trois cas : la reconnaissance mutuelle (les deux participants s'identifient mutuellement comme des humains), la reconnaissance unilatérale (seul un des deux participants identifie l'autre comme un humain), et la non-reconnaissance (aucun des deux participants ne reconnaît l'autre comme un humain).

Tous groupes confondus, nous ne pouvons pas conclure à une différence significative entre la fréquence de reconnaissance mutuelle et la

fréquence de reconnaissance unilatérale (p > .05). En revanche, la fréquence de reconnaissance mutuelle est statistiquement supérieure à la fréquence de non-reconnaissance (d = .27 ; t(15) = 4.1167 ; p < .01). De la même façon, la fréquence de reconnaissance unilatérale est statistiquement supérieure à la fréquence de non-reconnaissance (d = .17 ; t(15) = 2.8131 ; p = .01) (voir Figure 5).

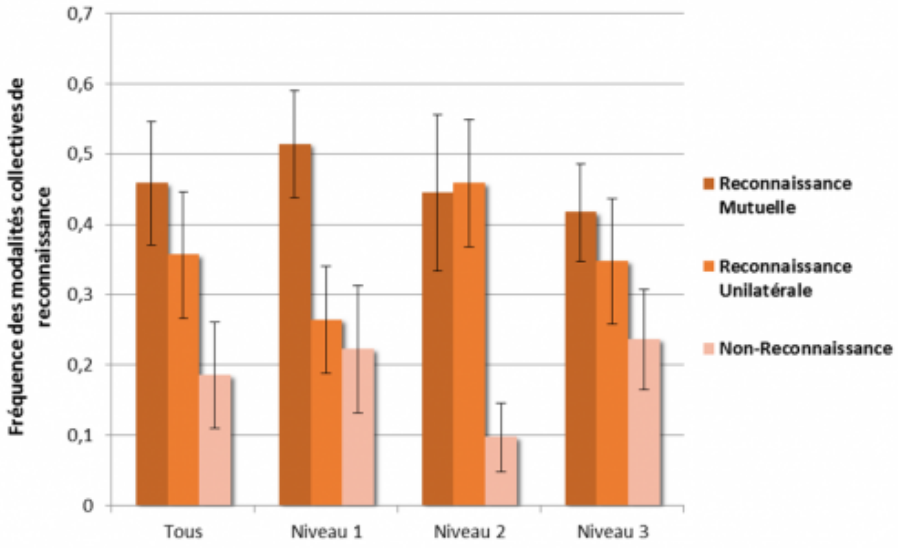


Figure 5

Fréquences des modalités collectives de reconnaissance (Reconnaissance mutuelle, Reconnaissance unilatérale et Non-reconnaissance) dans la condition Humain en fonction du niveau de complexité des robots

D'une façon générale, nous retrouvons les différences globales entre les modalités collectives de reconnaissance dans chaque groupe expérimental, excepté pour le groupe confronté aux robots du niveau 2, où la fréquence de reconnaissance mutuelle et la fréquence de reconnaissance unilatérale sont équivalents (d = .01). D'un point de vue statistique, nous ne pouvons pas conclure à un effet du niveau de complexité des robots pour aucune des modalités de reconnaissance, que ce soit pour la reconnaissance mutuelle, pour la reconnaissance unilatérale ou pour la non-reconnaissance (p > .05).

En ce qui concerne les modalités collectives de reconnaissance en fonction des deux blocs d'essais, la fréquence de reconnaissance mutuelle est statistiquement supérieure dans le second bloc par rapport au premier bloc d'essais (d = .28 ; t(15) = 4.7730 ; p < .01). De la même façon la fréquence de reconnaissance unilatérale est statistiquement inférieure dans le second bloc d'essais par rapport au premier bloc

d'essais (d = .20 ; t(15) = 2.7020 ; *p* = .01). En revanche, nous ne pouvons pas conclure à une évolution de la fréquence de non-reconnaissance entre les deux blocs d'essais (*p* > .05) (voir Figure 6).

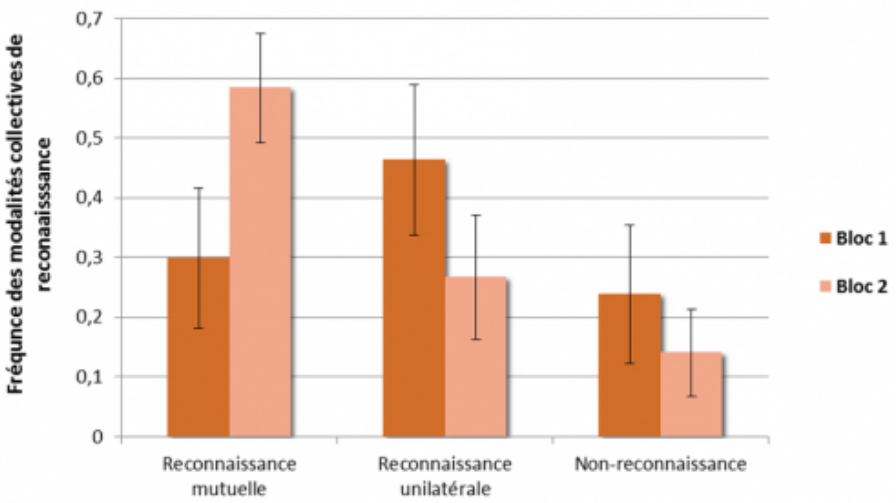


Figure 6

Fréquences des modalités collectives de reconnaissance (Reconnaissance mutuelle, Reconnaissance unilatérale et Non-reconnaissance) dans la condition Humain en fonction des deux blocs d'essais

