Université Paris Nanterre
200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

Membre de l'université Paris Lumières

École doctorale 139 :
Connaissance, langage, modélisation
Laboratoire Éthologie Cognition Développement (LECD)

Alice Araguas

Interagir avec des robots sociaux : Une approche pluridisciplinaire chez l'enfant et l'oiseau

Thèse présentée et soutenue publiquement le 28 janvier 2022
en vue de l'obtention du doctorat de Ethologie et cognition comparées de l'Université Paris Nanterre
sous la direction de M. Sébastien Derégnaucourt (Université Paris Nanterre)
et de Mme Bahia Guellaï (Université Paris Nanterre)

Jury :

Directeur de thèse:	M. Sébastien Derégnaucourt	PR, Université Paris Nanterre
Directrice de thèse :	Mme Bahia Guellaï	MCF, Université Paris Nanterre
Rapporteuse :	Mme Marine Grandgeorge	MCF, Université Rennes 1
Rapporteur :	M. Alexandre Pitti	MCF, Université de Cergy-Pontoise
Membre du jury :	Mme Erika Parlato de Oliveira	PR, Universidade Federal de Minas Gerais
Membre du jury :	M. Pascal Mallet	PR, Université Paris Nanterre

Interagir avec des robots sociaux :

Une approche pluridisciplinaire

chez l'enfant et l'oiseau

Alice Araguas

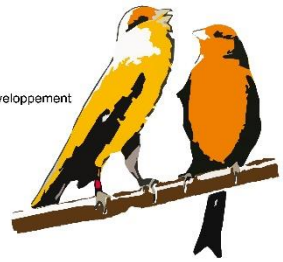
Sous la supervision de

Sébastien Derégnaucourt & Bahia Guellaï

Avant-propos

Ce projet de thèse a été rendu possible grâce à une bourse de l'Ecole Doctorale 139, Connaissance, langage, modélisation de l'Université Paris Nanterre.

Les études avec les enfants ont été menées au sein de différentes écoles dans le Sud-Ouest de la France et en Ile-de-France, ainsi que dans un Service d'Education Spéciale et de Soins A Domicile spécialisé dans l'accompagnement des enfants présentant un Trouble du Spectre Autistique. Les études avec les diamants mandarins ont été réalisées au sein du Laboratoire Ethologie Cognition Développement de l'Université Paris Nanterre.



Note à l'attention des lecteurs

Ce manuscrit est composé de plusieurs chapitres présentés sous forme d'articles.

CHAPITRE I

"Child-robot interactions in a learning context, an evaluation of body knowledge and a measure of prosociality: comparison with a human partner."

Alice Araguas, Arnaud Blanchard, Adrien Chopin, Sébastien Derégnaucourt, and Bahia Guellaï

En préparation.

CHAPITRE II

"Body awareness and prosocial behaviour in Autism Spectrum Disorder children interacting either with a human or a robot."

Alice Araguas, Arnaud Blanchard, Adrien Chopin, Sébastien Derégnaucourt, and Bahia Guellaï

En préparation.

CHAPITRE III

"Design of a robotic zebra finch for experimental studies on developmental song learning."

Alice Araguas, Bahia Guellaï, Philippe Gauthier, Florian Richer, Guglielmo Montone and Sébastien Derégnaucourt.

En révision à *Journal of Experimental Biology*.

CHAPITRE IV

"Of birds and robots: influence of behavioural contingency on developmental song learning in the Zebra Finch (*Taeniopygia guttata*)."

Alice Araguas, Bahia Guellaï, Adrien Chopin and Sébastien Derégnaucourt.

En préparation.

Remerciements

Au cours de mon cursus universitaire, j'ai eu la chance de rencontrer des personnes passionnées et inspirantes qui m'ont fait mûrir professionnellement, bien au-delà de ce que j'avais imaginé, je leur adresse toute ma reconnaissance.

Tout d'abord, je remercie sincèrement les membres du jury, Marine Grandgeorge, Alexandre Pitti, Pascal Mallet et Erika Parlato de Oliveira d'avoir accepté de lire et d'évaluer mon travail.

Je remercie vivement le directeur et la directrice de ma thèse, Sébastien Derégnaucourt et Bahia Guellaï, pour leur grande disponibilité, leur rigueur ainsi que leur ouverture d'esprit. Je suis heureuse d'avoir pu travailler avec eux sur un sujet aussi novateur et stimulant.

Je remercie Philippe Gauthier et Florian Richer de l'Institut des Systèmes Intelligents et de Robotique de Sorbonne Université pour leur travail sur le robot-diamant mandarin et leur disponibilité pour la mise en place des expériences, les réparations et ajustements du robot et également, Sofiane Boucenna et Arnaud Blanchard du Laboratoire Equipe Traitement de l'Information et Systèmes (ETIS) de l'Université Cergy-Pontoise pour leurs conseils au début de mon projet avec le robot humanoïde NAO. Merci beaucoup à Arnaud Blanchard pour nos échanges sur la robotique et les journées bien plus longues que prévu à programmer NAO.

Merci beaucoup à Philippe Groué, Emmanuelle Martin, Ophélie Bouillet, Marie Huet, Priscilla Roussel, Anaëlle Brunet et Margaux Laurent pour m'avoir accompagnée pendant mes expériences au laboratoire, pour m'avoir appris à m'occuper des oiseaux et avoir veillé à leur santé. Un merci tout particulier à Philippe pour son sens de l'humour et ses compétences en construction de dispositifs expérimentaux, à Emmanuelle pour sa bonne humeur et sa patience infinie et à Marie pour son ingéniosité et sa détermination à prendre soin de tous les animaux qui croisent sa route.

Je remercie tous les enfants qui ont participé aux études ainsi que leurs parents, les équipes éducatives des écoles et du SESSAD. Merci aux directrices et directeur des écoles d'avoir permis aux expériences d'avoir lieu malgré la complexité de la situation sanitaire : Marie-Christine Charlassier, Cédric Doisneau, Cathy Araguas, Claire Fitzgerald, Kühl Raynaud et Karine Pinchelimouroux ainsi que toute l'équipe du SESSAD et, en particulier, Fanny Lesage, pour sa persévérance à toute épreuve et son expertise dans l'accompagnement en psychomotricité des enfants avec un Trouble du Spectre Autistique.

Les étudiants Adeline Depierreux, Camille Nozières, Antoine Thomas, Khedidja Athamnia, Margaux Capezzera, Lorine Grosos, Vincent Basse et Delphine Doléans ont toute ma gratitude pour leur investissement et leur minutie dans l'analyse des vidéos de diamants mandarins sans oublier Oumaima Doumeksa, Pauline Bernard et Aïcha Bekkaoui pour les analyses chez les enfants. Un grand merci à Louisane Araguas pour son adaptabilité et sa polyvalence, de la peinture des robots-diamants mandarins à la passation des expériences en école.

Je remercie également Adrien Chopin de l'Institut de la Vision de Sorbonne Université pour tout le temps passé à décortiquer mes données et à réaliser des analyses statistiques que je n'aurais pu faire seule.

Merci beaucoup à Sofiane Boucenna du laboratoire ETIS de l'Université Cergy-Pontoise, à Rana Esseily du Laboratoire Ethologie Cognition Développement de l'Université Paris Nanterre et à Isabelle George du laboratoire EthoS de l'Université Rennes 1, pour leurs précieuses réflexions lors des comités de suivi de thèse.

Un merci chaleureux à toute l'équipe du Laboratoire Ethologie Cognition Développement de l'Université Paris Nanterre pour leur accueil depuis mon stage de master 2, leur enthousiasme et les discussions passionnantes sur la recherche en éthologie et en psychologie. Je remercie Nathalie Béguin, Lauriane Rat-Fischer, Frédérique Hallé, qui ont pris

le temps d’accompagner mes premiers dans l’enseignement, ainsi que Rana Esseily et Pascal Mallet pour les riches partages d’expériences en pédagogie.

Je souhaite à mes collègues doctorant.e.s et ex-doctorant.e.s, ami.e.s brillant.e.s et déterminé.e.s, de s'épanouir dans la suite de leur parcours. Un immense merci à Lucille Le Maguer avec qui j’ai tant ri et appris.

Je remercie également Stéphane Chameron, Heiko Rödel ainsi que toute l’équipe du Laboratoire d’Ethologie Expérimentale et Comparée (LEEC) de l’Université Sorbonne Paris Nord pour m’avoir permis de terminer mon master 2 et pour leurs enseignements captivants en éthologie. Je suis particulièrement reconnaissante envers Cristina Lorenzi, qui a encadré mon stage de master 1 au LEEC avec beaucoup de bienveillance, m’a formée et transmis son goût pour les études comportementales rigoureuses.

Je n’aurais pas entrepris ce long voyage sans Denis Grabot, directeur de l’Institut de Formation en Psychomotricité de Bordeaux, qui m’a toujours encouragée dans cette voie. Ce sont les séances avec les personnes transidentitaires et le mémoire que j’ai écrit sous sa supervision qui ont fait naître mon envie de devenir chercheuse.

Enfin, je remercie ma famille et mes ami.e.s pour leur soutien inestimable et indéfectible.

Résumé

Cette thèse pluridisciplinaire avait pour but de comparer les interactions agent biologique-robot à celles entre deux agents biologiques de la même espèce chez de jeunes individus de deux espèces hautement sociales : l'humain et une espèce d'oiseau, le Diamant Mandarin (*Taeniopygia guttata*). Trois populations ont été testées : des enfants au développement typique (DT) âgés de 3 à 6 ans, des enfants présentant un Trouble du Spectre Autistique (TSA) léger âgés de 4 à 8 ans et des diamants mandarins âgés de 35 à 100 jours. Chez les enfants, nous avons comparé l'apprentissage de vocabulaire (uniquement pour le groupe DT), la connaissance des parties du corps et des émotions ainsi que la tendance à aider selon que la démonstratrice était une adulte non familière ou le robot humanoïde NAO féminisé (voix humaine) et renommé NAOMIE. Nos résultats montrent que les scores des enfants (DT et TSA) étaient comparables quelle que soit la démonstratrice, suggérant que le robot NAO peut être utilisé en situation d'interaction sociale auprès de jeunes enfants. Pour réaliser nos expériences avec les diamants mandarins, nous avons développé un robot-diamant mandarin interactif : MANDABOT. Nous avons comparé l'apprentissage du chant selon que les jeunes avaient été élevés en présence d'un mâle adulte non familier ou de MANDABOT. Nous avons trouvé que l'imitation des chants produits par les tuteurs était aussi élevée dans les deux groupes suggérant que MANDABOT peut être utilisé comme tuteur de chant. Nos trois études nous ont permis de valider des protocoles en utilisant une approche expérimentale innovante et prometteuse en psychologie du développement, en psychomotricité et en éthologie : la robotique sociale.

Mots-clefs : interactions enfant-robot, interactions oiseau-robot, apprentissage vocal, conscience corporelle, émotions, comportement pro-social

Abstract

In this thesis, we developed a pluridisciplinary approach to compare interactions between a biological agent and arobot to interactions between two conspecific individualsin two highly social species: the human and the Zebra Finch (*Taeniopygia guttata*). We conducted three studies: with 3-to-6 years old typically developing (TD) children, 4-to-8 years old children with mild Austim Spectrum Disorder (ASD) and zebra finches from 35 to 100 days post hatch. In children, we evaluated vocabulary learning (for TD children only), knowledge on body parts and emotions and the tendency to help a demonstrator (an unfamiliar adult or the humanoid NAO robot feminized (human voice) as NAOMIE. We found similar results regardless the demonstrator type, in both TD children and ASD children which suggest that the NAO robot can be successfully used to study social behaviours of young children. For our experiments with zebra finches, we developed MANDABOT, an interactive robot finch. We compared song imitation in two groups of young zebra finches, whereas they were exposed to MANDABOT or to an unfamiliar adult male during their sensitive period for song learning. We found that both groups of zebra finches produced high imitations of songs models which suggest that MANDABOT can be an efficient tutor for song learning. Our results allowed us to validate experimental protocols using an innovative and promising tool in developemental psychology, psychomotricity and ethology: social robotics.

Key words: child-robot interactions, bird-robot interactions, vocal learning, body awareness, emotions, prosocial behaviour

Table des matières

Avant-propos 5

Note à l'attention des lecteurs 7

Remerciements 9

Résumé 13

Abstract 15

INTRODUCTION GENERALE 25

1. Du robot serf au robot compagnon : évolution des robots dans notre société 28

2. Les robots sociaux 33

2.1. Approche intra-spécifique des robots sociaux 33

2.1.1. L'apparence des robots sociaux 35

2.1.2. Habiletés sociales 41

2.1.2.1. Communication 41

2.1.2.2. Emotions 42

2.1.2.3. Personnalité 50

2.1.3. Autonomie et intelligence 52

2.1.4. Proximité physique 54

2.1.5. Profil temporel de l'exposition au robot 54

2.1.5.1. Le court terme 55

2.1.5.2. Le moyen-terme 55

2.1.5.3. Le long-terme 56
2.1.6 Rôle relationnel 56
2.1.7 Domaines d'application de l'utilisation des robots sociaux 57
2.2. Approche inter-spécifique des robots sociaux 58
2.3. Apprentissage à partir des robots sociaux 59
2.3.1 Apprentissage social 59
2.3.2 Utilisation des robots sociaux pour l'apprentissage 61
2.3.2.1. Chez les humains 61
2.3.2.2. Chez les animaux non humains 64
3. Choix des modèles 66
3.1. Modèles biologiques 66
3.1.1 Les enfants 66
3.1.1.1. Compétences des enfants dans les domaines étudiés dans nos expériences 68
3.1.1.2. Le langage oral 68
3.1.1.3. La connaissance des parties du corps 69
3.1.1.4. L'identification et l'expression des émotions 71
3.1.1.5. Les comportements pro-sociaux 72
3.1.2 Les diamants mandarins (Taeniopygia guttata) 74
4. Modèles robotiques 79
4.1. Le robot humanoïde NAO 79
4.2. Le robot diamant-mandarin « MANDABOT » 81
5. Objectifs de la thèse et hypothèses de travail 82

CHAPITRE I 85

1. Introduction 89

2. Materials and methods 91

2.1. Participants 91

2.2. Experimental procedure 92

2.2.1. General setup 92

2.2.2. Human group: exposure to an adult demonstrator 93

2.2.3. Robot group: exposure to the NAO robot demonstrator 93

2.2.4. Experimental tasks and scores 96

2.2.4.1. Questionnaires about child's representations of the demonstrator 96

2.2.4.2. Warm-up phase 97

2.2.4.3. Vocabulary task 97

2.2.4.4. Body task 99

2.2.4.5. Spontaneous helping situation 101

2.2.5. Adaptations of the experimental protocol 101

2.3. Behavioural analysis 102

2.3.1. Video recordings and analysis 102

2.3.2. Ethogram 102

2.4. Statistical analysis 102

2.4.1. Imitation of body parts sequences 104

2.4.2.Gaze time task by task 104

2.4.3. Pre and post-test questionnaires 104

3. Results 105

3.1. Experimental tasks and scores 105

3.1.1. Questionnaire about child's representations of the demonstrator 105

3.1.1.1. Mean number (average between pre and post-tests) of correct answers in the questionnaire 105

3.1.1.2. Number of "yes" answers in the questionnaire in pre-test 108

3.1.2. Vocabulary task 110

3.1.3. Body task 112

3.1.3.1. Comprehension of body parts labels 112

3.1.3.2. Identification of emotions 116

3.1.4. Spontaneous helping situation 120

3.2. Gaze times and other behaviours 121

3.2.1. Time spent looking at the demonstrator during the whole session 121

3.2.2. Time spent looking at the demonstrator during the key poses sub task 122

3.2.3. Direction of gaze when the demonstrator questioned the child 123

3.2.4. Number of times the child touched the demonstrator 123

4. Discussion 124

CHAPITRE II 147

1. Introduction 151

2. Materials and methods 155

2.1. Participants 155

2.2. Experimental procedure 155

2.2.1 General setup 155

2.2.2. Human condition: exposure to an adult demonstrator 157

2.2.3. Robot condition: exposure to the NAO robot demonstrator 157

2.2.4. Experimental tasks and scores 160

2.2.4.1. Warm-up phase 160

2.2.4.2. Body parts task 161

2.2.4.3. Emotions task 162

2.2.4.4. Spontaneous helping situation 163

2.2.5. Ethogram 163

2.3. Statistical analysis 163

3. Results 164

3.1. Scores 164

3.1.1. Body parts task 164

3.1.1.1. Comprehension of body parts labels 164

3.1.1.2. Imitation of body parts' sequences 165

3.1.2. Emotions task 165

3.1.2.1. Identification on images 165

3.1.2.2. Expression of emotions by the child 166

3.1.2.3. Identification of emotional key poses on the demonstrator 166

3.1.3 Spontaneous helping situation 168

3.2. Eye gaze 168

3.3. Other behaviours 169

3.3.1 Direction of gaze when the demonstrator questioned the child 169

3.3.2 Number of tactile contacts 170

4. Discussion 170

CHAPITRE III 187

1. Introduction 191

2. Materials and methods 193

2.1. Design of the robot 193

2.2. Interactive vocal loops between the bird and the robot 195

3. Results and discussion 196

CHAPITRE IV 201

1. Introduction 205

2. Materials and methods 206

2.1. Subjects 206

2.1.1. Experimental procedure 207

2.1.2. Live group: exposure to a live tutor 208

2.2. Exposure to a robotic finch, the MANDABOT 208

2.2.1. Contingent group 209

2.2.2. Non-contingent group 210

2.3. Behavioural analysis 211

2.3.1.Sound recording and analysis 211

2.3.1.1. Song segmentation 211

2.3.1.2. Similarity to the song model: analysis based on the motif 212

2.3.1.3. Preparation of acoustic stimuli for the robot 214
2.3.2.Video recording and analysis 214
2.3.2.1. Video analysis for the live group 214
2.3.2.2. Video analysis for the contingent group 215
2.4. Statistical analysis 218
2.4.1.Acoustic parameters 218
2.4.2.Influences of song activity on song imitation 218
2.4.3.Evolution of behaviours over the experiment 219
2.4.3.1.Live group 219
2.4.3.2. Contingent group 219
2.4.4.Effect of behaviours on song similarity 221
2.4.4.1. Live group 221
2.4.4.2. Contingent group 221
3. Results 222
3.1. Acoustic parameters 222
3.1.1.Song similarity between tutors and pupils 222
3.1.2.Comparison of acoustic parameters in the two groups 224
3.2. Influences of song activity on song learning 226
3.3.Evolution of behaviours over the experiment 227
3.3.1. Live group 227
3.3.2. Contingent group 227
3.3.2.1. Times spent on the platform, in close vicinity and in clumping 227

3.3.2.2. Number of pecking on the robot 229

3.3.2.3. Reactions to the robot songs 230

3.3.2.4. Evolution of number of calls over the experiment 232

3.4.Effects of behaviours on song similarity 232

3.4.1.Live group 232

3.4.2.Contingent group 233

4.Discussion 236

DISCUSSION GENERALE 247

1. Résumé des résultats principaux, interprétations et limites 249

1.1. Interactions enfant-robot dans un contexte d'apprentissage, de contrôle des connaissances sur le corps et de mesure de la pro-socialité: comparaison avec un partenaire humain 250

1.2. Conscience corporelle et comportement pro-social chez des enfants TSA, interagissant avec un humain ou un robot 255

1.3. Utilisation d'un robot comme tuteur de chant pour de jeunes mâles de Diamant Mandarin 258

2. Perspectives pour de futures études 262

2.1. Chez les enfants DT et présentant un TSA 262

2.2. Chez le Diamant Mandarin 265

REFERENCES 271

INTRODUCTION GENERALE

En décembre 2019, un café unique en son genre a ouvert ses portes à Tokyo. Au Pepper Parlor[1], les clients sont accueillis et servis par des robots Pepper, divertis par les chorégraphies des robots NAO pendant que le robot Whiz se charge de passer l'aspirateur (de gauche à droite sur les photographies ci-dessous).