Enfin, il est important de noter que les fréquences observées de reconnaissance mutuelle ne peuvent pas être expliquées par la propension individuelle des participants à répondre "Humain". En effet, la fréquence globale de reconnaissance mutuelle est de .45, alors que la probabilité conditionnelle que les deux participants répondent mutuellement "Humain" (comprise comme la probabilité qu'un participant donné réponde "Humain" quand son partenaire a également répondu "Humain") est de .25. Dans le second bloc, la fréquence de reconnaissance mutuelle est de .58, alors que la probabilité conditionnelle que les deux participants répondent mutuellement "Humain" est de .28.

5.2 - Analyse des trajectoires perceptives

Parallèlement à l'analyse des réponses, nous avons mené une analyse des trajectoires perceptives déployées par les participants au cours des essais. Ces analyses ont été rendues possibles par la méthodologie minimaliste utilisée, qui force le déploiement des activités perceptives des participants, ainsi que par la médiation technique, qui a permis l'enregistrement précis de ces trajectoires. Nous avons analysé une multitude d'indicateurs pour tenter de comprendre les réponses des

participants et caractériser les indicateurs de l'interaction qui prédisent le mieux la réponse « Humain ».

D'un point de vue méthodologique, nous sommes partis d'une analyse qualitative des trajectoires, à partir de leur représentation graphique (voir Figure 7).

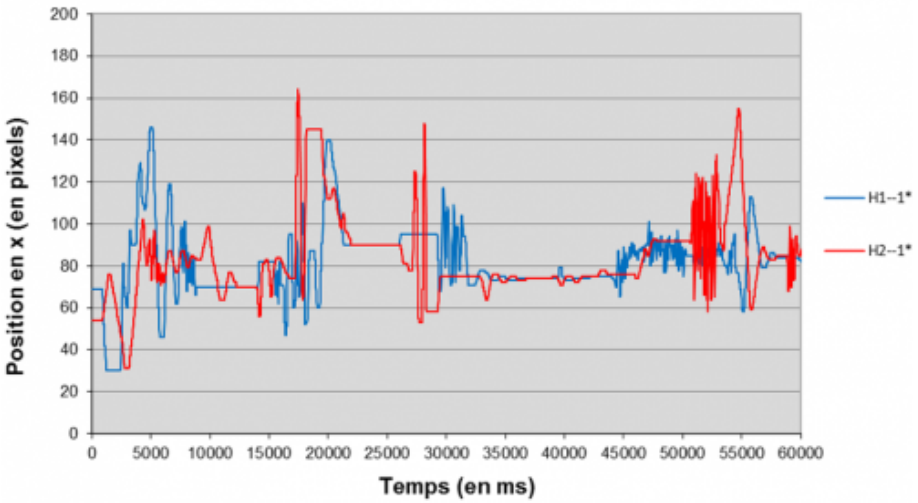


Figure 7

Exemple de trajectoires perceptives déployées par deux participants lors d'un essai de la condition Humain dans le groupe confronté aux robots de niveau 2 (Robots intermédiaires). Les positions en x (axe des ordonnées) des deux participants sont représentées en fonction du temps de l'essai (axe des abscisses). La courbe bleue représente la trajectoire du premier participant (« H1 ») et la courbe rouge représente la trajectoire du deuxième participant (« H2 »). À la fin de cet essai, les deux participants ont répondu « Humain ».

Cette première approche qualitative des trajectoires nous a permis de poser un certain nombre d'hypothèses, de façon à guider une analyse quantitative des indicateurs qui nous semblaient pertinents. Ainsi, nous avons repéré que l'interaction s'organisait souvent en séquences distinctes. Par souci de concision, nous ne présenterons ici que les indicateurs qui concernent ces séquences d'interaction, à la fois d'un point de vue microscopique (durée et variabilité de l'accroche mutuelle entre les partenaires d'interaction) et macroscopique (alternance d'engagements et de désengagements dans les séquences d'interaction au cours des essais). La distribution des données issues de ces indicateurs ne suivant pas une loi normale, nous avons effectué une analyse non-paramétrique (tests de Mann-Whitney), pour comparer les trajectoires associées à une réponse « Humain » avec les trajectoires associées à une réponse « Robot ».

5.2.1 - Durée de l'accroche mutuelle

Tout d'abord, nous nous sommes intéressés à la durée de l'accroche mutuelle entre les participants, soit la durée de l'ensemble des séquences d'interaction pendant lesquelles les écarts entre les stimulations sont inférieurs à 2 secondes. Notons que dans ce calcul, nous prenons en compte la durée des stimulations, mais pas la durée des arrêts mutuels (cas où les deux partenaires s'arrêtent l'un sur l'autre) lorsque ceux-ci sont supérieurs à 2 secondes.