Cet exemple illustre bien l'essor qu'a connu la robotique ces dernières décennies : les robots se sont perfectionnés et ont progressivement intégré le quotidien des humains.

Afin de comprendre ce qui a rendu possible un tel développement, nous résumerons dans une première partie de cette introduction, l'évolution des caractéristiques des robots depuis le milieu du XXème siècle. Nous nous focaliserons ensuite sur les robots sociaux, conçus spécialement pour interagir avec des agents biologiques. La recherche en robotique connait un développement exponentiel et les applications de ces robots dans nos sociétés ne cessent de se multiplier. Pour que le robot du XXIème siècle soit accepté comme un partenaire d'interaction, il est nécessaire de déterminer les caractéristiques qu'il doit posséder pour remplir cette fonction, contexte dans lequel s'inscrit cette thèse.

Nous évoquerons ainsi, dans un premier temps, des éléments saillants des recherches sur l'interaction agent biologique-robot. Nous présenterons ensuite les modèles biologiques de nos études en psychologie du développement, en psychomotricité et en éthologie : des enfants au développement typique et atypique ainsi que des jeunes d'une espèce d'oiseau, le Diamant Mandarin (*Taeniopygia guttata*). Nous poursuivrons brièvement avec les modèles robotiques. Nous conclurons cette introduction par les objectifs de ce travail de thèse

[1] https://pepperparlor.com/

pluridisciplinaire, à savoir la comparaison de l'interaction sociale et de l'apprentissage d'un individu selon que son partenaire est un adulte non familier ou un robot social.

Ce manuscrit comporte ensuite deux articles consacrés aux expériences menées chez les enfants au développement typique et chez les enfants présentant un Trouble du Spectre Autistique. Les deux chapitres suivants correspondent à un article méthodologique présentant le robot-diamant mandarin MANDABOT puis à l'article étudiant son utilisation en situation d'apprentissage du chant. Enfin, dans le dernier chapitre, nous résumerons les principaux résultats et présenterons les perspectives pour de futures expériences.

1. Du robot serf au robot compagnon : évolution des robots dans notre société

Le terme de robot vient du tchèque « *robota* » qui signifie « travail pénible, corvée » et vient de la même famille que le mot « serf ». Il a été inventé par le dramaturge tchèque Karel Čapek en 1920 pour décrire une « *machine à l'aspect humain, capable de parler et de se mouvoir* » (dans « R.U.R. Rossum's universal robots », 1920). On attribue par la suite la popularisation de la robotique à Isaac Asimov qui publia en 1942, une nouvelle de science-fiction dans lesquelles il décrivit trois lois gouvernant les comportements des robots : (1) un robot ne doit pas porter atteinte à un humain, ni le laisser dans une situation dangereuse sans intervenir ; 2) Un robot doit obéir aux ordres donnés par les humains sauf si cela va à l'encontre de la première loi ; 3) un robot doit protéger son existence dans la mesure où cette protection n'entre pas en contradiction avec les première et seconde lois. Le développement technologique de la seconde moitié du XXème siècle a progressivement conduit au passage des robots du domaine de la science-fiction aux chaînes de production des usines. A l'époque, les robots permettaient de réaliser des tâches mécaniques afin de minimiser les erreurs humaines et les risques de blessure tout en augmentant la productivité (Hockstein et al., 2007).

En 2012, la Fédération Internationale de la Robotique (IRF en anglais), un groupe industriel spécialisé dans les robots commerciaux, définit les robots industriels comme des « *manipulateurs contrôlés automatiquement, reprogrammables et polyvalents. Ils peuvent être programmés sur 3 axes ou plus qui peuvent être fixés sur place ou mobiles et sont utilisés pour des applications d'automatisation industrielle* ». L'automatisation, concept clef de la robotique, correspond à l'utilisation de systèmes et d'équipements automatiques contrôlés notamment par ordinateur et qui remplacent certaines tâches effectuées par des humains (Raj & Seamens, 2019).

Ces dernières années, les domaines d'application des robots n'ont cessé de se diversifier : des objets du quotidien à l'exploration spatiale et des fonds marins, de l'agriculture aux missions de secourisme, dans le domaine militaire, en médecine ou encore dans l'éducation… Les robots se développent avec une telle variété de formes (voir Baraka et al., 2020 pour une classification) et de fonctions que les définir de manière consensuelle constitue un réel défi.

Nous retiendrons qu'un robot : 1) a la capacité d'exécuter un programme lui permettant de réaliser des tâches spécifiques pour lesquelles il a été conçu (Leenes et al., 2017) et 2) présente un certain niveau d'autonomie. L'autonomie est la capacité à décider d'un comportement et de l'exprimer (Gunderson & Gunderson, 2004). Le degré d'autonomie des robots est très variable allant de la télé-opération à distance par l'humain (aussi appelée Wizard-of-Oz, Riek, 2012), comme le robot chirurgical daVinci (Francis et al., 2018) à une autonomie totale comme le robot Spot qui peut explorer des environnements extrêmes (Bouman et al., 2020). L'autonomie est un enjeu majeur de la robotique et la dernière décennie a été marquée par d'importantes avancées en la matière. De plus en plus de tâches spécialisées peuvent être exécutées par des robots de manière autonome : voler dans des environnements naturels

(Floreano & Wood, 2015), nager (Chen et al., 2014), ramasser des objets au sol (Kappassov et al., 2015) ou encore conduire un véhicule (Zhao et al., 2018).

En parallèle des améliorations concernant le degré d'autonomie, le champ de la robotique a progressivement convergé vers celui de l'Intelligence Artificielle (IA) pour donner naissance à des robots « intelligents » (Westerlund, 2020). De plus en plus de robots intègrent ainsi une IA ce qui augmente leurs capacités et la complexité des tâches qu'ils peuvent accomplir.

En Grande Bretagne, entre 1951 et 1952, apparaissent les premiers programmes informatiques intelligents, regroupés sous le terme d'intelligence de machine (Copeland, 2000). Le terme d'Intelligence Artificielle est créé quelques années plus tard, par le mathématicien John McCarthy qui l'utilise, pour la première fois, en 1956 dans le cadre d'un programme de recherche organisé au Darmouth College, aux Etats-Unis. Il établit une nouvelle science dont l'ambition est de doter les machines de systèmes informatiques leur permettant d'utiliser le langage, l'abstraction et les concepts et de résoudre des problèmes habituellement réservés aux humains (McCarthy et al., 2006). La même année, Allan Newell, Hebert Simon et Cliff Shaw, développent un programme appelé le théoricien logique (the Logic Theorist) capable de résoudre automatiquement une partie des théorèmes du traité de mathématiques *Principia Mathematica* d'Alfred Whitehead et Bertrand Russell (1913, cité dans Copeland, 2000). C'est la première fois qu'une machine est capable de raisonnement et il est considéré, à ce titre, comme le premier programme d'IA.

A la même époque, le mathématicien britannique Alan Turing développe un test emblématique auquel il donnera son nom et visant à déterminer si les machines peuvent penser (Turing, 1950). Ce test est mené à l'aveugle : un humain converse à l'écrit sans le savoir soit avec un autre humain soit avec une machine. S'il n'arrive pas à déterminer lequel de ses

interlocuteurs est une machine, on considère que cette dernière a réussi le test. Bien que critiqué, ce test a marqué un tournant dans le domaine de l'évaluation de l'Intelligence Artificielle.

Actuellement, les IA sont considérées comme des agents interactifs, autonomes, capables d'apprendre par eux-mêmes et pouvant accomplir des tâches nécessitant traditionnellement l'intelligence et l'intervention humaines (Floridi, 2019). « Apprendre » fait ici référence à la notion d'apprentissage automatisé ou « machine learning » : la machine peut améliorer les tâches programmées, routinières et automatisées pour lesquelles elle a été fabriquée mais, « *malgré ce que le terme d'apprentissage évoque, elle ne peut ni acquérir des connaissances, ni une forme de clairvoyance, ni faire preuve d'agentivité comme le fait un humain* » (Broussard, 2018). On parle également d'IA restreinte pour souligner qu'il s'agit seulement d'un logiciel informatique utilisant des algorithmes hautement sophistiqués lui permettant de trouver des motifs de données sur lesquels il base ses prédictions.

Richards & Smart (2013) proposent une définition légèrement différente du robot « intelligent » qu'ils décrivent comme un « *système construit faisant preuve d'agentivité à la fois physique et mentale mais qui n'est pas vivant au sens biologique du terme* ». Ils rappellent ainsi qu'une des particularités des robots est d'avoir un corps physique et ils leur attribuent une capacité fondamentale, l'agentivité. Dans ce champ, le terme d'agent définit une entité dont les actions sont contrôlées par un programme informatique autonome à la différence d'un avatar dont les actions sont contrôlées par un humain en temps réel. (Nowak & Biocca, 2003).

De plus, les robots sont capables de percevoir leur environnement de plus en plus finement à la fois grâce à l'amélioration des capteurs sensoriels utilisés, à l'intégration des IA et de la « vision par ordinateur », c'est-à-dire la capacité à interpréter des données visuelles pour en extraire des informations (Andreu-Perez et al., 2017). Dans le futur, les robots pourront, par leurs interactions avec leur environnement et les agents qui s'y trouvent, se construire une représentation sensori-motrice d'eux-mêmes comme des agents différenciés (Chatila et al.,

2018). Cette capacité à se distinguer d'autrui, c'est-à-dire à intégrer certains évènements comme des expériences internes et d'autres comme des expériences appartenant au monde externe (Birch et al., 2020), est le niveau le plus élémentaire de la conscience de soi et est applicable à la fois aux humains, aux animaux non-humains et aux robots. Si, à l'heure actuelle, la majorité des chercheurs en neurosciences s'accordent à dire que les robots ne sont pas conscients d'eux-mêmes (Hildt, 2019), ils soulignent que dans les années à venir, les robots pourraient le devenir (Manzotti & Chella, 2018). Cette conscience artificielle de soi-même serait alors cruciale pour que le robot interagisse de manière fluide et adaptée avec les autres entités de son environnement.

Ces considérations nous conduisent au champ de la robotique sociale qui recouvre d'une part, la robotique cognitive qui s'intéresse à la manière dont les robots apprennent et comprennent leur environnement de manière autonome et, d'autre part, l'interaction humain-robot (HRI en anglais) qui étudie les interactions et la communication entre les humains et les robots (Fong et al., 2003). Si cette discipline est née de la rencontre de la psychologie et de la robotique, les outils et concepts propres à l'éthologie (l'étude scientifique du comportement) sont également pertinents pour étudier les interactions humain-robot (Grandgeorge, 2020) comme nous le présenterons dans nos études (Chapitres I à IV).

Enfin, pour Miriyev & Kovač (2020), les robots actuels sont encore limités dans leurs capacités et leurs interactions du fait d'un déséquilibre entre, d'une part, le niveau de performance de leurs IA et leurs corps, leurs morphologies et les matériaux utilisés qui sont, en comparaison, moins développés. Ces auteurs proposent d'adopter une démarche holistique et pluridisciplinaire, au carrefour de l'ingénierie mécanique, la science des matériaux, la chimie, l'informatique et la biologie. Les comportements intelligents des robots sont ainsi appréhendés comme l'expression de l'interaction entre la structure de leur « cerveau », la morphologie de leur corps et ce, dans un environnement donné. C'est ce qu'ils nomment le champ de

l'Intelligence Artificielle Physique (IAP). Les robots faisant preuve d'IAP seront ainsi capables d'agir mais aussi de s'auto-réguler en permanence pour conserver un état d'équilibre (ce qui correspond au principe biologique d'homéostasie, Man & Damasio, 2019) de manière autonome et dans des environnements non structurés en exploitant leurs caractéristiques physiques et leurs compétences « cérébrales ». Ils pourront contrôler par eux-mêmes leur adaptation et leurs apprentissages (Cully et al., 2015), ils seront physiquement résilients aux dégradations qu'ils pourront subir (Bilodeau & Kramer, 2017) et seront intégrés dans des systèmes collectifs (Petersen et al., 2019). Ces évolutions laissent imaginer que les robots tisseront des relations sur le long-terme avec les agents biologiques, occupant également des rôles de compagnon (Cominelli et al., 2018).

Comme nous venons de le présenter, les humains sont amenés à rencontrer des robots de plus en plus fréquemment dans leur quotidien. Ceci pose la question suivante : pour être accepté comme partenaire d'interaction, est-il préférable que le robot ressemble le plus possible à l'espèce de l'individu avec lequel il interagit (approche intra-spécifique) ou bien est-il, au contraire, préférable qu'il s'en éloigne et ne manifeste les caractéristiques de l'espèce que si elles sont indispensables pour l'interaction (approche inter-spécifique) ?

Dans la partie suivante, nous donnerons des pistes de réflexion et aborderons ces deux approches à la fois chez les humains et les animaux non-humains.

2. Les robots sociaux

2.1. Approche intra-spécifique des robots sociaux

L'approche intra-spécifique est la plus courante en robotique : les robots sont conçus comme étant les plus proches possibles - physiquement et comportementalement - des espèces avec lesquelles ils sont amenés à interagir. Certains auteurs considèrent, en effet, que le robot

doit être perçu comme faisant partie de la même espèce afin de favoriser des interactions sociales fluides et pérennes entre agents biologiques et robotiques (Romano et al., 2019).

Nous nous appuierons pour cette partie sur le très récent modèle de l'interaction humain – robot décrit par Baraka et ses collègues (2020). Dans ce modèle, sept caractéristiques permettant de considérer un robot comme social sont présentées (Figure 1). Trois concernent directement le robot : 1) son apparence, 2) ses habiletés sociales, 3) ses degrés d'autonomie et d'intelligence ; deux sont en lien avec l'interaction robot-humain : 4) la proximité physique et 5) le profil temporel ; et les deux dernières dépendent du contexte d'utilisation du robot, 6) son rôle relationnel et 7) son domaine d'application.

Estimant que les champs disciplinaires de l'interaction humain-robot et animal non-humain-robot ont de nombreux points communs, nous utiliserons ce modèle, théorisé à partir d'études auprès d'humains, pour définir des concepts généraux de l'interaction animal (toutes espèces confondues) – robot.

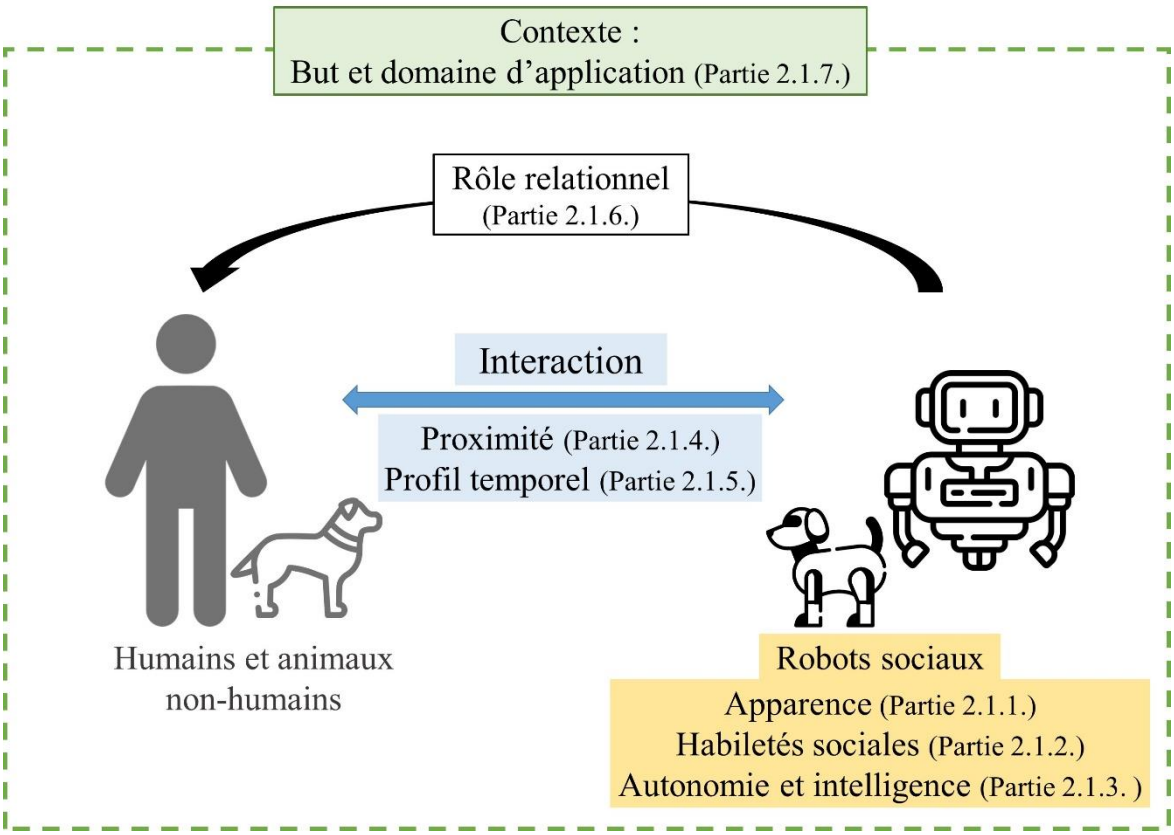


Figure 1. Résumé visuel des sept dimensions du modèle de Baraka et collègues (2020) en fonction du rôle du robot, de l'interaction et du contexte. Chacune des dimensions est discutée séparément dans les parties suivantes.