Tous niveaux de complexité confondus, et quelle que soit la nature réelle du partenaire, la durée de l'accroche mutuelle est statistiquement inférieure dans les essais qui ont conduit à une réponse « Humain » par rapport aux essais qui ont conduit à une réponse « Robot » ($U = 1128$; $Z_{adj.} = -2.02$; $p = .04$) (voir Figure 8).

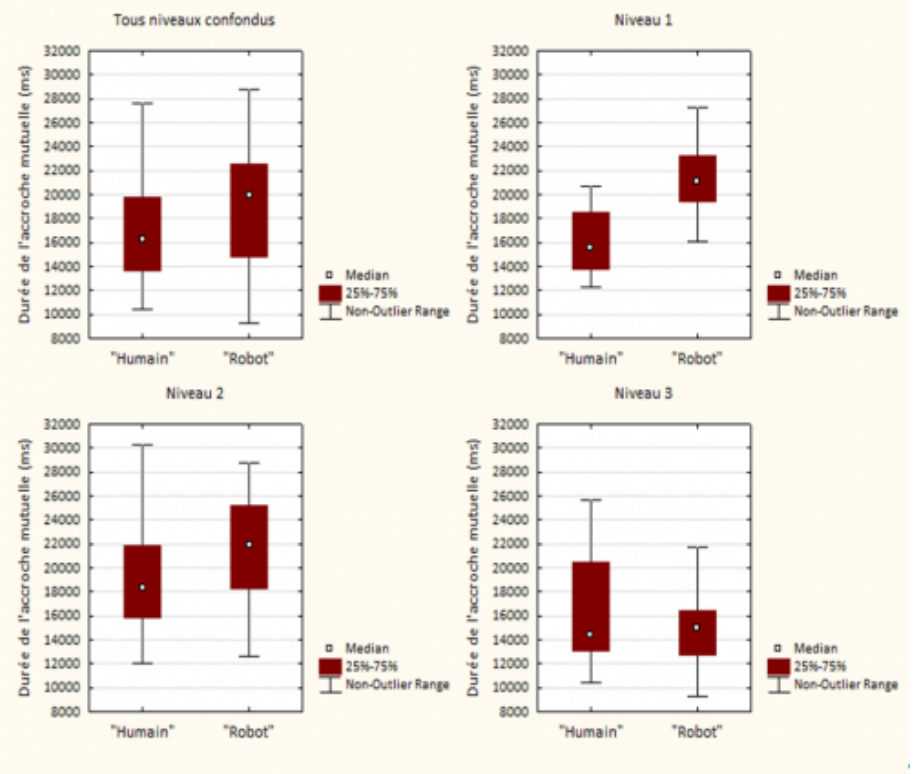


Figure 8

Durée de l'accroche mutuelle dans les essais qui conduisent à une réponse « Humain » et dans les essais qui conduisent à une réponse « Robot » en fonction des niveaux de complexité des robots

Mais l'analyse des niveaux de complexité révèle que cette différence n'est significative que pour le groupe confronté aux robots du niveau 1 (U

= 66 ; $Z_{adj.}$ = -3.02 ; p < .01). Nous ne pouvons pas conclure à une différence en ce qui concerne les niveaux 2 et 3.

5.2.2 - Variabilité de l'accroche mutuelle

De façon à caractériser les séquences d'accroche que nous avons identifiées, nous avons calculé la variabilité des écarts temporels entre les stimulations, en prenant comme base de données les séquences où ces écarts sont inférieurs à 2 secondes. Pour cela, nous avons calculé l'écart-type des écarts entre les stimulations dans les séquences d'accroche mutuelle pour chaque essai. Ceci nous a permis de caractériser la dispersion des écarts temporels entre les stimulations dans un continuum allant d'une accroche mutuelle modérée (écarts de 2 secondes) à une accroche mutuelle forte (écarts tendant vers zéro).

Tous niveaux de complexité confondus, la variabilité de l'accroche mutuelle est statistiquement plus importante dans les essais où les participants ont répondu « Humain » que dans les essais où ils ont répondu « Robot », quel que soit la nature réelle du partenaire (U = 582 ; $Z_{adj.}$ = 5.37 ; p < .01) (voir Figure 9).