2.1.1. L'apparence des robots sociaux

Les robots, en comparaison, à d'autres dispositifs technologiques ont un corps : leur expérience du monde physique et des situations sociales ainsi que leurs actions sont des faits tangibles (Wainer et al., 2006). Cette présence physique leur confère des avantages en situation d'interaction : les réactions humaines sont plus positives et les performances plus élevées lorsque les humains peuvent interagir physiquement et directement avec les robots que lorsqu'ils les voient sur des écrans (Li, 2015 pour une revue).

Les corps des robots sociaux ont une multitude de formes différentes. Nous nous focaliserons sur les robots bio-inspirés, conçus d'après des organismes biologiques. Ils

possèdent les caractéristiques corporelles propres à l'espèce dont ils s'inspirent : structure et forme du corps et de la tête et caractéristiques du revêtement extérieur (plastique, peau artificielle, fourrure, plumes ou vêtements).

Les robots peuvent correspondre seulement à des parties du corps, comme c'est le cas de la tête robotique Furhat dont le degré de réalisme très élevé (Figure 2). La surface du visage de Furhat[2] est, en effet, composée d'un masque translucide sur lequel une animation en trois dimensions d'un visage est projetée depuis l'intérieur de la tête. Cette projection animée permet des mouvements du visage plus fluides que ceux des robots mécatroniques. De plus, les caractéristiques physiques du visage (genre, âge, couleur de peau, position des sourcils, taille des yeux et des lèvres, maquillage) sont totalement personnalisables par l'utilisateur.



Figure 2. A gauche, le robot Furhat et Samer Al Moubayed, directeur général de Furhat Robotics, et à droite, certains des masques qui peuvent être projetés sur le robot.

Les robots qui possèdent un corps entier ressemblant à celui d'un humain ou d'un animal non-humain sont qualifiés respectivement d'humanoïdes et d'animaloïdes. Le degré de ressemblance peut être limité, intermédiaire, élevé voire très élevé.

[2] https://furhatrobotics.com/

A présent, décrivons les différents types d'apparence des robots selon leur degré de ressemblance avec les individus de l'espèce considérée (Figure 3).

2.1.1.1.Degré de ressemblance limité

C'est le cas du robot humanoïde LOLA dont la structure mécanique est apparente.

2.1.1.2. Degré de ressemblance intermédiaire

AIBO, le robot-chien, a une apparence proche du jouet en plastique alors que PARO, le robot-bébé phoque se rapproche de celle d'une peluche. Les robots que nous avons utilisés pour nos études, présentés dans la partie §4 (Modèles robotiques), appartiennent à cette catégorie. Le robot humanoïde NAO possède un corps articulé, très mobile, recouvert de plastique. Son visage arrondi aux grands yeux lumineux rappelle celui d'un petit enfant. Le robot-oiseau MANDABOT a été conçu spécialement pour ce projet de recherche. Son apparence (taille, plumage coloré peint sur la structure en plastique) est similaire à celle d'un mâle adulte Diamant Mandarin (*Taeniopygia guttata*).

Dans plusieurs études sur l'interaction poisson-robot, les propriétés morphologiques et locomotrices des robots sur l'attraction sont testées simultanément (Romano et al., 2019). Par exemple, une étude menée chez le Guppy (*Poecilia reticulata* Peters) décrit que l'ajout d'yeux réalistes et d'un mode de déplacement plus naturel conduit à une augmentation significative du niveau d'acceptation du robot-guppy (Landgraf et al. 2016). Chez le Méné jaune (*Notemigonus crysoleucas* Mitchill,) et le Poisson zèbre (*Danio rerio* Hamilton), plus la fréquence des battements de queue du robot est élevée, plus les congénères s'approchent du robot (respectivement Marras & Porfiri, 2012 et Kopman et al., 2013; Bonnet et al., 2016a, b, Bonnet et al., 2017a, b). Les préférences spatiales et la cohésion de petits bancs de Gambusies (*Gambusia affinis* Baird & Girard) varient également selon la morphologie (altération ou respect des proportions corporelles) et la profondeur de nage des robots-gambusies. Quelles

que soient leurs caractéristiques, les robots semblent ne jamais attirer les bancs de poissons. On observe même un effet répulsif du robot dont la taille est supérieure à celle d'un Gambusie, les auteurs faisant l'hypothèse que cela le fait alors ressembler à un prédateur de l'espèce (Polverino & Porfiri, 2013).

2.1.1.3.Degré de ressemblance élevé

Les robots ERICA et SOPHIA, par exemple, présentent une peau artificielle, portent une perruque et des vêtements. On retrouve également le robot-écureuil (*Sciurus carolinensis*, Partan et al., 2009) recouvert d'un pelage de lapin, le robot-grenouille (*Physalaemus pustulosus*, Taylor et al. 2008) ainsi que les 34 robots-animaux[3] développés par la British Broadcasting Corporation (BBC) National History Unit et équipés de caméras dans le cadre des documentaires « *Spy in the Wild* » visant à intégrer des groupes d'animaux, dans leur environnement naturel.

Par ailleurs, intégrer le genre augmente le réalisme du robot. Chez l'humain, un robot est gynoïde lorsque son genre est identifié comme féminin et d'androïde lorsqu'il est identifié comme masculin. Les robots utilisés dans le cadre d'expériences de parade reproduisent les dimorphismes sexuels propres aux espèces étudiées : chez les oiseaux, une femelle robotique Jardinier satiné (*Ptilonorhynchus violaceus*) (Patricelli et al., 2006) et une femelle robotique Tétras des armoises (*Centrocercus urophasianus*) (Patricelli & Krakauer, 2009) ont été conçues et chez les poissons, il existe un mâle robotique killifish (*Lucania goodei* Jordan) (Phamduy et al. 2014).

2.1.1.4. Degré de ressemblance très élevé

Plus récemment, les robots dits géminoïdes sont apparus ; il s'agit de répliques robotiques d'humains identifiables (Hiroshi Ishiguro, roboticien japonais a construit Geminoid HI-4, son

[3] https://www.bbc.co.uk/programmes/m000dlxj

«jumeau robotique» et Geminoid-F, « jumelle robotique » d'une femme dont l'identité est n'a pas été dévoilée).

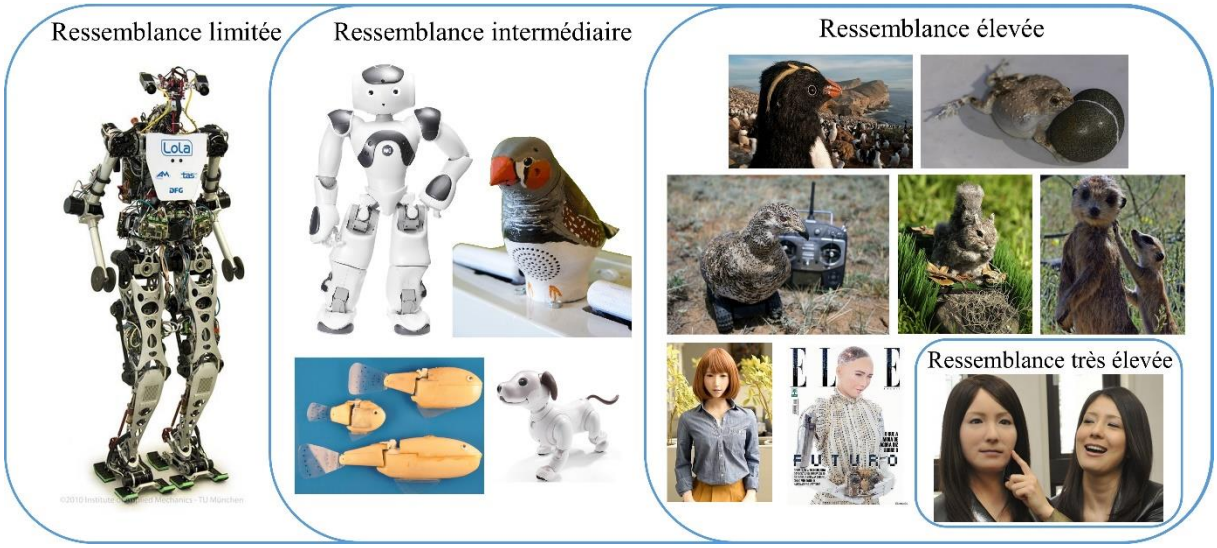


Figure 3. Robots classés selon leur degré de ressemblance avec l'espèce d'interaction. De gauche à droite et de bas en haut : le robot LOLA[4] puis le robot-gambusie[5], le robot-chien AIBO[6], le robot NAO[7] et le robot-diamant mandarin MANDABOT. Dans la catégorie très ressemblant, le robot ERICA[8], le robot SOPHIA[9], le robot tétra des Armoises[10], le robot-écureuil[11], le robot-suricate[12], le robot-pingouin[13] et le robot-grenouille[14]. Concernant la ressemblance très élevée : le robot Geminoid-F[15].

L'acceptation du robot dans l'interaction est grandement conditionnée par son apparence (Phillips et al., 2017). En psychologie, les participants peuvent être interrogés directement sur leur perception des robots et leur attrait. Par exemple, le robot social iSocioBot

[4] Source : https://www.mec.ed.tum.de/en/am/research/completed-projects/humanoid-robotics/lola/
[5] Polverino & Porfiri, 2013
[6] Source : https://us.aibo.com/
[7] Source : https://www.softbankrobotics.com/emea/fr/nao
[8] Source : https://sciencepost.fr/un-robot-dote-dune-ia-va-jouer-dans-un-film-a-gros-budget/. Credits : capture YouTube / Inventions World
[9] Source : https://www.huffingtonpost.fr/2016/12/04/magazine-elle-bresil-robot-couverture_a_21619847/
[10] Source : Patricelli & Krakauer, 2009
[11] Source : Partan et al., 2009
[12] Source : https://www.bbc.co.uk/programmes/m000dlxj
[13] Source : https://spectrum.ieee.org/robot-penguins-spy-on-bird-colony
[14] Source : Taylor et al. 2008
[15] Source : https://vieartificielle.com/androide-geminoid-valle-de-l-etrange

est préféré comme partenaire d'interaction lorsqu'il a une apparence humanoïde plutôt que lorsqu'il est présenté dans sa version mécanique, non humanoïde (Hameed et al., 2016).

Le roboticien Masahiro Mori est le premier à avoir théorisé, en 1970, les réactions humaines possibles face à des agents non-vivants dont les robots. Plus les robots sont proches des humains en apparence et dans leur manière de se mouvoir, plus ils suscitent de la sympathie. Cependant, lorsque la ressemblance est très élevée mais sans être parfaite, le robot provoque alors de la répulsion. C'est l'effet de la « vallée de l'étrange » (Mori et al., 2012) (Figure 4).

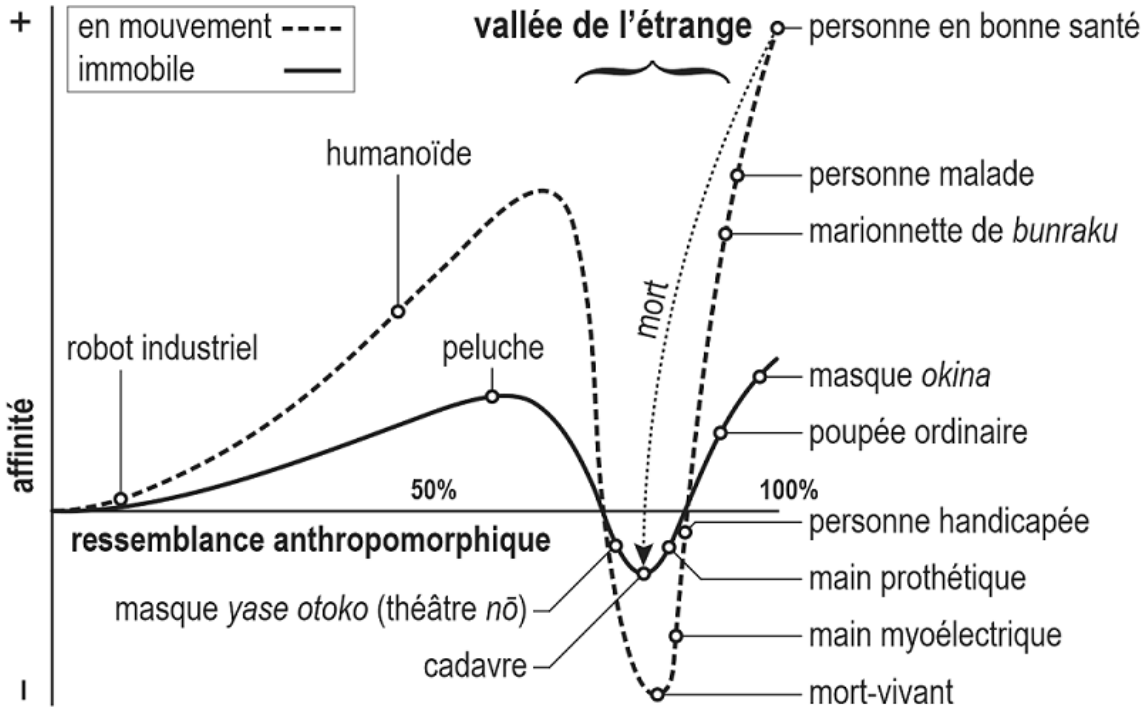


Figure 4. Graphique illustrant la vallée de l'étrange, c'est-à-dire le lien entre le degré de ressemblance anthropomorphique d'une entité (immobile ou en mouvement) et le sentiment d'affinité qu'elle induit chez l'observateur. Les pentes de la vallée de l'étrange pour les entités en mouvement sont plus abruptes que celles pour les entités immobiles. À sa mort, l'individu suit le trajet de la flèche en pointillé du graphique et sort de la courbe des entités en mouvement pour aller vers celle des objets inanimés (Mori et al., 2012).

Les robots-animaux semblent provoquer des réactions similaires chez les humains: les robots ayant un niveau de ressemblance élevé ou bas sont préférés par rapport à ceux qui combinent des caractéristiques réalistes et non réalistes (Löffler et al., 2020). D'autre part, les adultes préfèrent les robots sociaux qui ressemblent à des animaux ou des humains par rapport à ceux qui ont l'apparence de créatures imaginaires ou qui ne ressemblent à aucune créature identifiable (Hosseini et al., 2017).

A notre connaissance, une seule étude a démontré l'effet de la « vallée de l'étrange » chez des animaux non-humains, les macaques crabiers (*Macaca fascicularis*) lorsqu'ils étaient confrontés à des photographies plus ou moins réalistes de macaques (Steckenfinger & Ghazanfar, 2009).

Dans la partie suivante, nous aborderons le fait qu'au-delà de leur apparence, les robots sociaux doivent être socialement compétents. Pour Baraka et collègues (2020), cela repose notamment sur leur capacité à communiquer efficacement, à exprimer des émotions et à manifester des traits de personnalité.

2.1.2. Habiletés sociales

2.1.2.1. Communication

Pour communiquer, le robot peut utiliser le langage ou des modalités non verbales comme les mouvements (regards, gestes et expressions faciales), les lumières et les sons, de manière isolée ou combinée. La communication kinesthésique (appelée couramment « langage corporel ») correspond aux informations transmises par les mouvements des mains, des bras, du corps et de la tête. Le robot Furhat est considéré comme un des robots conversationnels les plus sophistiqués actuellement disponibles. Son visage offre une large variété d'expressions faciales et il peut hocher la tête ce qui lui confère d'importantes capacités

de communication non verbale. Il intègre une intelligence conversationnelle comprenant un système de reconnaissance vocale, la synchronisation labiale automatique dans plus de 40 langues différentes lui permettant de se rapprocher d'une conversation entre humains, avec des tours de parole rapides[16]. Il a notamment été entraîné, dans une étude préliminaire, à jouer le rôle de psychothérapeute chargé de clarifier l'aide dont le patient avait besoin (Johansson et al., 2017).

2.1.2.2. Emotions

Approche comparative et définitions :

Dans la littérature, nombreux sont les auteurs qui soulignent la nécessité de développer une approche comparative des émotions en ligne avec l'idée de continuité phylogénétique du système des émotions décrite par Darwin en 1872 (Anderson & Adolphs, 2014 ; de Waal, 2011 ; Kremer et al., 2020 pour une revue ; Lang, 2010 ; LeDoux, 2016 ; Rolls, 2014)). Paul & Mendl (2018) donnent une définition commune aux humains et aux animaux non humains : une émotion est une réponse multi composantes (subjective, physiologique, comportementale, cognitive) faisant suite à un stimulus ou à un évènement important pour l'individu. La composante consciente, subjective de l'émotion est généralement vue comme sa caractéristique centrale. Une émotion a toujours une valence (agréable ou désagréable). Elle peut être intense ou faible, durer longtemps ou être brève. Le type d'émotion expérimentée (tristesse, douleur, remords) dépendra de la nature précise de l'évènement produisant l'émotion. Un évènement émotif peut être interne, généré par l'imagination ou un souvenir ou venir de l'environnement. Si les animaux non-humains ne peuvent pas décrire verbalement leurs états internes, les corrélats physiologiques, neurobiologiques et comportementaux liés à ces états sont mesurables (Panksepp, 2004).