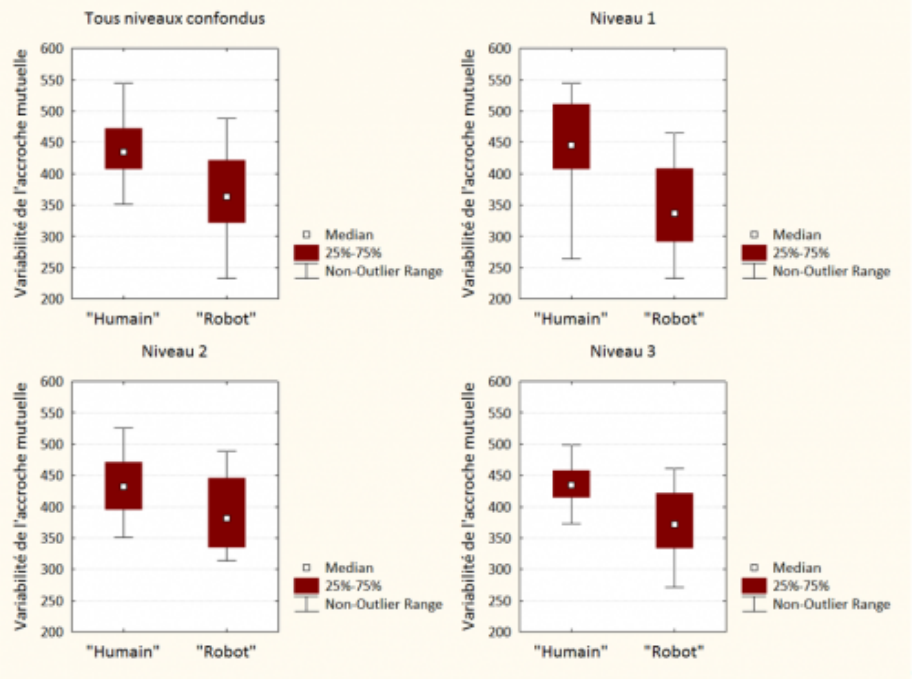


Figure 9

Variabilité de l'accroche mutuelle dans les essais qui conduisent à une réponse « Humain » et dans les essais qui conduisent à une réponse « Robot » en fonction des niveaux de complexité des robots.

Des analyses plus spécifiques révèlent que cette différence est statistiquement vérifiée pour le niveau 1 (U = 49 ; $Z_{adj.}$ = 3.55 ; p < .01), pour le niveau 2 (U = 94 ; $Z_{adj.}$ = 2.13 ; p = .03), et pour le niveau 3 (U = 58 ; $Z_{adj.}$ = 3.27 ; p < .01).

5.2.3 - Désengagements de l'accroche mutuelle

Enfin, nous avons analysé les interactions de façon plus macroscopique, de façon à rendre compte de l'organisation des séquences d'interaction lors des différents essais. En particulier, nous nous sommes intéressés au nombre de désengagements, qui signent les ruptures des séquences d'interaction. Nous avons considéré comme une rupture les cas où une séquence d'accroche mutuelle se termine, soit parce que l'écart entre les stimulations augmente au-delà de 2 secondes, soit parce qu'aucune autre stimulation ne suit cette séquence. Notons que nous ne considérons pas comme un désengagement les cas où une séquence d'interaction est en cours lorsque l'essai prend fin.

Tous niveaux de complexité confondus, et quelle que soit la nature réelle du partenaire, le nombre de désengagements est statistiquement plus élevé dans les essais où les participants ont répondu « Humain » que dans les essais où ils ont répondu « Robot » (U = 419 ; $Z_{adj.}$ = 6.38 ; p < .01) (voir Figure 10).

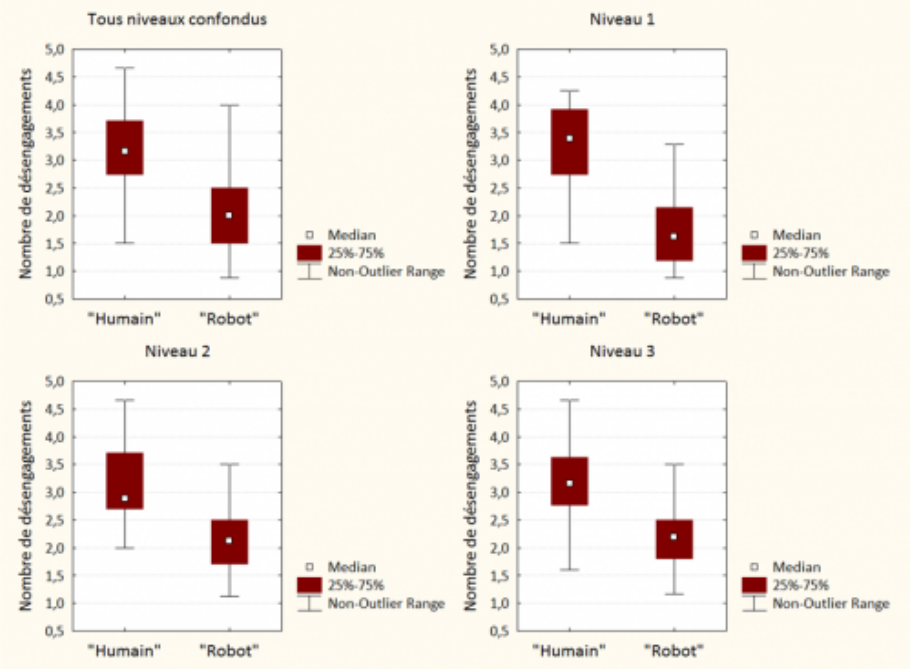


Figure 10

Nombre de désengagements de l'accroche mutuelle dans les essais qui conduisent à une réponse « Humain » et dans les essais qui conduisent à une réponse « Robot » en fonction des niveaux de complexité des robots.

Là encore, des analyses plus spécifiques révèlent que cette différence est statistiquement vérifiée pour le niveau 1 ($U = 36.5$; $Z_{adj.} = 3.95$; $p < .01$), pour le niveau 2 ($U = 50$; $Z_{adj.} = 3.52$; $p < .01$), et pour le niveau 3 ($U = 58.5$; $Z_{adj.} = 3.26$; $p < .01$).