[16] https://furhatrobotics.com/blog/what-are-social-robots

La fonction principale des émotions est d'orienter les décisions comportementales. Eviter le stimulus déclencheur de l'émotion ou s'en approcher (Briefer, 2018 ; Panksepp, 2010) augmente la probabilité de survie et/ou le succès de reproduction d'un individu (Frijda, 1986 ; Lazarus, 1991 ; Levenson, 1999 ; Tooby & Cosmides, 2008). Par exemple, au sein d'un groupe, les individus qui ont la capacité d'expérimenter la peur sont plus vigilants et ont tendance à s'enfuir davantage lorsqu'un prédateur apparaît (Tooby & Cosmides, 2008).

Les émotions ont également des propriétés socio-communicatives analogues au langage formel (Van Kleef, 2017). Elles sont cruciales pour le bien-être social à tel point qu'elles sont décrites comme permettant la « survie sociale » de l'individu en aidant 1) à développer des relations sociales positives et harmonieuses basées sur la coopération avec autrui (fonction affiliative) et 2) à établir ou maintenir une position sociale par rapport à autrui, préserver l'estime de soi et l'identité parfois au dépend d'autrui (fonction de distanciation) (Fischer et al., 2016).

Un consensus scientifique quant à l'existence de cinq émotions primaires a récemment était proposé (Ekman, 2016). La colère, la peur, le dégoût, la tristesse et la joie, seraient présentes de manière universelle, chez tous les humains, indépendamment des facteurs culturels et sociaux. Panksepp (2005, 2010) défend l'existence de sept systèmes émotionnels primaires communs, a minima, à tous les mammifères et répartis en deux catégories. Lorsque la sollicitation des systèmes équivaut à une « récompense » au niveau cérébral, on parle d'émotions agréables : 1) « seek » : le système de recherche/désir, 2) « lust » : le système sexuel, 3) « care » : le système des soins parentaux et 4) certains aspects du système « play » : le jeu, qui sous-entend notamment des comportements avec un engagement social et physique. Au contraire, les trois autres systèmes poussent les individus à fuir la stimulation et correspondent aux émotions désagréables : 5) « rage » : le système de colère, 6) « fear » : le

système de l'anxiété, 7) « grief » ou « panic » : le système lié à la détresse causée par la séparation.

Les systèmes secondaires résultent de la mémorisation et de l'apprentissage et les systèmes tertiaires, comme la rumination ou la contemplation, mettent en jeu des habiletés cognitives de haut niveau permettant à l'individu d'agir intentionnellement (Panksepp, 2010). Les émotions rattachées aux systèmes tertiaires seraient propres aux humains (Turner & Ortony, 1992). Cependant, des approches comparatives ont nuancé cette séparation notamment concernant la contagion émotionnelle, mécanisme central dans l'empathie (de Waal, 2008). Un individu qui exprime une émotion forte déclenche instantanément une émotion similaire associée à un comportement spécifique chez les individus se trouvant à proximité (de Waal, 2008). C'est un puissant mécanisme de partage d'informations (Preston et de Waal, 2002) qui favorise la vie en groupe (Decety et al., 2016). Bien documentée chez l'humain, la contagion émotionnelle a été décrite à plusieurs reprises chez des animaux non humains (Palagi et al., 2020) comme les rats (Church, 1959 ; Rice & Gainer, 1962), les macaques rhésus (Masserman et al., 1964) mais également, plus récemment, chez le cochon (Reimert, 2013), le chien (Palagi et al., 2015) et le corbeau (Adriaense et al., 2019).

Capacité de reconnaissance des émotions :

Exprimer et reconnaître des émotions chez les humains passe par les expressions faciales (Ekman et al., 1987), le langage corporel (Dael et al., 2012) et l'intonation de la voix (Koolagudi & Rao, 2012). Les primates non humains peuvent également identifier les émotions sur des dessins (Parr, 2003 ; Parr et al., 2007), images, films ou en interactions directes avec leurs congénères (Parr et al., 1998 ; Nahm et al., 1997). Des capacités de reconnaissance des expressions faciales et posturales de conspécifiques ont été démontrées chez le cheval, le

mouton et le chien (Ferretti & Papaleo, 2019). Les positions et mouvements des oreilles et de la queue sont également utilisés comme indicateurs émotionnels (Neethirajan et al., 2021).

Par ailleurs, il existe une reconnaissance inter-spécifique des émotions, ce qui souligne que la capacité à reconnaître des émotions permet à la fois la survie au sein d'un groupe de conspécifiques (survie intra-espèce) mais également dans les interactions entre espèces différentes (Ferretti & Papaleo, 2019). Les humains peuvent reconnaître certaines expressions faciales des singes (Marechal et al. 2017) et des chiens (Wan et al., 2012). Tandis que les expressions humaines faciales de la joie et de la colère peuvent être reconnues par les singes (Kanazawa, 1996), les chiens (Müller et al., 2015), les chevaux (Proops et al., 2018), les chèvres et les pandas géants (Li et al., 2017) (Guo et al., 2019 pour une revue).

Développer des robots émotionnels pour favoriser les interactions :

Comme nous venons de le décrire, les émotions ont un fort impact sur le déroulement des interactions sociales. Aussi, développer des robots expressifs et qui suscitent des réactions émotionnelles est un des challenges majeurs de la robotique actuelle notamment pour les robots au contact de clients ou interagissant de manière répétée avec les utilisateurs. D'après Spezialetti et ses collègues (2020), les futurs robots auront des capacités comparables à celles des humains concernant l'observation, l'interprétation et l'expression des émotions. Ils proposent trois angles d'approche. Le premier est de concevoir des robots ayant leurs propres états émotionnels pour améliorer leur capacité d'adaptation ainsi que leur crédibilité (Hudlicka, 2011). Le second correspond à l'intégration d'intelligence artificielle émotionnelle dans les robots. Abordé comme un levier majeur pour rendre les interactions humain-robot plus efficaces et plus naturelles, ce champ est actuellement en pleine expansion. Il s'agit de rendre les robots capables de détecter, en temps réel, l'état émotionnel de leurs interlocuteurs puis d'y répondre de manière adaptée. Bien que la majorité des IA actuelles se basent sur la

reconnaissance des expressions faciales, une reconnaissance multimodale serait plus pertinente (Lan et al., 2020) car la combinaison d'informations complémentaires augmente la robustesse et la performance. Ainsi, lorsqu'une modalité n'est pas disponible à cause des conditions et/ou du bruit, le système de reconnaissance s'appuie sur les modalités restantes (Hong et al., 2020). Enfin, le troisième volet correspond à la capacité pour les robots d'exprimer des émotions reconnaissables (Mavridis, 2015). Il existe plusieurs manières de rendre les robots expressifs. L'apparence des robots non humanoïdes contraint l'utilisation de canaux sensoriels non anthropomorphiques comme les couleurs des LED, les mouvements et les sons (Löffler et al., 2018). Là encore, exprimer une émotion de manière multimodale semble plus efficace pour la reconnaissance par les humains. La joie, par exemple, est mieux reconnue quand elle est exprimée en utilisant à la fois la couleur et le mouvement (Löffler et al., 2018). Chez les robots humanoïdes, les émotions peuvent se transmettre par les expressions faciales (Rawal et Stock-Homburg, 2021 pour une revue), les postures (Beck et al., 2013 ; Erden, 2013), les gestes (Li & Chignell, 2011) ou la voix. Concernant les expressions faciales, le robot SOPHIA est, par exemple, capable d'exprimer 62 expressions différentes de manière très réaliste (Faraj et al., 2021). La capacité d'exprimer des émotions par la posture du robot a été plus particulièrement étudiée chez les robots ne pouvant pas ou peu exprimer d'émotions avec le visage : le robot NAO (Beck et al., 2010 ; Beck et al., 2013 ; Coulson, 2004 ; Erden, 2013), le robot PEPPER (Tuyen et al., 2018), le robot WE-4RII (Miwa et al., 2004), le robot KOBIAN qui correspond à la partie supérieure du robot précédent, WE-4RII, assemblée à la partie inférieure du robot WABIAN-2R (Zecca et al., 2009) mais aussi des robots sociaux non humanoïdes comme iCat (Cohen et al., 2014) et Keepon (Kozima et al., 2005).

Soulignons qu'en matière d'émotions, il est également important de prendre en compte le contexte culturel dans lequel le robot interagit et la manière dont cela influence l'interaction.

Les émotions exprimées par les robots peuvent, en effet, être interprétés différemment selon la culture d'appartenance (Dang et al., 2017).

Des robots qui suscitent des émotions et de l'empathie :

Le robot social phoque, PARO, a été conçu spécialement pour apaiser les patients atteints de la maladie d'Alzheimer ou de troubles apparentés (Shibata & Wada, 2011, Figure 5). Il répond aux caresses en bougeant ses nageoires et en ouvrant/fermant les yeux, il est capable de chercher le contact oculaire de son interlocuteur et produit des sons similaires à ceux d'un bébé phoque.



Figure 5. Une soignante utilise le robot PARO avec une personne ayant une maladie d'Alzheimer[17].

Geva et ses collègues (2020) ont récemment démontré que, lorsque des adultes en bonne santé caressent PARO, ils se sentent plus joyeux, leur taux d'ocytocine salivaire diminue (dont l'augmentation signe un stress physiologique et psychologique) comme leur sensation de

[17] https://www.phoque-paro.fr/

douleur provoquée expérimentalement. Auprès de patients atteints de démence, bien que limitées, certaines études montrent que PARO réduit les émotions désagréables et les troubles du comportement, améliore l'humeur et la perception des soins (Hung et al., 2019 pour une revue).

Il a même été démontré que les robots pouvaient susciter de l'empathie. De Waal & Preston (2017) définissent l'empathie chez les mammifères à la fois comme la capacité des individus à partager des émotions et à adopter le point de vue d'autrui. Rosenthal-von der Pütten et ses collègues (2013) ont comparé les réactions d'adultes visualisant le robot-dinosaure Pleo, traité avec douceur (caressé et nourri) ou torturé (frappé, étranglé, cogné contre une table) par un humain. Face à la vidéo du robot maltraité, les participants ont eu à la fois des réponses physiologiques (augmentation de la conductance de la peau) et psychologiques significatives : ils ont dit avoir un ressenti plus négatif et avoir éprouvé de l'empathie pour le robot s'exprimant à la fois par de la pitié pour le robot et de la colère contre l'humain le torturant.

Eprouver de l'empathie joue un rôle essentiel dans les comportements pro-sociaux définis comme des comportements volontaires bénéficiant à autrui (Eisenberg, 2006). Des rats ont, par exemple, des comportements différents selon qu'un robot-rat les avait aidés par le passé ou non (Quinn et al., 2018). La séquence d'interaction est décrite sur la Figure 6, ci-dessous. Les rats testés ont préférentiellement libéré de la boîte, le robot-rat qui les avait précédemment eux-mêmes libérés par rapport au robot-rat qui n'avait pas eu de comportement pro-social à leur égard.

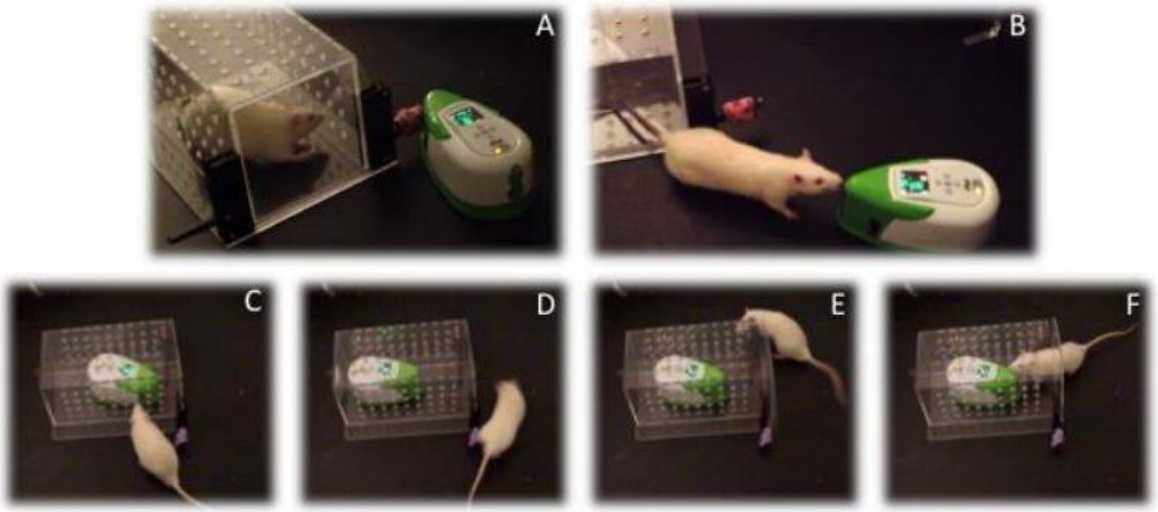


Figure 6. Séquence du comportement d'un rat face au robot-rat. (A) Le robot-rat pousse un levier ce qui libère le rat de la boîte. (B) Le rat pose son nez sur celui du robot juste après que la porte s'est ouverte. (C) Le rat voit que le robot est à son tour enfermé dans la boîte. (D) Le rat se déplace sur le côté de la boîte pour atteindre le levier. (E) Le rat appuie sur le levier ce qui permet d'ouvrir la porte de la boîte (le rat regarde en direction du robot). (F) Le rat interagit immédiatement avec le robot suite à l'ouverture de la porte (Quinn et al., 2018).

Dans leur revue, Oliveira et collègues (2021) ont montré que 12 études sur les 23 analysées rapportaient que les robots et agents virtuels pouvaient induire des comportements pro-sociaux de la part d'adultes. Par exemple, des adultes aidaient significativement plus le robot NAO et étaient plus empathiques avec lui lorsqu'au préalable, le robot avait fait preuve d'adaptation émotionnelle - en disant être dans le même état émotionnel ou avoir la même couleur préférée que l'adulte - que lorsqu'il ne s'était pas adapté (Kühnlenz et al., 2018).

De plus, Connoly et ses collègues (2020) ont montré que les comportements pro-sociaux des adultes à l'égard des robots variaient selon les émotions que les robots exprimaient. Dans leur étude, trois robots Cozmo interagissaient avec le participant dans une tâche de construction en Lego : un robot-cible et deux robots-spectateurs. Lorsque le robot-cible faisait une erreur

dans la construction, l'expérimentateur le maltraitait en disant que ses idées étaient stupides, en le secouant ou en le poussant vigoureusement. Suite à cela, le robot-cible affichait une expression de tristesse et s'éteignait pendant 10 secondes (Figure 7). Les robots-spectateurs pouvaient soit se tourner vers le robot-cible et également afficher une expression de tristesse soit ne rien manifester. Les chercheurs ont montré que dans le cas où les robots-spectateurs exprimaient de la tristesse suite aux comportements maltraitants, les participants venaient significativement plus en aide au robot-cible maltraité, à la fois physiquement en déplaçant le robot-cible pour éviter qu'il ne soit secoué ou verbalement en demandant à l'expérimentateur d'arrêter de maltraiter le robot.



Figure 7. A gauche, le robot Cozmo exprime de la tristesse. A droite, dans la condition expérimentale, les robots-spectateurs bleu et vert expriment de la tristesse suite à l'épisode de maltraitance du robot-cible jaune[18] (Connoly et al., 2020).

2.1.2.3. Personnalité

Chez l'humain, les traits de personnalité sont communément définis comme des modèles de pensées, de sentiments et de comportements relativement stables dans le temps, qui

[18] https://spectrum.ieee.org/can-robots-keep-humans-from-abusing-other-robots

différencient les individus les uns des autres et s'expriment dans des situations particulières (Roberts, 2009). La plupart des psychologues travaillent avec le modèle de la personnalité en cinq facteurs - « Big Five » de McCrae & Costa (1997) - regroupés sous l'acronyme OCEAN. O correspond à « l'Ouverture à l'expérience », soit le degré d'imagination, de curiosité et d'ouverture d'esprit (McCrae & Costa, 1997). Le C désigne la « Conscience », la tendance à être prudent, à délibérer et à être conscient de ses actions (Tasa et al., 2011) ; 3) Le E correspond à « Extraversion/introversion » : un individu extraverti est assertif, bavard et sociable. De plus, l'extraversion est fortement corrélée à l'expressivité émotionnelle, autrement dit, la capacité à exprimer avec précision les ressentis (Riggio & Riggio, 2002). L'introversion a des caractéristiques opposées à l'extraversion et la personne aura tendance à préférer être seule (Driskell et al., 2006). Le A correspond à « l'Agréabilité » soit la mesure dans laquelle l'individu se montre coopératif et amical (Peeters et al. 2006). Enfin, le N est le « Névrosisme », la tendance à se mettre facilement en colère, à ne pas être correctement adapté aux situations rencontrées mais aussi le manque d'assurance et de confiance en soi (Driskell et al., 2006). Le névrosisme est le contraire de la stabilité émotionnelle, qui décrit une personne calme et bien adaptée socialement (Peeters et al., 2006).

Chez les animaux non-humains, les termes de tempérament, de syndrome comportemental et de personnalité sont utilisés. Gosling (2008) propose de n'utiliser que le terme de personnalité afin de s'appuyer sur la littérature foisonnante qui existe chez l'humain. La personnalité est définie comme la présence de variations consistantes entre les individus, stables dans le temps. Les traits les plus étudiés sont le courage, le comportement exploratoire, l'agressivité et la sociabilité (White et al., 2020).

Au vu de ces définitions, comment peut-on rendre tangible la personnalité des robots sociaux dans l'interaction ?

Dou et ses collègues (2019) ont étudié comment la voix du robot PEPPER impactait la manière dont sa personnalité était perçue par les utilisateurs dans un scenario commercial. Ils ont notamment démontré que lorsque le robot avait une voix enfantine, les utilisateurs lui associaient préférentiellement une personnalité extravertie, passionnée ou détendue. Le robot paraissait inoffensif et était mieux accepté que lorsqu'il avait une voix adulte féminine ou masculine.

Dans une revue publiée en 2020, Robert et collègues font état des interactions entre personnalité de l'humain et du robot et ressentis dans l'interaction. Ils en retiennent quatre résultats principaux : 1) l'extraversion semble jouer un rôle fondamental dans l'interaction humain-robot : les individus plus extravertis s'engagent plus volontiers dans les interactions avec les robots, 2) faire varier le comportement (caractéristiques gestuelles et vocales) du robot permet de lui conférer une personnalité extravertie : 3) la préférence des humains pour les robots plus extravertis est modérée et dépend du rôle du robot. En effet, pour les utilisateurs, un robot chargé d'assurer la sécurité n'a pas autant besoin d'être extraverti qu'un robot chargé d'assurer des soins médicaux (Tay et al., 2014).

Nous venons de décrire, en nous basant sur la classification de Baraka et collègues (2020), les compétences sociales (communicatives, émotionnelles et en terme de personnalité) que les robots sociaux devaient présenter dans l'interaction. En plus de l'apparence et des habiletés sociales, l'autonomie et l'intelligence sont les dernières caractéristiques propres au robot qui influencent l'interaction avec des agents biologiques.

2.1.3. Autonomie et intelligence

L'équipe de Baraka (2020) s'est appuyée sur les définitions d'autonomie et d'intelligence de leurs prédécesseurs (respectivement Beer et al., 2014 et Gunderson &

Gunderson, 2004) et les ont étendues pour les rendre applicables au champ de la robotique sociale.

L'autonomie correspond à « *la mesure dans laquelle un robot peut réaliser les tâches pour lesquelles il a été conçu (ou bien les tâches qu'il s'est attribué par lui-même) sans intervention externe* » et soulignent l'utilisation du terme intervention et non pas interaction.

L'intelligence correspond à *« la capacité à déterminer le comportement qui maximisera la probabilité d'atteindre un objectif dans des conditions dynamiques et incertaines, en lien avec l'environnement et l'interaction avec d'autres agents, potentiellement humains* ». Il existe parfois un décalage considérable entre l'intelligence perçue du robot et son niveau d'intelligence réel (Bartneck et al., 2008). Minimiser cet écart permet aux personnes amenées à interagir avec le robot d'avoir des attentes réalistes et des niveaux de confiance dans le robot, appropriés.

Comme souligné précédemment, les intelligences artificielles se perfectionnent dans le but de rendre les interactions humain-robot plus fluides et naturelles. Récemment, Cominelli et ses collègues (2018) ont mis au point un système cognitif spécialement conçu pour les robots sociaux et émotionnels appelé SEAI (Social Emotional Artifical Intelligence). Ce système est bio-inspiré : les auteurs se sont basés sur les travaux fondateurs de Damasio (1994) démontrant une dépendance stricte entre le corps, les émotions, les sentiments, les pensées et les décisions (allant à l'encontre d'une vision de l'intelligence imprégnée du dualisme corps-esprit cartésien). C'est également un système hybride : il permet à la fois modeler des émotions et il présente un haut niveau de capacité de raisonnement. L'équipe de Cominelli (2018) a intégré le système SEAI dans un robot gynoïde, aux expressions faciales particulièrement sophistiquées, FACE (Facial Automaton for Conveying Emotions) et a testé ses réactions face à trois humains manifestant des comportements différents à son égard (positifs, impolis ou mixtes combinant les comportements des deux premières situations). Le système SEAI a permis au robot de se

construire automatiquement des souvenirs au sujet de ces trois rencontres qui ont influencé de manière visible son comportement dans les interactions suivantes avec les mêmes individus. Le robot a ainsi spécifiquement fixé du regard en affichant une expression faciale de contentement l'individu qui avait était chaleureux avec lui.

Au-delà des caractéristiques propres au robot social (apparence, habiletés sociales, degrés d'autonomie et d'intelligence), les caractéristiques de l'interaction en elle-même, à savoir la proximité physique et le profil temporel de l'exposition au robot, semblent cruciales pour l'acception et l'intérêt porte aux robots.

2.1.4. Proximité physique

Le type d'interaction possible et la perception de l'interaction sont grandement influencés par les caractéristiques physiques de l'interaction. De plus, la proximité physique porte une signification sociale, appelée proxémie, et constitue une importante part du comportement non-verbal du robot. On distingue 1) les interactions co-localisées dans lesquelles le sujet et le robot interagissent directement, dans un espace partagé et sans contact physique explicite (ce sont les plus fréquemment retrouvées en robotique sociale) et 2) les interactions physiques.

2.1.5. Profil temporel de l'exposition au robot

Trois variables temporelles peuvent influencer le comportement d'un humain (ou d'un animal non-humain) face à un robot : la période, la durée d'exposition et la fréquence des interactions. On distingue trois catégories de profils temporels, dont les limites précises dépendent du contexte : court, moyen et long-terme.

2.1.5.1.Le court terme

Le court-terme correspond à une ou plusieurs interactions consécutives. La première impression faite par le robot (dépendant notamment de l'apparence, de la taille, des mouvements, de la proxémie, des comportements d'approche et d'initiation de l'interaction) impacte fortement ce type d'interaction. De plus, l'effet de nouveauté est la première réponse face à la technologie et non une réaction qui persistera au cours du temps, lorsque le produit utilisé cesse d'être nouveau pour l'utilisateur (Sung et al., 2009). L'effet de nouveauté est décrit comme « *une motivation, un intérêt ou une participation accrus de la part des individus simplement parce qu'ils réalisent une tâche différente ou nouvelle ce qui affecte leurs réponses d'une manière qui n'est pas liée à la variable dépendante* » (Benge et al., 2012). Cet effet de nouveauté est particulièrement marqué et entraîne des performances initiales plus élevées, non pas à cause d'une réelle amélioration mais à cause d'un intérêt important du sujet pour les outils technologiques (Clark, 1983). Au contraire, un effet de néophobie a également été démontré en début d'interaction : des enfants de 5 ans placés face au robot humanoïde Infanoid se montrent néophobes pendant les 3-4 premières minutes d'interaction, ils sont embarrassés et ne savent pas comment réagir face au robot. Cette phase est suivie d'une exploration tactile du robot puis d'une phase d'interaction dans laquelle l'enfant donne un jouet que le robot montre du doigt (Kozima et al., 2005).

2.1.5.2.Le moyen-terme

Le moyen-terme correspond à des interactions de plusieurs jours à plusieurs semaines. C'est souvent le cas dans les études concernant l'apprentissage des enfants à partir de robots. L'effet de nouveauté disparaît rapidement au cours des interactions. Le nombre d'interactions spontanées d'élèves japonais avec le robot humanoïde Robovie, utilisé pour apprendre du vocabulaire anglais, a par exemple chuté dès la seconde semaine de présence du robot en classe, signant une perte d'intérêt de la part des enfants (Kanda et al., 2004). Cet exemple illustre bien

que plus la durée et la fréquence des interactions augmentent, plus il est difficile de maintenir un investissement de qualité, notamment de la part de sujets ayant des capacités attentionnelles limitées comme les enfants (Baraka et al., 2020).

2.1.5.3.Le long-terme

Susciter un engagement régulier est un enjeu important lorsque les interactions sont répétées au long-terme. Leite et ses collègues (2013) ont présenté dans leur revue de la littérature plusieurs facteurs à prendre en compte lorsque humain et robot interagissent sur le long-terme comme c'est le cas pour les robots présents à domicile. Le robot, par son comportement, doit à la fois donner une sensation de continuité et de nouveauté dans l'interaction. Il doit présenter des comportements routiniers comme saluer son interlocuteur, des comportements dits stratégiques comme rappeler les activités précédentes faites ensemble mais également exprimer progressivement de nouveaux comportements dans les interactions. Dans les interactions à long-terme, la personnalisation du robot devient plus importante. Le robot compagnon Care-O-Bot3 utilisé à domicile est, par exemple, capable de s'ajuster aux différences individuelles des utilisateurs comme leur personnalité, leurs préférences et leurs besoins (Saunders et al., 2016).

Maintenant que nous avons présenté les paramètres à prendre en compte dans l'interaction (proximité physique et profil temporel), abordons le dernier point permettant de définir un robot comme social, le contexte d'utilisation du robot, regroupant à la fois son rôle relationnel et son domaine d'application.

2.1.6 Rôle relationnel

Le rôle relationnel est le rôle pour lequel le robot a été conçu dans l'interaction. Dans nos études, nous nous sommes focalisés essentiellement sur les robots utilisés comme outils de

recherche en sciences sociales qui permettent d'affiner et de valider les théories sur l'interaction sociale, le développement social, psychologique et biologique et les théories concernant la communication émotionnelle et non-verbale. Ces robots imitent les caractéristiques sociales ou psychologiques particulières de l'espèce « *comme s'ils étaient des individus de la même espèce* » (Baraka et al., 2020). Face aux robots sociaux, les utilisateurs ont tendance à faire preuve d'anthropomorphisme : ils leur attribuent des caractéristiques habituellement propres au vivant (Duffy, 2003) et leur prêtent des intentions (Gunkel, 2018). Hildt (2019) voit en l'obtention de la nationalité saoudienne en 2017 par le robot social humanoïde SOPHIA (Katz, 2017), le point culminant de cet anthropomorphisme.

2.1.7 Domaines d'application de l'utilisation des robots sociaux

Les champs d'application des robots sociaux sont extrêmement variés. Même si un robot a été créé pour accomplir une tâche en particulier, ses compétences peuvent être utiles dans d'autres domaines, parfois très éloignés de celui d'origine. C'est le cas du robot NAO qui est utilisé à la fois en recherche et dans l'industrie, mais également pour assister des personnes ayant des troubles cognitifs et enseigner aux enfants. De même, des robots différents peuvent être utilisés dans un même but. Les domaines les plus couramment cités sont l'industrie, les services publics, pour le divertissement et les Arts, pour l'assistance à domicile et au travail, et pour le secourisme.

En résumé, l'approche intra-spécifique a été très développée chez l'humain et conçoit le robot plutôt comme un partenaire de la même espèce. Dans cette perspective, les robots tendent à avoir une apparence, des habiletés sociales (capacités communicatives, émotions, personnalité), des degrés d'autonomie et d'intelligence ressemblant à ceux des humains.

2.2. Approche inter-spécifique des robots sociaux

Miklósi et ses collègues (2017) défendent l'idée de concevoir et développer les robots sociaux, non pas en imitant le plus possible les humains (approche intra-spécifique) mais plutôt en considérant que les robots forment des espèces à part entière, hautement adaptées à leur niche écologique. Ces auteurs proposent de fonder une nouvelle science interdisciplinaire appelée l'étho-robotique et reliant évolution, écologie et éthologie. Bien que ce modèle ait initialement été conçu pour l'interaction humain-robot, nous pensons qu'il est également possible de l'appliquer aux autres espèces animales. En étho-robotique, les apparences, les comportements et les capacités de résolution de problèmes des robots sont spécifiques aux environnements dans lesquels ils évoluent. En d'autres termes, les ressemblances physiques et comportementales qui n'ont pas de lien avec leur fonction spécifique ne sont donc pas pertinentes. Les étho-robots doivent présenter des caractéristiques de l'espèce uniquement si cela est requis et optimal pour leur fonction (comme un feedback verbal ou la présence d'une main humaine).

Ceci permettrait de dépasser l'effet de la vallée de l'étrange car ils seraient des partenaires sociaux plus acceptables que des androïdes/gynoïdes imparfaits. Ils permettraient également de se départir des considérations liées au genre.

Par ailleurs, ils doivent avoir des compétences sociales minimales pour établir une relation et coopérer avec les humains. Parmi ces compétences sont citées : l'attachement aux humains (Topál et al., 2005), des manières simples de communiquer et d'interagir (Miklósi et al., 2000), un degré de réponse à l'apprentissage et à l'entraînement (Topál et al., 2009) et avoir des applications utiles de manière spécifique pour les humains (Ostojic & Clayton, 2014). Pour ces auteurs, ces compétences sont suffisantes pour former la base d'une interaction sociale dès lors qu'il y a un gain mutuel d'expérience.

La relation humain-robot est ainsi abordée comme inter-spécifique plutôt que comme intra-spécifique. Ce type de relation n'est pas unique et peut-être comparée aux relations inter-spécifiques chiens-humains par exemple. Elle correspond à une forme de symbiose dans laquelle les humains bénéficient de conséquences positives pour leur fitness (mutualisme). Ces éthorobots seraient ainsi utiles aux sociétés humaines tout en limitant les dangers potentiels liés au développement technologique : bien qu'ils puissent coopérer avec les humains, ils ne sont pas en compétition avec eux comme ils pourraient l'être s'ils étaient considérés comme faisant partie de la même espèce. Enfin, développer des robots moins ressemblants aux humains permet d'avoir des attentes plus réalistes des capacités des robots dans les interactions (Löffler et al., 2018).

Maintenant que nous avons décrit les caractéristiques des robots sociaux dans l'interaction à la fois dans une approche intraspécifique et une approche interspécifique, nous présentons plus en détail un cas particulier d'interaction, aux applications particulièrement prometteuses : les situations d'apprentissage social. Comme nous le décrirons dans la partie §5 (Objectifs de la thèse et hypothèses de travail), certaines des tâches proposées nous ont permis d'analyser l'apprentissage de l'individu à partir d'un robot social.

2.3. Apprentissage à partir des robots sociaux

2.3.1 Apprentissage social

L'apprentissage social est « un apprentissage guidé par l'observation ou l'interaction avec un autre animal, généralement de la même espèce ou ses produits » (Box, 1984 ; Galef, 1988). Si les espèces avec des organisations sociales complexes comme les humains utilisent très fréquemment l'apprentissage social, son existence a aussi été démontrée chez des espèces de vertébrés et d'invertébrés plus solitaires comme les drosophiles (Danchin et al., 2018), les

tortues (Wilkinson et al., 2010), les pieuvres communes (*Octopus vulgaris*, Fiorito & Scotto, 1992), les corbeaux de Nouvelle-Calédonie (*Corvus moneduloides*, Holzaider et al., 2010) et les orangs-outans (*Pongo spp.*, Schuppli et al., 2016).

Pour Allen (2019), certains comportements se prêtent plus à l'apprentissage social que d'autres. L'apprentissage vocal est une forme d'apprentissage social dans le domaine de la communication. De nombreux animaux sont capables de mémoriser des sons et d'y répondre de manière appropriée. Mais la capacité à modifier, imiter voire inventer de nouveaux sons est une capacité rare dans le règne animal. Cette composante essentielle du langage chez l'humain n'a été décrite que chez quelques espèces de mammifères : les éléphants (Stoeger & Manger, 2014), les chauves-souris (Boughman, 1998), les pinnipèdes (Reichmuth & Casey, 2014) et les cétacés (Janik, 2014), On la retrouve également chez trois groupes d'oiseaux : les colibris, les perroquets et les oscines (aussi appelés « oiseaux chanteurs »), correspondant au taxon des Passeriformes qui englobe quasiment la moitié des espèces d'oiseaux existantes (Catchpole et al., 2008 ; Bolhuis et al., 2010) La dichotomie traditionnelle entre espèces apprenantes et non apprenantes cède néanmoins la place à un continuum (Petkov & Jarvis, 2012): des influences sociales sur le développement vocal ayant été mises en évidence chez certaines espèces de primates (Marshall et al., 1999 ; Takahashi et al., 2017) mais également des oiseaux (Desmedt et al., 2020).

L'apprentissage social est plus utilisé par les individus jeunes que par les adultes, que ce soit chez les humains (Carr et al., 2015 ; Flynn et al., 2016) ou chez d'autres espèces comme le Singe Capucin (*Sapajus spp.)* (Coelho et al., 2015). De plus, tous les individus n'ont pas les mêmes capacités d'apprentissage et tous les tuteurs ne sont pas équivalents, l'apprentissage social est sélectif. Chez l'humain, l'apprentissage par observation est influencé par différentes caractéristiques du démonstrateur. Plusieurs auteurs suggèrent que l'identification des enfants au démonstrateur facilite l'apprentissage social (Meltzoff, 2005; Meltzoff, 2007; Moore, 2006 ;

Over & Carpenter, 2011 ; Uzgiris, 1981 cités dans Shimpi et al., 2013). Ils imitent préférentiellement les individus qui parlent la même langue qu'eux (Buttelmann et al., 2013). Le genre du démonstrateur ainsi que son âge impactent l'imitation (Perloff, 1982). Les enfants préfèrent copier des adultes, plutôt que des enfants du même âge qu'eux, et cela même lorsque les enfants sont plus experts que les adultes pour réaliser la tâche en question (Wood et al., 2012). Les enfants s'appuient également sur le statut social du démonstrateur (un statut social élevé est notamment induit par un âge plus élevé, la popularité, la dominance sociale, d'après Flynn & Whiten, 2012). Chez les animaux non-humains, on retrouve également un effet du sexe du démonstrateur (Benskin et al., 2002; Katz & Lachlan, 2003). De plus, les individus vont préférer copier le comportement exprimé par le plus grand nombre de démonstrateurs au sein d'un groupe (appelé biais de conformité, Lachlan et al., 1998). Ils vont également imiter préférentiellement des individus dont ils sont familiers (Benskin et al., 2002; Guillette et al., 2016) et les individus dominants dans le groupe (Nicol & Pope, 1994, 1999).

Depuis des années, le nombre de supports technologiques disponibles (ordinateurs, tablettes, robots) pour partager des connaissances et apprendre n'a cessé de croître, posant la question de l'intérêt et des limites de ces nouveaux outils. Dans cette partie, nous présenterons des études utilisant des robots sociaux pour l'apprentissage, chez l'humain et les autres espèces animales.