6 - Discussion

Dans cette expérience, nous avons mobilisé les principes du paradigme du croisement perceptif minimaliste (Lenay, 2017 ; Lenay & Stewart, 2012), dans une tâche impliquant des interactions strictement dyadiques. Comme pour les études dont elle s'inspire (Auvray et al., 2009 ; Lenay et al., 2006 ; Lenay & Stewart, 2012), il est question d'étudier des interactions en temps réel (Auvray & Rohde, 2012), en mettant en jeu de façon cruciale l'*embodiment* et l'engagement des partenaires sociaux (Gallagher, 2001 ; Reddy & Morris, 2004). L'objectif de notre étude consistait à moduler la dynamique d'interaction via l'implémentation de comportements plus ou moins contingents dans des robots de complexité variable. Pour cela, nous avons mis en œuvre une situation expérimentale originale, selon le principe d'un test de Turing non-verbal (Turing, 1950). L'objectif général de cette étude était de caractériser le processus d'interaction et d'évaluer son impact sur la performance individuelle dans le contexte d'un tâche minimaliste de cognition sociale (Lenay, 2017 ; Lenay & Stewart, 2012).

Nos hypothèses étaient basées sur l'idée que les interactions sociales donnent lieu à l'émergence d'une dynamique d'interaction particulière, qui fournit un support pertinent aux participants pour discriminer les partenaires humains de robots, et ce, quel que soit le niveau de complexité de ces derniers. Néanmoins il était attendu que la complexité de certains robots pouvait ponctuellement induire les participants en erreur, dans la mesure où ces robots pouvaient participer à l'émergence d'une telle dynamique. En particulier, nous pensions que les robots conçus pour interagir de façon contingente avec les activités perceptives des participants entraîneraient plus d'erreurs d'identification, et ce, d'autant plus pour les niveaux de complexité élevés. En outre, la répétition des essais devait conduire à une stabilisation progressive de la

dynamique interpersonnelle, qui favoriserait la reconnaissance mutuelle des participants dans le deuxième bloc d'essais de l'expérience.

Les résultats tendent à montrer que l'implémentation de degrés de complexité dans les robots suffit à tromper les participants. En effet, la fréquence de réponses « Humain » données par erreur dans la condition Robot augmente à mesure que la complexité des robots augmente. Toutefois, l'effet de la complexité des robots ne semble pas pouvoir être réduite au seul fait d'interagir de façon contingente. Au contraire, nous observons que les robots élaborés en termes de programmes clos sont plus souvent identifiés comme des humains que les robots élaborés en termes de programmes d'interaction. Si nous détaillons ce résultat, nous observons que les programmes d'interaction du niveau de complexité le plus bas (robots simples) conduisent à très peu d'erreurs d'identification, contrairement aux programmes d'interaction des niveaux de complexité supérieurs (robots intermédiaires et robots complexes). De fait, les programmes d'interaction simples ne manifestent aucune variation rythmique. Quels que soient les mouvements du participant, ces robots « traquent » la position de leur curseur, et inversent leur trajectoire quand ils l'ont croisé, à un rythme constant. En conséquence, ils manifestent une contingence parfaite en oscillant périodiquement autour du participant. Le fait que ces programmes soient assez facilement identifiés comme des robots est cohérent avec certaines observations en psychologie du développement, qui mettent en évidence une préférence pour les contingences élevées mais imparfaites dans un contexte social, et cela, dès la première année (Gergely & Watson, 1996, 1999 ; Rochat & Striano, 1999). Pour autant, les résultats montrent également que dans les niveaux de complexité supérieurs, les robots conçus pour interagir de façon imparfaitement contingente ne conduisent pas à une fréquence de réponses « Humain » plus élevée que pour les robots non-contingents. Par ailleurs, la fréquence de réponse « Humain » obtenue dans la condition Humain reste significativement supérieure à la fréquence de réponses « Humain » obtenue dans la condition Robot, quelle que soit la complexité de ces derniers. Deux hypothèses peuvent être avancées pour expliquer ce résultat.

Tout d'abord, nous pourrions considérer que la variabilité implémentée dans les robots intermédiaires et complexes était trop en-deçà de la variabilité propre aux interactions entre des partenaires humains pour simuler de façon convaincante leur comportement. Dans cette optique, il serait intéressant de poursuivre cette expérience, en implémentant dans

des robots de complexité supérieure un degré de variabilité dans leurs oscillations de plus en plus élevé : on pourrait alors observer si les participants sont encore capables de distinguer les humains des robots, et jusqu'à quel point cette variabilité pourrait les tromper systématiquement, ou du moins les conduire à une indécision.