2.3.2 Utilisation des robots sociaux pour l'apprentissage

2.3.2.1. Chez les humains

Dans leur revue concernant les robots sociaux dans le cadre des interactions humains-robots, Lambert et ses collègues (2020) indiquent que les domaines les plus fréquemment abordés sont : 1) le robot comme compagnon social (24% des études), 2) les effets sociaux des robots (évaluer les effets de l'interaction sociale avec un robot conçu pour provoquer un

comportement défini, 24% également), suivis de 3) la définition sociale (les publications qui évaluent et définissent ce qui est socialement attendu d'un robot, 20%), de 4) la santé (17%) et enfin, de 5) l'éducation (15%). On retrouve dans ce dernier domaine les études dans lesquelles les robots sociaux interagissent avec les apprenants pour améliorer leurs capacités d'apprentissage et de rétention d'informations (Lambert et al., 2020).

Une seconde revue (Johal, 2020, extension de la revue de Belpaeme et al., 2018) présente les tendances en matière de recherche pour l'éducation basée sur les robots sociaux : 1) Une majorité des études concerne les enfants âgés de 6 à 11 ans. 2) Les protocoles expérimentaux commencent à se standardiser : le dispositif le plus couramment déployé correspond au robot NAO dans le rôle du tuteur d'apprentissage, qui interagit en face à face avec l'apprenant via une tablette tactile. D'autres objets médiateurs sont utilisés comme des interfaces faites sur mesure avec des boutons par exemple, des cartes, des livres et des ordinateurs. Par ailleurs, Chen et al. (2020), suggèrent de développer les IA des robots sociaux pour en faire des agents flexibles capables d'alterner, selon les besoins, entre le rôle de tuteur pour l'enfant et le rôle d'élève face à l'enfant tuteur. Ce dernier scénario pédagogique, appelé « apprendre en enseignant » (Roscoe & Chi, 2007) est déjà utilisé dans les écoles et permet aux enfants d'améliorer et de consolider leur apprentissage en le transmettant à leurs camarades (Biswas et al., 2005; Brophy et al., 1999; Chin et al., 2013; Park & Howard, 2014; Tanaka & Matsuzoe, 2012, études citées dans Chen et al., 2020) ou à des robots (Jamet et al., 2018 pour une revue). 3) Depuis 2016, les études reportent les impacts à la fois au niveau cognitif et au niveau affectif de l'interaction sur les apprenants, plutôt qu'une seule des deux variables comme précédemment. L'impact cognitif est mesuré par l'évaluation des connaissances des élèves en pré- et en post-test alors que la mesure des aspects affectifs est beaucoup plus variée et dépend des expériences (mesure des expressions faciales de l'élève, mesure de la perception du robot, des temps de regard vers le robot…) (Belpaeme et al., 2018). 4) De plus en plus d'études

cherchent à personnaliser, à individualiser l'apprentissage en adaptant le contenu des séquences au niveau de l'élève et le comportement social du robot, notamment sa personnalité, aux réactions de l'élève. Ceci conduit à une augmentation importante des protocoles basés sur un apprentissage « un pour un » (un tuteur par élève). 5) Du fait de leurs capacités motrices limitées, les robots font en majorité des mouvements à valeur communicative (gestes non-verbaux correspondant au discours ou à des gestes à valeur affective) comparés aux gestes déictiques, comme le pointage, ou aux gestes de manipulation visant à agir directement sur l'environnement. Enfin, 6) le nombre d'études utilisant des robots autonomes plutôt que télé-opérés par l'humain a grandement augmenté.

Concernant les domaines d'apprentissage testés, le robot RASA (Robot Assistant for Social Aims) a été conçu pour apprendre la langue des signes perse à des enfants sourds ou malentendants iraniens (Meghdari et al., 2018). Il a été testé dans une étude pilote auprès d'adultes et devrait prochainement être utilisé auprès d'enfants (Figure 8).

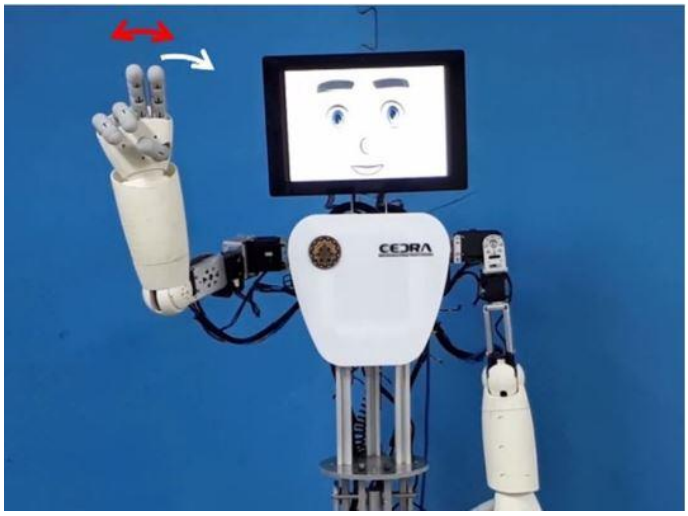


Figure 8. Le robot RASA fait les gestes correspondant aux mots « lundi » (flèche blanche) et « vingt » (flèche rouge) en langue des signes perse (Meghdari et al., 2018).

Dans l'étude de Kory Westlund et ses collègues (2017), le robot Tega (Figure 9) racontait une histoire contenant des mots-cibles à retenir à des enfants de 5 ans. Non seulement, les enfants ont immédiatement appris le vocabulaire sur le moment mais ils ont également été capables de raconter l'histoire immédiatement après l'interaction et entre 4 et 6 semaines plus tard.

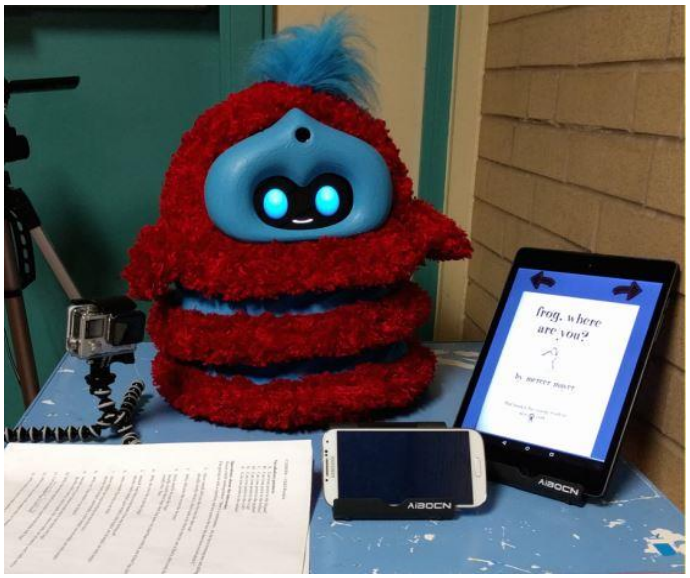


Figure 9. Dispositif expérimental dans lequel le robot Tega raconte une histoire à un enfant (Kory Westlund et al., 2017).

2.3.2.2. Chez les animaux non humains

Quelques études en laboratoire (voir Romano et al., 2019 pour une revue) ont montré que les robots pouvaient être utilisés comme tuteurs d'apprentissage notamment en contexte alimentaire. Par exemple, le robot-rat WM-2 peut conduire des groupes de rats vers des sources de nourriture (Takanishi et al., 1998). Dans deux autres études, il a été démontré que les rats pouvaient apprendre d'un robot-rat à pousser un levier pour obtenir de la nourriture (Laschi et al., 2006 ; Patanè et al., 2007). Enfin, chez une espèce d'oiseaux, le Roselin du Mexique (*Carpodacus mexicanus*), les individus passaient plus de temps à fourrager et à chercher de la nourriture quand des robots-oiseaux simulaient un mouvement corporel associés à un

fourragement fructueux (Fernández-Juricic et al., 2006). Les robots peuvent être utilisés pour étudier le phénomène d'empreinte défini comme une courte période précoce, après la naissance ou l'éclosion, durant laquelle l'animal est capable d'un apprentissage rapide et d'établir une réponse durable à un stimulus spécifique (Lorenz 1935). L'attachement pour un objet donné peut se faire via des stimulations visuelles, auditives ou tactiles. En conditions naturelles, les juvéniles s'attachent généralement à leurs parents mais dans en conditions expérimentales, d'autres animaux, des objets comme des robots peuvent être source d'empreinte. Plusieurs études ont mesuré un attachement au robot chez les poussins de poule (*Gallus gallus domesticus* Linnaeus, Gribovskiy et al., 2010, 2015, 2018) et de caille japonaise (*Coturnix coturnix japonica*, De Margerie et al. 2011, 2013 : Jolly et al. 2016, Figure 10).

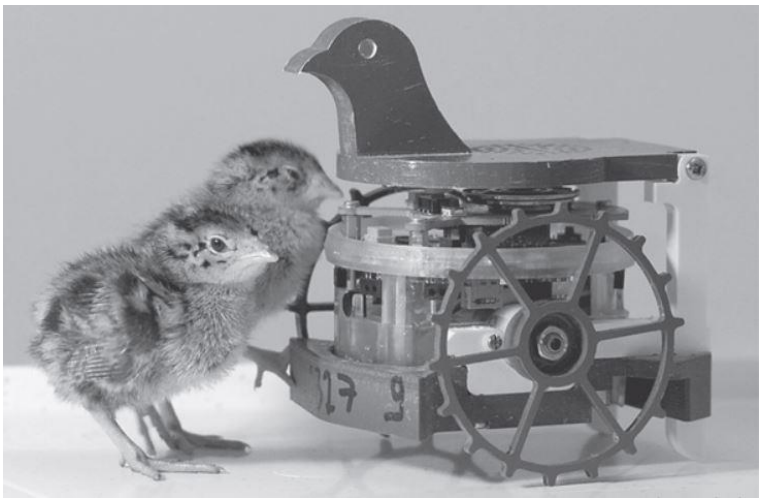


Figure 10. Le robot-caille et deux cailleteaux (*Coturnix coturnix japonica*) présentés dans l'étude de Jolly et collègues (2016).

Maintenant que nous avons détaillé les caractéristiques des robots sociaux, nous allons présenter nos deux modèles biologiques. Nous avons choisi d'étudier deux espèces hautement sociales, l'humain et une espèce d'oiseau, le Diamant Mandarin. Nos études s'inscrivent dans un contexte développemental et ont été menées auprès de jeunes individus car l'apprentissage précoce est déjà bien étudié chez ces deux espèces.

3. Choix des modèles

3.1. Modèles biologiques

3.1.1 Les enfants

Notre choix s'est porté sur les enfants au développement typique (DT) âgés de 3 à 6 ans ainsi que sur des enfants avec un diagnostic de Trouble du Spectre Autistique (TSA) de niveau 1 (léger), âgés de 4 à 8 ans.

D'après le DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) de l'Association Américaine de Psychiatrie, le TSA est un « trouble neuro-développemental caractérisé à la fois par des difficultés dans la sphère de la communication et de l'interaction sociale ainsi que des comportements/intérêts/activités restreints et répétitifs. Les symptômes sont présents précocement dans la vie de l'individu et affectent son fonctionnement au quotidien. Le terme de « spectre » illustre l'hétérogénéité dans les manifestations et la sévérité des symptômes ainsi que les compétences et le niveau de fonctionnement des individus présentant un TSA. Trois niveaux de sévérité ont été décrits selon les adaptations environnementales nécessaires au quotidien : 1) l'enfant a besoin d'un soutien, 2) l'enfant a besoin d'un soutien important et 3) l'enfant a besoin d'un soutien très important.»

Nous avons choisi ces tranches d'âges car les études sur l'interaction jeunes enfants-robots sont encore rares. Comme nous le présenterons au Chapitre II, plusieurs études ont mis en évidence que les enfants TSA étaient très intéressés par les robots ce qui en fait de potentiels médiateurs éducatifs et thérapeutiques. Souhaitant comparer les résultats obtenus auprès de ces deux populations, notre protocole expérimental initial a été pensé pour être accessible à la fois aux enfants DT et aux enfants TSA léger.

Les enfants DT ont été recrutés au sein de quatre écoles (trois dans le Sud-Ouest de la France et une en Ile-de-France). Au total, 63 enfants ont été recrutés. Tous les enfants ont été autorisés à participer par leurs parents via des formulaires de consentement écrits. Les sessions

se sont déroulées au sein des écoles, dans une pièce calme mise à disposition par les établissements (Figure 11).



Figure 11. Enfant réalisant l'expérience avec le robot humanoïde NAO en école.

De plus, une collaboration a été mise en place avec un Service d'Education Spéciale et de Soins A Domicile (SESSAD), dans le Sud-Ouest de la France. En mai 2021, 6 enfants âgés de 4 à 8 ans, avec un TSA de niveau 1, ont participé à l'étude. Tous les enfants ont été autorisés par leurs parents via des formulaires de consentement écrits. Les sessions se sont déroulées dans une pièce calme bien connue des enfants (Figure 12).

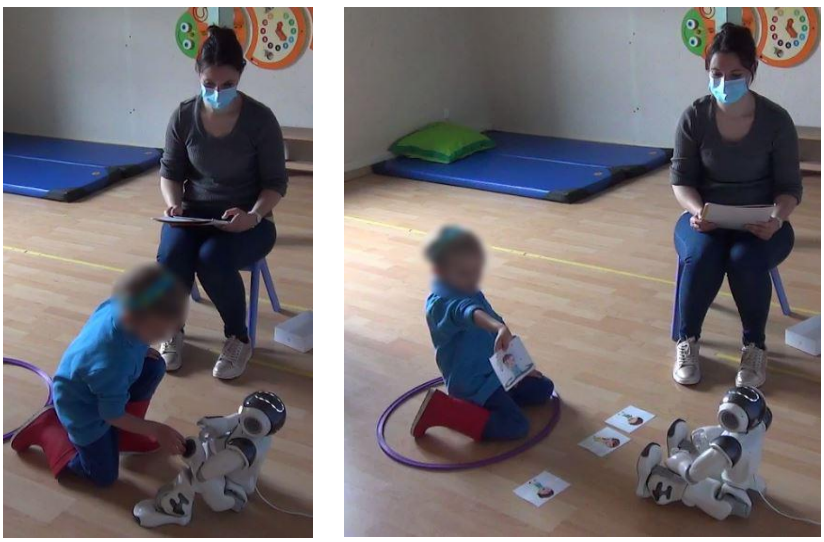


Figure 12. Enfant réalisant l'expérience avec le robot humanoïde NAO au SESSAD.

3.1.1.1. Compétences des enfants dans les domaines étudiés dans nos expériences

Dans cette partie, nous présentons les compétences des enfants dans les domaines évalués dans nos études (Chapitres I et II), à savoir le langage oral, la connaissance des parties du corps et des émotions, les comportements pro-sociaux, afin de mettre en perspective les comportements des enfants testés.

D'un point de vue théorique, le développement de l'enfant est abordé sous l'angle de la psychomotricité, c'est-à-dire en considérant que le développement des systèmes cognitif, émotionnel, symbolique et sensori-moteur sont étroitement liés et que leurs interactions influencent la capacité de l'enfant à s'exprimer dans un contexte psychosocial donné (Cueto et al., 2017).

3.1.1.2. Le langage oral

Durant les deux premières années de vie, les enfants font deux acquisitions majeures qui bouleversent leur rapport à autrui, au monde et aux objets : la locomotion autonome et le langage (Iverson, 2010). Le développement du langage oral passe par plusieurs étapes reprises par Alaria & Gentaz (2018). Deux mois après sa naissance, le bébé babille déjà, il imite avec de plus en plus de fidélité les intonations et la prosodie des adultes autour de lui. Au cours de la première année, sa perception des phonèmes (unités sonores qui permettent de différencier deux mots dans une langue) s'améliore en même temps que son répertoire de phonèmes perçus diminue et se limite à ceux de sa langue natale. A 18 mois, l'enfant est capable d'énoncer une cinquantaine de mots, surtout des noms, des verbes, des adjectifs et des onomatopées. A 24 mois, son répertoire s'est considérablement étoffé puisqu'il comprend en moyenne 600 mots qu'il peut agencer pour former des phrases. A 3 ans, les enfants connaissent environ 1000 mots et en apprennent 500 nouveaux par an jusqu'à leurs six ans. On parle à cette période d'explosion du langage : les enfants sont en mesure de s'exprimer avec de plus en plus d'aisance, ils utilisent

toutes les catégories grammaticales (noms, verbes, pronoms personnels, adjectifs, adverbes) et une syntaxe de plus en plus élaborée.

Comme défini dans la partie §3.1.1., la sphère de la communication est souvent largement impactée chez les individus avec un TSA qui produisent leurs premiers mots autour de 38 mois contre 8 à 14 mois chez les enfants DT (Howlin, 2003). En 2018, la Haute Autorité de Santé a publié une fiche de synthèse sur les bonnes pratiques face aux enfants avec TSA. Ainsi, les signes majeurs d'alerte du TSA chez un jeune enfant, comprennent : 1) une absence de babillage, de pointage à distance ou de gestes sociaux à visée communicative (dire « bonjour ») à 12 mois et au-delà, 2) une absence de mots à 18 mois et au-delà et 3) une absence d'association de mots (non écholaliques[19]) à 24 mois et au-delà. Les enfants avec un TSA ont des déficits variables concernant l'acquisition du langage. Sont touchés : 1) la sémantique soit la connaissance du sens des mots et leur correspondance avec le monde réel. 2) La syntaxe c'est-à-dire la capacité à combiner les mots entre eux pour former des phrases. 3) La phonologie, la production des sons et l'articulation du discours. 4) La morphologie : les morphèmes sont les plus petites unités signifiantes du langage. Le développement morphologique correspond au développement et à la compréhension de ces unités et de la manière dont ces unités se combinent pour former des mots. 5) La lecture et la capacité de narration ainsi que 6) l'utilisation du langage en contexte social (Tager Flusberg et al., 2005 ; Prelock & Nelson, 2012 cités dans Ahtam et al., 2020).

3.1.1.3. La connaissance des parties du corps

Pour Raimo et ses collègues (2019), les représentations corporelles se distinguent en trois catégories. La première, qualifiée de « schéma corporel » est une représentation dynamique des parties du corps liées à la fois aux informations motrices et sensorielles reçues

[19] L'écholalie est la répétition de syllabes, de mots ou de phrases (Stiegler, 2015).

par l'individu (Schwoebel et al., 2002 ; Schwoebel & Coslett, 2005). La seconde est qualifiée de carte corporelle visuo-spatiale ou de représentation structurelle du corps. C'est une représentation topographique du corps venant principalement des informations visuelles concernant les délimitations entre chaque partie du corps et les relations spatiales de proximité (Schwoebel & Coslett, 2005 ; Sirigu et al., 1991). Enfin, la dernière est la représentation lexico-sémantique du corps, elle inclut le nom des parties du corps, leur fonction et leur lien aux objets (Coslett et al., 2002 ; Schwoebel & Coslett, 2005). Les enfants DT de 2,2 ans à 3,4 ans savent montrer sur leur corps 15 à 33 parties sur les 50 testées. Les parties du corps les mieux identifiées sont la langue, le pied, la tête, le dos, le ventre, la bouche, l'œil, les dents, la main, la jambe, les doigts, le bras, l'oreille, les fesses et les cheveux. Les parties du corps les moins connues sont la tempe, la cuisse, l'aisselle, la paume de la main et la plante du pied (Camões-Costa et al., 2011). La connaissance d'une partie du corps par l'enfant dépend en premier de sa localisation, les parties du visage et les parties du corps les plus larges sont, en effet, connues en premier tandis que les articulations sont les moins bien repérées. Elle est aussi liée au nombre de représentations sensorielles (mais pas motrices) de chaque partie du corps dans le cortex humain et de la fréquence avec laquelle elle est nommée par les adultes (Camões-Costa et al., 2011).

Une partie du test d'imitation de gestes de Bergès et Lézine (1963) permet d'évaluer la compréhension et la production des noms des parties du corps chez les enfants de 3 à 6 ans. L'enfant doit montrer les parties du corps nommées par l'adulte puis nommer les parties du corps pointées par l'adulte, à la fois sur son corps à lui et sur le corps de l'‘adulte. Dans ces tâches, les enfants avec un TSA, âgés de 5 ans et demi, ont significativement plus de difficultés que les enfants DT, âgés de 4 ans et demi. La tâche de compréhension des noms est réussie à 42.65% et celle de production des noms à 37.28% chez les enfants avec TSA alors que chez les enfants DT, elles sont réussies respectivement à 70.44% et 56.76% (Russo et al., 2018).

3.1.1.4. L'identification et l'expression des émotions

La reconnaissance et l'expression des émotions sont deux compétences imbriquées l'une dans l'autre. En effet, lors des interactions en face à face, l'individu reconnaît l'émotion spécifique exprimée par son interlocuteur en le regardant puis il recrée l'émotion par un processus d'imitation (Denham et al., 2003). Dès 3 ans, les enfants sont en mesure d'associer l'expression faciale au nom de l'émotion exprimée pour les six émotions primaires (Widen & Russel, 2003). Les visages et les corps ont des propriétés physiques communes et transmettent les informations sociales et émotionnelles similaires ce qui sous-entend que l'encodage des postures et des visages suit un processus commun (Reed et al., 2003 ; Reed et al., 2006 ; Stekelenburg & de Gelder, 2004 ; Van den Stock et al., 2007 cités dans Parker et al., 2013). Par exemple, quand une personne est en colère, les muscles de son visage (notamment au niveau des sourcils) et de son corps (surtout les épaules et les bras) se tendent. Parker et ses collègues (2013) ont développé une échelle de mesure de la reconnaissance des émotions chez les enfants et les adolescents (CARE). Ils ont pour cela pris des photographies d'enfants exprimant les 6 émotions primaires ainsi qu'une expression neutre à la fois avec leur visage et avec leur corps. Ils les ont ensuite montrées à des enfants de 3 à 6 ans et ont trouvé que ces derniers étaient capables de nommer les émotions présentées sur les photographies. Chez les enfants TSA, le développement de la régulation émotionnelle est retardée à la fois en ce qui concerne la reconnaissance des expressions faciales émotionnelles et leur expression (Lecciso et al., 2021). Cependant, concernant les postures corporelles émotionnelles, les enfants de 9 ans, ont les mêmes capacités de reconnaissance des 6 émotions primaires qu'ils aient un TSA ou qu'ils aient un développement typique (Peterson et al., 2015).

3.1.1.5. Les comportements pro-sociaux

Les comportements pro-sociaux englobent tous les comportements volontaires dont l'intention commune est d'être bénéfique pour autrui. Dès 2 ans, les enfants expriment des comportements pro-sociaux à l'égard d'un adulte manifestant une variété de besoins différents comme un « besoin instrumental » ou de la détresse émotionnelle (Dunfield & Kuhlmeier, 2013). Dans l'étude de Dunfield & Kuhlmeier (2013), la situation de besoin instrumental était la suivante : l'expérimentateur faisait tomber une pièce de puzzle sans parvenir à la ramasser. Il regardait alors alternativement la pièce et l'enfant jusqu'à ce que celui-ci réagisse, sans lui demander explicitement de la ramasser. Les auteurs ont trouvé qu'à 2 ans la majorité des enfants aidaient l'adulte à retrouver la pièce de puzzle, soulignant que répondre au besoin instrumental d'autrui est l'un des premiers comportements pro-sociaux à émerger. Dans la situation de détresse émotionnelle, soit l'expérimentateur se blessait en se cognant le genou, soit il abimait un jouet en l'utilisant. Dans un second temps, il regardait l'enfant en lui demandant "que devrions-nous faire ?". Les auteurs ont trouvé que la tendance des enfants à répondre à la détresse émotionnelle de l'adulte augmentait entre 2 et 3 ans pour ensuite se stabiliser entre 3 et 4 ans. Ils ont également relevé que les enfants répondaient plus quand la détresse émotionnelle était liée au jouet cassé que quand elle était liée au fait de se blesser, faisant l'hypothèse qu'il est plus aisé pour les enfants de répondre à une demande concrète mettant en jeu un objet (le jouet cassé) qu'à une demande strictement affective, dans laquelle la réponse appropriée que l'enfant doit donner est moins évidente.

En 2019, Dunfield et ses collègues ont publié une nouvelle étude en comparant, cette fois-ci, les comportements pro-sociaux d'enfant avec un TSA (âgés en moyenne de 3,9 ans) et DT (âgés en moyenne de 2,8 ans). De nouveau, ils ont testé une situation de besoin instrumental nécessitant que l'enfant ramasse le jouet que l'expérimentateur avait fait tomber (inspirée de la tâche de la pince à linge de Warneken & Tomasello, 2006) et la même tâche de détresse

émotionnelle que dans leur précédente étude, en conservant uniquement la partie où l'expérimentateur se cognait le genou. Ils ont démontré que dans la tâche de besoin instrumental, les enfants DT ramassaient plus le jouet lorsque l'expérimentateur montrait qu'il ne pouvait pas l'atteindre (condition expérimentale) que lorsqu'il croisait ses bras, sans chercher à le récupérer (condition contrôle). Les enfants avec un TSA, quant à eux, ramassaient autant le jouet dans les deux conditions. Dans la condition expérimentale, 50% des enfants DT ont ramassé le jouet contre 21,4% des enfants avec TSA alors que dans la condition contrôle, aucun enfant quel que soit son groupe n'a ramassé le jouet. Concernant la tâche de détresse émotionnelle deux fenêtres temporelles ont été analysées, 10 sec et 30 sec après que l'adulte s'est cogné car il a été montré que les enfants mettaient plus de temps à répondre à une détresse émotionnelle qu'à une situation de besoin instrumental. Dans les 10 sec, les enfants DT et TSA ont autant réconforté l'adulte dans les deux conditions (dans la condition expérimentale, l'adulte exprime qu'il souffre après s'être cogné alors que dans la condition contrôle, l'adulte affiche une expression neutre après s'être cogné). En revanche, dans les 30 sec, les enfants DT et les enfants TSA réconfortaient plus l'adulte dans la condition expérimentale que dans la condition contrôle. De leur étude, Dunfield et ses collègues (2019) ont conclu que les enfants avec un TSA pouvaient se représenter les besoins d'autrui mais que, lorsque le comportement pro-social était coûteux, comme dans la situation de besoin instrumental, les enfants avec un TSA étaient moins enclins à manifester un comportement pro-social. Ces résultats suggèrent que s'engager dans un comportement pro-social nécessite à la fois de pouvoir reconnaître le besoin d'autrui et d'être motivé pour agir en sa faveur.

Notre second modèle biologique est une espèce d'oiseau, le Diamant Mandarin. Comme pour les enfants, nous allons présenter le domaine étudié dans notre étude (Chapitre IV) : l'apprentissage du chant.

3.1.2 Les diamants mandarins (Taeniopygia guttata)

Le Diamant Mandarin fait partie des Passériformes (sous-ordre des oscines et famille des Estrildidés). Même si le Diamant Mandarin (*Taeniopygia guttata castanotis*) vit encore à l'état sauvage en Australie, c'est sous sa forme domestique que cette espèce est plus connue. C'est une espèce granivore, qui, en milieu naturel, se nourrit occasionnellement d'insectes et fourrage en groupes pour trouver des ressources alimentaires dispersées (Zann, 1996). C'est une espèce hautement sociale organisée en colonies, qui présente un fort nomadisme (Zann, 1996). Mâles et femelles présentent un dimorphisme sexuel au niveau de leur plumage. Ils sont monogames et forment des couples stables qui durent toute leur vie. Ils se reproduisent aisément et rapidement, quelle que soit la période de l'année. Ils sont par ailleurs, très bien adaptés à la vie en captivité.

Depuis plus de 50 ans, le Diamant Mandarin est l'espèce la plus étudiée pour les recherches en laboratoire sur le chant des oscines. Chez cette espèce, seuls les mâles chantent. Les jeunes apprennent à chanter pendant une période sensible de la vie précoce en écoutant et en imitant principalement le chant de leur père, mais aussi en copiant certaines parties du chant d'autres mâles vivant dans leur environnement proche. Comme chez les autres oscines, on retrouve dans le cerveau du Diamant Mandarin un réseau de structures cérébrales appelé « système du chant » (« song control system ») permettant la production, la perception et l'apprentissage du chant (Nottebohm et al., 1976).

La période sensible s'étend du 25 au 90ème jour après l'éclosion et comprend une phase sensorielle pendant laquelle les jeunes mémorisent les chants de leurs tuteurs et une phase sensorimotrice pendant laquelle le jeune commence à produire des chants. Environ 3 mois après son éclosion, la structure du chant du jeune est stabilisée et ses caractéristiques acoustiques n'évolueront plus jusqu'à la fin de la vie de l'individu. Le Diamant Mandarin est ainsi classé

comme un « *age-limited learner* », on parle d'apprentissage fermé du chant (Brainard & Doupe, 2002; Catchpole & Slater, 2008).

Le chant du diamant mandarin est stéréotypé et commence généralement par un ensemble de sons appelés syllabes introductives suivies d'une ou plusieurs répétitions de l'unité de base du chant, le motif. Un motif est bref (0.5 à 1.5 sec) et correspond à la répétition dans un ordre donné d'une séquence de syllabes stéréotypées (Price, 1979 ; Sossinka & Böhner, 1980 ; Williams, 1990 ; Zann, 1996). Les motifs sont reliés entre eux par des connecteurs (Hyland Bruno & Tchernichovski, 2017) et l'ensemble forme des phrases de chant (« song bouts ») définies comme une suite de chants ininterrompue (Zann, 1996) comme présenté sur la Figure 13.

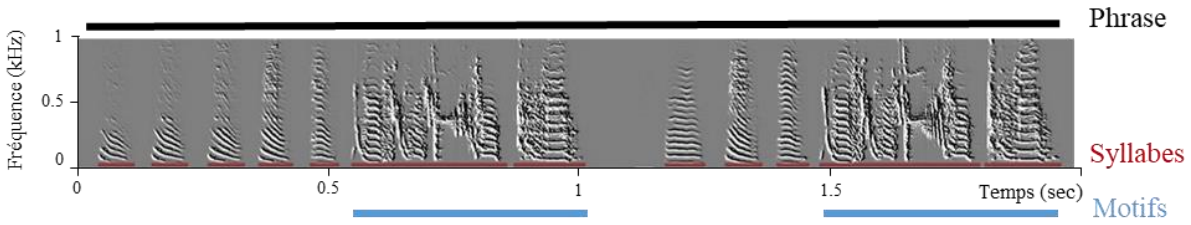


Figure 13. Spectrogramme d'un chant de diamant mandarin. La barre noire au-dessus du spectrogramme correspond à une phrase de chant et les barres bleues en-dessous aux motifs. Le chant peut être divisé en syllabes distinctes, représentées ici par les barres rouges et séparées entre elles par des silences.

Depuis les travaux pionniers d'Immelmann (1969), des milliers de diamants mandarins ont été élevés dans des environnements sonores contrôlés afin d'identifier les facteurs qui influencent leur apprentissage du chant. De jeunes mandarins élevés sans tuteur (isolement social) développent des chants présentant certaines des caractéristiques spécifiques de l'espèce mais également des sons aberrants (Price, 1979).

En laboratoire, différentes procédures expérimentales ont été utilisées pour entraîner de jeunes diamants mandarins à chanter (Derégnaucourt, 2011 ; Derégnaucourt, Araguas et

Guellaï, en préparation). Contrairement à d'autres espèces d'oscines comme le Canari (*Serinus canaria*) ou le Rossignol Philomèle (*Luscinia megarhynchos*), les jeunes mandarins exposés passivement à des enregistrements de chants de leur espèce diffusés par haut-parleur imitent peu les modèles qu'ils entendent (Derégnaucourt, 2011). Dans les dispositifs utilisant un conditionnement opérant, les oiseaux sont plus actifs dans leur apprentissage : ils peuvent entendre du chant diffusé par haut-parleur en donnant un coup de bec sur une clef ou en se posant sur une perche (Adret, 1992 ; ten Cate, 1991 ; Derégnaucourt, 2011 ; Houx & ten Cate, 1999). Ce système présente des résultats contrastés et une forte variabilité interindividuelle : certains oiseaux apprennent de manière significative, voire produisent une copie très fidèle du modèle de chant diffusé tandis que d'autres ne montrent pas de signe d'apprentissage (Derégnaucourt et al., 2005). Enfin, l'exposition passive à des vidéos de mâles adultes en train de chanter (Deshpande et al., 2014 ; Ljubičić et al., 2016) ne permet pas aux jeunes d'apprendre le chant de leur tuteur.

Finalement, élever un jeune mâle en compagnie d'un mâle adulte est la situation idéale pour obtenir une copie quasi-parfaite du chant. Cette relation dyadique jeune-adulte constitue une référence pour les recherches sur le chant d'oiseau (Derégnaucourt, 2011) et a notamment permis de déterminer l'impact des interactions entre le jeune et son tuteur sur l'apprentissage du chant : si l'on sépare les jeunes de leur tuteur par une cloison opaque ou une grille, ils copieront moins bien le chant de leur tuteur que des oiseaux pouvant interagir physiquement et visuellement (Derégnaucourt, 2011).

De plus, il a été récemment démontré que les jeunes imitaient plus fidèlement le modèle de chant lorsqu'ils été exposés de manière contingente à un feedback social (diffusion d'une vidéo de femelle diamant mandarin présentant une posture corporelle particulière impliquant un mouvement des ailes) au moment où ils commençaient à chanter que lorsque ce feedback arrivait de manière aléatoire (Carouso-Peck & Goldstein, 2019). Ce résultat laisse supposer que

la contingence comportementale des partenaires d'interactions influence également l'apprentissage du chant.

Bien que ces méthodes aient permis de mieux comprendre les mécanismes impliqués dans l'apprentissage du chant, elles ne permettent pas un contrôle précis des différentes variables impliquées dans l'apprentissage du chant, comme le contrôle du comportement du tuteur (contingent ou non de celui du jeune) ou le nombre de chants entendus, ce qui est rendu possible par l'utilisation d'un robot.

Comme mentionné ci-dessus, l'apprentissage du chant chez le Diamant Mandarin est un modèle biologique de choix pour étudier les bases neuro-éthologiques de l'apprentissage vocal, une composante importante du langage. Depuis plusieurs décennies, les analogies entre l'acquisition du langage chez l'humain et l'apprentissage du chant chez les oscines ont été mis en exergue et impliquent entre autres une intégration sensorimotrice, des périodes sensibles d'apprentissage, ou encore une spécialisation hémisphérique au niveau du contrôle neuronal (Bolhuis et al., 2010). Le développement rapide du chant chez le Diamant Mandarin permet également d'étudier les corrélats neurophysiologiques, hormonaux et génétiques de l'apprentissage vocal. Ainsi, on a pu par exemple démontrer l'influence du sommeil sur l'apprentissage vocal à la fois au niveau comportemental (Derégnaucourt et al., 2005) mais aussi au niveau cellulaire et moléculaire (Shank et al., 2009 ; Roberts et al., 2010). La production de diamants transgéniques permet d'étudier les corrélats génétiques de certains troubles neuro-développementaux chez l'humain comme par exemple la maladie de Huntington (Liu et al., 2015), ou encore des troubles de l'acquisition du langage impliquant le gène FoxP2 (Haesler et al., 2007).

Certaines de ces procédures restent compliquées (ex : transgenèse) et la règle des 3R (Réduire, Raffiner, Remplacer) fixent le cadre éthique de l'expérimentation animale. Ainsi, il faut à la fois réduire le nombre d'animaux utilisés, raffiner les protocoles expérimentaux (par

exemple, en identifiant les points limites ou en enrichissant les conditions d'hébergement) et éventuellement songer à remplacer les animaux par des méthodes alternatives ou des modèles. Concernant ce dernier point, l'apprentissage vocal fait l'objet de travaux visant à modéliser ces caractéristiques à la fois chez l'humain (Pitti et al., 2021) mais aussi chez les oiseaux (Pagliarini et al., 2018).