Une deuxième hypothèse consiste à tenir que la dynamique des interactions humaines présente une organisation spécifique, qui ne se réduit pas à de simples relations contingentes, fussent-elles imparfaites. L'analyse des trajectoires perceptives peut permettre de comprendre ce qui singularise cette organisation. De façon minimale, nous pourrions caractériser l'interaction humaine par l'émergence d'une accroche perceptive mutuelle. C'est en effet cette accroche qui semble caractériser la reconnaissance de l'autre dans les croisements perceptifs minimalistes, qui peut ainsi être décrite comme le produit d'un attracteur de la dynamique collective (Auvray et al., 2009 ; Lenay et al., 2006). Mais dans notre cas, c'est l'inverse qui semble s'être produit : la durée de l'accroche mutuelle est inférieure dans les essais qui ont conduit à une réponse « Humain » par rapport aux essais qui ont conduit à une réponse « Robot ». Nous avons toutefois noté que cet effet ne se vérifie que pour les participants confrontés aux robots du niveau 1 (robots simples). Ce résultat pourrait s'expliquer par une trop grande régularité des écarts temporels qui séparent les rencontres entre les participants et les robots simples, notamment lorsqu'il s'agit des programmes d'interaction simples. Pour ce qui est des niveaux de complexité supérieurs, l'impossibilité de prédire la réponse du participant par la durée de l'accroche mutuelle pourrait s'expliquer par l'effet combiné d'une proportion d'accroche élevée avec une variabilité temporelle de cette accroche. L'irrégularité de ces robots semble en effet avoir entraîné les sujets dans une dynamique d'interaction qui les a induits en erreur. Effectivement, la variabilité de l'accroche semble mieux prédire la réponse « Humain ». Quel que soit le niveau de complexité des robots, plus l'accroche mutuelle présente des variabilités temporelles, plus les contingences sociales peuvent être considérées comme imparfaites, et plus le participant répond « Humain ».

Parallèlement, la succession temporelle de séquences d'interaction semble également être un bon prédicteur de la réponse « Humain ». En effet, plus l'interaction a subi de ruptures au cours d'un essai, et plus les participants répondent « Humain ». La plupart du temps, les ruptures de l'interaction peuvent être interprétées par une incapacité individuelle à

participer au maintien de la dynamique d'interaction. L'un des deux partenaires a pu essayer d'entraîner son partenaire dans une direction donnée, pour voir si celui-ci était capable de le suivre dans l'espace, tandis que l'autre n'a pas perçu ce déplacement. En outre, il est également possible que l'un des deux partenaires ait perçu à tort une dérive de l'interaction, et l'ait anticipée dans l'espace, ce qui a conduit à une perte de contact. Quoi qu'il en soit, l'engagement mutuel des deux participants dans la recherche active de stimulations rend chaque fois possible une restauration de la dynamique, qui peut ensuite conduire à de nouveaux désengagements. C'est précisément cette capacité de réengagement qui nous semble cruciale. Les participants ont l'expérience d'une activité perceptive intentionnelle dirigée vers leur propre activité perceptive, qui s'accroche et se désengage périodiquement. Ils peuvent alors tenter de moduler cette expérience, en provoquant intentionnellement la rupture pour tester la capacité de réengagement de l'autre. Les deux participants entrent alors dans une stratégie active de vérification, qui leur permet de donner du sens à la dynamique d'interaction et de la comprendre non comme un engagement unilatéral (comme dans le cas des programmes d'interaction), mais comme le résultat d'un engagement mutuel intentionnel.

Au regard de ces résultats, nous pouvons envisager que la réponse « Humain » est déterminée par une organisation particulière de l'interaction, qui repose sur la combinaison de caractéristiques microscopiques (comme la variabilité de l'accroche mutuelle) et de caractéristiques macroscopiques (comme l'alternance d'engagements et de désengagements). Nous suggérons que cette organisation résulte de l'engagement mutuel des participants, qui influence par elle-même les stratégies déployées individuellement. Les participants semblent en effet entraînés dans une dynamique qui les oblige à des ajustements continus de leurs mouvements, pour ne pas rompre l'interaction, chaque participant étant individuellement engagé pour le maintien du contact interpersonnel. Ce faisant, ils sont activement impliqués dans la co-régulation de l'interaction pour garantir sa persistance dans le temps (De Jaegher & Di Paolo, 2007), en saisissant localement les opportunités d'action disponibles (Krueger, 2011 ; Lenay & Stewart, 2012).

La détection des contingences sociales reste essentielle pour la compréhension interpersonnelle (Crown et al., 2002 ; Stern, 1985). La rupture brutale de la contingence est détectée par le bébé lorsque la mère adopte brusquement un visage impassible (Tronick et al., 1979), ou

lorsque la dynamique d'interaction est perturbée (Murray & Trevarthen, 1985 ; Jacqueline Nadel et al., 1999). Mais si la rythmicité des comportements expressifs est cruciale, une stricte régularité perturbe également l'engagement mutuel. Les ajustements dynamiques du participant, résultant de la régulation active des mouvements de son partenaire, fait émerger une structure temporelle particulière qui semble caractériser le processus d'interaction (De Jaegher et al., 2010). Cette co-régulation dynamique nous semble rejouer l'altéroception décrite dans les interactions précoces (Braten, 1998 ; Trevarthen, 1993). La rythmicité de l'interaction sociale renvoie alors à la variabilité des coordinations temporelles qui définissent l'interaction dès les premiers mois de la vie (Bateson, 1979 ; Stern, 1985). Selon l'approche énactive (Stewart et al., 2010 ; Varela et al., 1991), le rythme en question se réfère à l'auto-organisation de l'interaction comme un processus autonome : quand il a émergé, il exerce sa propre influence sur les ajustements individuels et organise donc le processus d'interaction (De Jaegher & Di Paolo, 2007).

Nos résultats suggèrent que la qualité de l'interaction est cruciale pour la discrimination entre les humains et les robots. Les humains sont plus souvent identifiés comme des humains que comme des robots, quel que soit le niveau de complexité de ces derniers, ce qui suggère que les interactions perceptives humaines sont qualitativement différentes. Par ailleurs, nous pouvons conclure à une évolution différenciée des modalités collectives de reconnaissance durant la tâche, qui se traduit par une augmentation de la fréquence de reconnaissance mutuelle et par une diminution de la fréquence de reconnaissance unilatérale entre les deux blocs d'essais. Les interactions entre les participants semblent présenter une signature particulière, qui est de mieux en mieux identifiée par les participants à mesure de la répétition des essais. Ainsi, il semble que la dynamique d'interaction se stabilise progressivement, comme cela a été suggéré ailleurs (Di Paolo et al., 2008), et devient un indice suffisant pour la discrimination des humains et des robots. Cette stabilisation de la dynamique pourrait progressivement rendre plus saillant ses éléments caractéristiques, qui seraient alors plus facilement ressaisis par les participants en tant qu'indice pour la reconnaissance d'opportunités d'interaction caractéristiques d'une coordination interpersonnelle (Lenay, 2017 ; Lenay & Stewart, 2012).

7 - Conclusion

En accord avec les résultats issus du paradigme du croisement perceptif

minimaliste (Lenay & Stewart, 2012), notre étude suggère que c'est la dynamique collective de l'interaction qui est reconnue, plus que l'intentionnalité individuelle des individus impliqués dans cette interaction (Auvray et al., 2009 ; Lenay et al., 2006). En outre, si la détection de contingences caractéristiques peut être une des composantes de la compréhension interpersonnelle, il semble que ces contingences résultent d'un engagement mutuel, qui fait émerger une dynamique autonome pouvant influencer elle-même les conditions de son maintien. Ainsi, la cognition sociale semble reposer sur une capacité à participer activement à des coordinations interpersonnelles, plutôt que sur la seule capacité individuelle à agir de façon contingente. De façon à caractériser ces coordinations, et à expliquer les réponses des participants, nous avons mené une analyse des trajectoires perceptives déployées par les participants. Celle-ci indique que la réponse « Humain » peut être prédite à partir de caractéristiques objectives de la dynamique d'interaction, que ce soit à un niveau microscopique (variabilité de l'accroche perceptive) ou à un niveau macroscopique (alternance de désengagements et de réengagements dans les séquences d'interaction). Des analyses complémentaires pourront offrir de préciser ces paramètres pour une meilleure compréhension de la dynamique d'interaction.

Ces résultats encouragent à replacer le processus d'interaction au cœur des études de la cognition sociale, de façon à impliquer le fait même d'interagir dans les mécanismes de la compréhension sociale (De Jaegher & Di Paolo, 2007). Cette approche énactive de la cognition sociale suggère que le croisement perceptif, même lorsqu'il est réduit à son expression la plus simple, implique des facteurs individuels et des facteurs collectifs qui participent conjointement à l'émergence d'une dynamique relationnelle constitutive de la cognition sociale (De Jaegher et al., 2010 ; Di Paolo & De Jaegher, 2016 ; Lenay, 2017).

Remerciements

Les auteurs remercient la Région Hauts-de-France et le Fond Européen de Développement Régional (FEDER) pour le financement de cette recherche via le projet SPACEI (2014/2018).

8 - Références

Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*(1), 32-47.

https://doi.org/10.1016/j.newideapsych.2007.12.002

Auvray, M., & Rohde, M. (2012). Perceptual crossing : the simplest online paradigm. *Frontiers in Human Neuroscience*, *6*, 181. https://doi.org/10.3389/fnhum.2012.00181

Bateson, M. C. (1979). The epigenesis of conversational interaction : A personal account of research and development. In M. Bullowa (éd.), *Before Speech : The Beginning of Interpersonal Communication* (p. 63-77). Cambridge : Cambridge University Press.

Braten, S. (1998). Intersubjective Communication and Emotion in Early Ontogeny. Cambridge University Press.

Crown, C. L., Feldstein, S., Jasnow, M. D., Beebe, B., & Jaffe, J. (2002). The cross-modal coordination of interpersonal timing : six-week-olds infants' gaze with adults' vocal behavior. *Journal of Psycholinguistic Research*, *31*(1), 1-23.

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 Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition ? *Trends in Cognitive Sciences*, *14*(10), 441-447. https://doi.org/10.1016/j.tics.2010.06.009

Di Paolo, E. (2009). Extended Life. *Topoi*, *28*(1), 9–21.

Di Paolo, E., & De Jaegher, H. (2016). Neither individualistic, nor interactionist. In C. Durt, T. Fuchs, & C. Tewes (éd.), *Embodiement, enaction, and culture*. MIT Press.

Di Paolo, E., Rohde, M., & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction ? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, *26*(2), 278-294.

Fogel, A. (1993). Developing through relationships : origins of communication, self, and culture. Harvester Wheatsheaf.

Fogel, A., & Garvey, A. (2007). Alive communication. *Infant Behavior &*

Development, *30*(2), 251-257.
https://doi.org/10.1016/j.infbeh.2007.02.007

Frith, C. D.(2008). Social cognition. *Philosophical Transactions of the Royal Society B : Biological Sciences*, *363*(1499), 2033-2039.
https://doi.org/10.1098/rstb.2008.0005

Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction ? *Journal of Consciousness Studies*, *8*(5-7), 83-108.

Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (éd.), *The shared mind : Perspectives on intersubjectivity. Converging Evidence in Language and Communication Reserach* (John Benjamins Publishing Company, Vol. 12, p. 17–38).

Gallotti, M., & Frith, C. D.(2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160-165.
https://doi.org/10.1016/j.tics.2013.02.002

Gergely, G., & Watson, J. S.(1996). The social biofeedback theory of parental affect-mirroring : the development of emotional self-awareness and self-control in infancy. *The International Journal of Psycho-Analysis*, *77 (Pt 6)*, 1181-1212.

Gergely, G., & Watson, J. S.(1999). Early socio–emotional development : Contingency perception and the social–biofeedback model. In P. Rochat (éd.), *Early Social Cognition : Understanding Others in the First Months of Life* (p. 101–136). Hillsdale, NJ : Lawrence Erlbaum Associates.

Hobson, R. P.(2002). The Cradle of Thought : Exploring the Origins of Thinking. Oxford University Press.

Hutto, D. D.(2004). The Limits of Spectatorial Folk Psychology. *Mind & Language*, *19*(5), 548–573.
https://doi.org/10.1111/j.0268-1064.2004.00272.x

Krueger, J. (2011). Extended cognition and the space of social interaction. *Consciousness and Cognition*, *20*(3), 643-657.
https://doi.org/10.1016/j.concog.2010.09.022

Lenay, C. (2006). Enaction, externalisme et suppléance perceptive. *Intellectica*, *43*, 27–52.

Lenay, C. (2017). Schèmes interactionnistes de la cognition sociale. *Cahiers COSTECH*, numéro 1 (2017) - 06/2017, http://www.costech.utc.fr/CahiersCOSTECH/spip.php?article52.

Lenay, C., Auvray, M., Sebbah, F., & Stewart, J. (2006). Perception of an intentional subject : An enactive approach. In *Third International Conference on Enactive Interfaces* (p. 37-38). Montpellier, France.

Lenay, C., & Stewart, J. (2012). Minimalist approach to perceptual interactions. *Frontiers in Human Neuroscience*, *6*, art. 98. https://doi.org/10.3389/fnhum.2012.00098

McGann, M., & De Jaegher, H. (2009). Self–other contingencies : Enacting social perception. *Phenomenology and the Cognitive Sciences*, *8*(4), 417-437. https://doi.org/10.1007/s11097-009-9141-7

Michael, J., & Overgaard, S. (2012). Interaction and social cognition : A comment on Auvray et al.'s perceptual crossing paradigm. *New Ideas in Psychology*, *30*(3), 296-299. https://doi.org/10.1016/j.newideapsych.2012.02.001

Murray, L., & Trevarthen, C. (1985). Emotional Regulation of Interaction Between Two-Month-Olds and Their Mothers. In T. Field & N. Fox (éd.), *Social perception in infants* (p. 101-125). Norwood, NJ : Ablex.

Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communications. *Journal of Child Language*, *13*(01), 15-29. https://doi.org/10.1017/S0305000900000271

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, *2*(2), 164–173. https://doi.org/10.1111/1467-7687.00065

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, *24*(5), 939-973 ; discussion 973-1031.

Reddy, V. (2003). On being the object of attention : implications for self–other consciousness. *Trends in Cognitive Sciences*, *7*(9), 397-402. https://doi.org/10.1016/S1364-6613(03)00191-8

Reddy, V., & Morris, P. (2004). Participants Don't Need Theories Knowing Minds in Engagement. *Theory & Psychology*, *14*(5), 647-665. https://doi.org/10.1177/0959354304046177

Rochat, P., & Striano, T. (1999). Social-cognitive development in the first year. In P. Rochat (éd.), *Early Social Cognition : Understanding Others in the First Months of Life* (p. 3-34). Hillsdale, NJ : Lawrence Erlbaum Associates.

Stern, D. N.(1985). The interpersonal world of the infant : a view from psychoanalysis and developmental psychology. New York : Basic Books.

Stewart, J., Gapenne, O., & Di Paolo, E. (2010). *Enaction : Toward a New Paradigm for Cognitive Science*. Cambridge, MA : MIT Press.

Trevarthen, C. (1993). The function of emotions in early infant communication and development. In J. Nadel & L. Camaioni (éd.), *New Perspectives in Early Communicative Development* (p. 48-81). London : Routledge.

Trevarthen, C. (2001). Intrinsic motives for companionship in understanding : Their origin, development, and significance for infant mental health. *Infant Mental Health Journal*, *22*(1-2), 95–131. https://doi.org/10.1002/1097-0355(200101/04)22:1<95::AID-IMHJ4>3.0.CO ;2-6

Tronick, E. Z., Als, H., & Adamson, L. (1979). Structure of early face-to-face communicative interactions. In M. Bullowa (éd.), *Before Speech : The Beginning of Interpersonal Communication* (p. 349-370). Cambridge : Cambridge University Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind : Cognitive Science and Human Experience* (6th edition). Cambridge, MA : MIT Press.