Comme nous l'avons précisé, la meilleure méthode pour obtenir une imitation quasi-parfaite du chant chez le Diamant Mandarin consiste à élever un jeune avec un tuteur adulte. Cette situation n'est pas naturelle car un jeune diamant mandarin est en général exposé aux chants d'autres individus conspécifiques, jeunes et adultes. Mais cette procédure constitue un standard car elle permet de tester les effets d'une lésion (ex : pinéalectomie ; Derégnaucourt et al., 2012) ou encore d'une transfection génique (Haesler et al., 2007) sur l'apprentissage du chant chez un petit nombre d'individus en maintenant un degré de puissance statistique suffisant. A ce jour, les méthodes utilisées en laboratoire ne permettent pas de garantir une imitation parfaite du chant et une forte variabilité inter-individuelle est observée (Derégnaucourt et al., 2013). Le développement d'un robot qui serait utilisé comme tuteur de chant permettrait d'étudier les aspects multimodaux de l'apprentissage vocal et éventuellement de reproduire ce résultat observé lorsqu'un jeune est élevé seul avec un adulte : une imitation quasi-parfaite du chant.

Au total, 36 diamants mandarins mâles ont été utilisés pour ces expériences. Ils sont tous issus de la population captive hébergée au sein du Laboratoire Ethologie Cognition Développement de l'Université Paris Nanterre. En dehors des expériences, les diamants mandarins vivent dans des volières mixtes. Ils sont placés en couple dans des batteries pour la reproduction. Enfin, ils sont hébergés en cage individuelle dans des caissons insonorisés pour la passation des expériences (Figure 14).



Figure 14. Dispositif expérimental pour l'expérience d'apprentissage du chant chez le Diamant Mandarin. Le caisson insonorisé contient la cage expérimentale dans laquelle le jeune oiseau et le robot diamant mandarin interagissent. Ils sont filmés par deux caméras placées à droite et à gauche et l'activité vocale est enregistrée en permanence grâce à un système d'enregistrement automatisé (logiciel Sound Analysis Pro).

Enfin, nous présentons les caractéristiques les plus saillantes des deux modèles robotiques (le robot humanoïde NAO et le robot-diamant mandarin MANDABOT) que nous avons utilisés, une description plus complète étant faite dans les articles dédiés.

4. Modèles robotiques

4.1. Le robot humanoïde NAO

La première version du robot humanoïde NAO a été créée en 2005 par la société française Aldebaran Robotics. C'est le robot le plus utilisé en recherche et en institution : il est en effet accessible facilement à un prix abordable, son apparence de jouet avec son visage rond et ses grands yeux le rendent attractif, il est robuste d'un point de vue technologique et facilement programmable (Belpaeme et al., 2018). Pour nos études, nous avons fait l'acquisition

au laboratoire du robot NAO robot version 6. Ses 25 degrés de liberté lui permettent une grande mobilité. Il est équipé de deux caméras ainsi que de deux haut-parleurs et de 4 microphones qui lui confèrent des capacités de reconnaissance vocale et de synthèse vocale. De plus, ses capteurs tactiles situés au niveau de sa tête, ses mains, sa poitrine et ses pieds lui permettent de réagir au contact. Il peut être utilisé à la fois sur un mode autonome et télé-opéré à distance. Ci-dessous (Figure 15) quelques exemples des capacités et des domaines d'applications du robot NAO.



Figure 15. Le robot NAO soutient les enfants hospitalisés (projet Alize[20]), joue au football avec d'autres robots[21] et peut animer des cours de gymnastique[22]. Il sait aussi dessiner[23], il aide les enfants à l'école, participe à des spectacles de danse[24], assure la téléprésence des enfants hospitalisés (projet AvatarKid[25]), a de très bonnes capacités d'équilibre[26] , rend visite à des usagers en EHPAD[27].

[20] https://fp7alizeproject.wordpress.com/photos/#jp-carousel-107
[21] https://www.clubic.com/sante/actualite-844922-coupe-monde-football-robotique-tiendra-bordeaux.html
[22] https://www.villa-sully.fr/2015/06/10/maisons-de-retraite-a-chaque-region-son-prix/
[23] https://www.intuitive-robots.com/fr/a-propos-intuitive-robots/
[24] https://www.researchgate.net/figure/Dancer-Gael-Rougegrez-of-Blanca-Li-Dance-Company-performs-with-a-NAO-robot-at-Barbican_fig15_336312417
[25] https://www.sciencesetavenir.fr/high-tech/robot/avatar-kids-le-robot-nao-a-l-hopital-pour-garder-le-lien-entre-copains-de-classe_114315
[26] https://www.prepose.ch/fr/2017/05/11/robot-mon-ami/
[27] https://lejournal.cnrs.fr/billets/rire-avec-les-robots-pour-mieux-vivre-avec

Pour nos expériences, le robot a été entièrement pré-programmé par Arnaud Blanchard, du Laboratoire Equipe Traitement de l'Information et Systèmes de l'Université de Cergy-Pontoise. Il était télé-opéré à l'aide d'une tablette tactile pendant les expériences.

4.2. Le robot diamant-mandarin « MANDABOT »

Ce robot a été conçu sur mesure en collaboration avec Philippe Gauthier et Florian Richer de l'Institut des Systèmes Intelligents et de Robotique de Sorbonne Université. Guglielmo Montone du Laboratoire de Psychologie de la Perception de l'Université Paris Descartes, a développé une boucle d'interactions vocales oiseau-machine qui a été implémentée au robot. Ainsi lorsqu'un oiseau placé en présence du robot vocalise, le robot émet un cri en réponse. MANDABOT est un robot animaloïde qui ressemble physiquement à un diamant mandarin mâle adulte (Figure 16).

Dans l'abdomen du robot, un haut-parleur permet de diffuser des cris et des chants préenregistrés. Un système de boucle d'interaction vocale a également été implanté dans le robot : MANDABOT peut effectuer différents mouvements comme tourner la tête et bouger son corps autour d'un axe rotatif. Deux perches équipées de capteurs ont été positionnées de part et d'autre du robot. Lorsque l'oiseau se pose sur la perche, le robot se tourne vers lui et fait un mouvement de tête ou produit un chant. C'est le mode contingent qui permet d'obtenir à la fin de l'interaction un script de toutes les vocalisations et de tous les mouvements produits par le robot (contrôle en boucle fermée). Le robot peut par la suite rejouer le script à l'identique, c'est le mode non-contingent (contrôle en boucle ouverte) puisque le robot suit le script enregistré au lieu de réagir en temps réel au comportement de l'oiseau. Une description plus détaillée du robot est proposée au lecteur au Chapitre III : il s'agit d'un article scientifique soumis dans un journal international avec comité de lecture.



Figure 16. Le robot-diamant mandarin MANDABOT utilisé pour l'expérience d'apprentissage du chant. Il a la taille et l'apparence d'un mâle mandarin adulte.

5. Objectifs de la thèse et hypothèses de travail

Le but de cette thèse était d'explorer les différences et les similitudes de comportement ainsi que les performances d'apprentissage des individus selon que leur partenaire d'interaction était un robot ou un individu de la même espèce. Nous avons, pour cela, employé une approche pluridisciplinaire à la fois en éthologie, en psychologie du développement et en psychomotricité.

Les Chapitres I et II présentent les résultats obtenus respectivement chez les enfants DT et les enfants TSA. Nos hypothèses étaient les suivantes :

➔ Chez les enfants DT, les performances sont comparables dans le groupe robot et dans le groupe humain et s'améliorent avec l'âge, dans les deux groupes. De plus, les enfants regardent autant la démonstratrice robotique que la démonstratrice humaine.

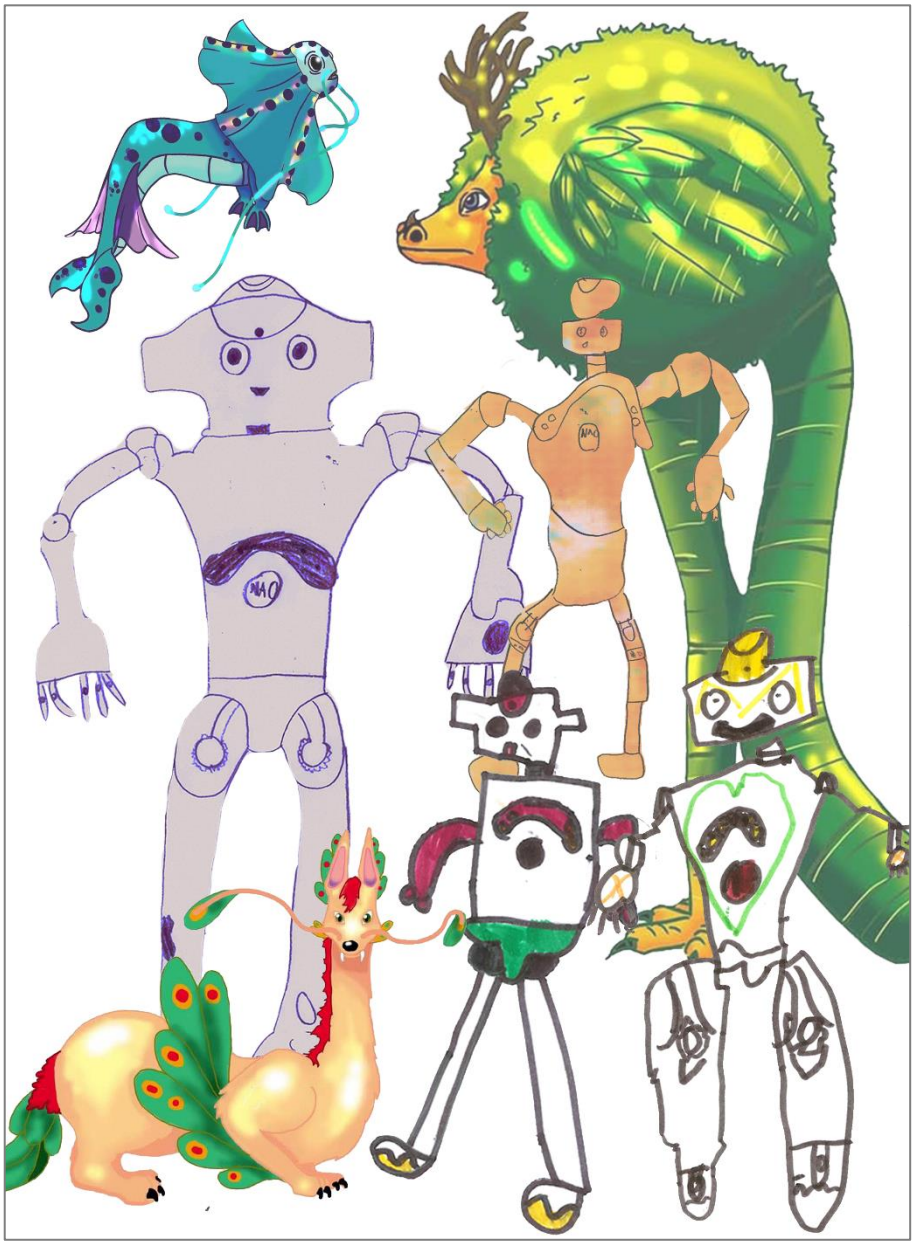
- ➔ Chez les enfants TSA, les performances sont comparables dans la condition robot et dans la condition humaine. En revanche, les enfants regardent plus la démonstratrice robotique que la démonstratrice humaine.

Dans le Chapitre III, nous présentons les caractéristiques du robot diamant-mandarin MANDABOT et, dans le Chapitre IV, les résultats d'une expérience visant à tester son efficacité comme tuteur de chant. Comme mentionné dans la partie §4.2., nous avons la possibilité de tester le robot en mode contingent ou en mode non contingent. Dans le premier cas, MANDABOT répond de manière contingente aux comportements du jeune (ex : il produit un cri lorsque le jeune produit un cri) alors que dans le second cas, les actions du robot sont produites de manière aléatoire par rapport aux comportements produits par le jeune oiseau.
Nos hypothèses étaient les suivantes :

- ➔ Les jeunes exposés au MANDABOT programmé en mode contingent (mouvements, production sonore) imitent aussi bien le chant que des jeunes exposés à un diamant mandarin adulte non familier.
- ➔ Les jeunes exposés au MANDABOT programmé en mode non contingent imitent moins bien le chant que des jeunes exposés à un robot en mode contingent ou à un diamant mandarin adulte non familier.

CHAPITRE I

Child-robot interactions in a learning context, an evaluation of body knowledge and a measure of prosociality: comparison with a human partner



Le robot NAO dessiné par les enfants à l'école
& les images des animaux imaginaires utilisées dans l'expérience.

Child-robot interactions in a learning context, an evaluation of body knowledge and a measure of prosociality: comparison with a human partner

Alice Araguas [1*], Arnaud Blanchard [2], Adrien Chopin[3], Sébastien Derégnaucourt [1, 4] and Bahia Guellaï[1,4]

[1] *Laboratoire Ethologie Cognition Développement, Université Paris Nanterre, 200 Avenue de la République, F92001 Nanterre Cedex, France*

[2] *Laboratoire Equipe Traitement de l'Information et Systèmes, Université Cergy-Pontoise, F95000 Pontoise, France*

[3] *INSERM, CNRS, Institut de la Vision, Sorbonne Université, F75012, Paris, France*

[4] *Institut Universitaire de France*

* Correspondence: Alice Araguas, Laboratoire Ethologie Cognition Développement, UPL, Univ Paris Nanterre, F92000 Nanterre, France. E-mail address: alice.araguas@parisnanterre.fr

Abstract

The aim of our study was to compare 3- to 6-years old child-robot interactions to child-human ones in different domains: a vocabulary learning task, the evaluation of the child body awareness (including body parts labelling and emotions identification) and a situation of spontaneous help. We also evaluated children's representation of the demonstrator (human or robotic) and children gaze directions during the experiment. We found no effect of the demonstrator type on the vocabulary learning and the body awareness tasks except for the respect of body parts order in the imitation sequence. We found that children looked more at the robotic demonstrator only during the subtask of emotional key poses identification. Regarding effect of age, we found that older children obtained higher scores than younger ones in questionnaires about their representations of the demonstrator, in the vocabulary task, in labelling body parts on the demonstrator's body, in the whole imitation task, in the recall of emotions and the identification of emotional key poses. Moreover, compared to younger children, older ones looked more at the robotic demonstrator than at the human demonstrator during the identification of emotional key poses.

Keywords: child-robot interaction, vocabulary learning, body parts comprehension, emotional body postures, prosocial behaviour

1. Introduction

Early in their life, children use their social learning capacities to acquire crucial skills such as language (known as “vocal learning abilities”) and various social behaviours such as initiating a conversation in any subject, following the rules, helping, demanding help or information, awaiting their turn and collaborating (Aksoy & Baran, 2010). Children exhibit different developmental trajectories in acquiring these social skills and they also show preferences concerning tutors’ characteristics such as gender (Perloff, 1982), age (Perloff, 1982), familiarity (Olson & Spelke, 2008), perceived degree of expertise (Zmyj et al., 2010) or language (Kinzler et al., 2007).

In the past few years, the number of technological supports for learning increased significantly and exist under various forms: computers, tablets, smartphones and, more recently, robots. Social robots are specifically designed to interact and communicate with people, either semi autonomously or autonomously (i.e., with or without a person controlling the robot in real time) (van de Berghe et al., 2019), following behavioural norms that are typical for human interaction (Bartneck & Forlizzi, 2004). The robot’s main purpose is to engage in an affective or helpful interaction directly with the person (Sheridan, 2020). Social robots are used for educational purposes and in human-robot interactions research projects. In this perspective, one challenge is to determine whether a social robot can be considered as a social partner and, in addition, as a relevant learning tutor for young children. Interest for the effectiveness of robots in teaching tasks in comparison to human teachers is growing. Even though robots are typically not developed to replace human teachers, comparisons between robot and human teachers or peers are useful to investigate areas in which robots can complement humans (van den Berghe 2019). So far, there is a large heterogeneity concerning thematic areas, experimental protocols and types of robots used for this purpose (Belpaeme et al., 2018; van den Berghe et al., 2019

for reviews). However, across studies using social robots as teachers, the humanoid robot NAO is the most frequently used (in 48% of 89 articles reviewed, Belpaeme et al., 2018).

We conducted a developmental study with children aged from 3 to 6 years old and compared their reactions and performances in several tasks with a robotic or a human demonstrator. For half of the children, the demonstrator was the NAO robot, to which we attributed a feminine gender (i.e., robot group) and for the others, the demonstrator was an unfamiliar woman (i.e., human group). We proposed three situations to children as described in following parts.

The field of robot-assisted language learning (RALL) which encompasses vocabulary, reading skills, speaking skills, grammar, and sign language to learn first - and second- language learning is growing over the years (van den Berghe et al., 2019). In the first task of our study, we partially replicated the experimental protocol of Kory Westlund and colleagues (2017) who found that a robot tutor was as performant as a human one for pre-schooler children to learn new vocabulary in the form of names of unknown animals in their native language.

The second experimental task we designed was an awareness body task including a) the evaluation of the comprehension of body parts labels, b) an imitation task and c) an emotions' identification and expression task. As body parts knowledge is mostly tested through using imitation task of a robot and is restricted to children with Autism Spectrum Disorder (Costa et al., 2014; Suzuki & Lee, 2016; Suzuki et al., 2017), we decided to test, in a first part, production of body parts labels on the child's body and on the demonstrator's body apart following the procedure described by Bergès & Lézine (1963). In a second part, we evaluated body parts labels production and imitation of designated body parts sequences. The following sub-tasks we designed concerned the identification of emotions in different supports (images and the demonstrator body postures). In two studies, Beck and colleagues (2010; 2013) explored the identification of the NAO robot's emotional key poses - defined as static body postures - by

adults and 11 to 13 years old children. They concluded that the lack of facial expressions was not a barrier to recognize emotional key poses. In the present study, we decided to test the most recognized NAO key poses (corresponding to anger, sadness, fear, happiness) with younger children.

The third task concerned prosocial behaviours which refer to a large range of voluntary behaviours that benefit to other individuals (Dunfield & Kuhlmeier, 2013) and appear around 2 years old (Warneken & Tomasello, 2006). In our study, we focused on instrumental need which, for Dunfield and colleagues (2019), occurs when an individual is unable to complete a goal directed behaviour. As proposed by these authors, we used the term of "spontaneous helping" to refer to "other-oriented acts aimed at alleviating another's instrumental need" (Dunfield et al., 2019). We tested whether children will help the demonstrator reaching a stick that felt after having interacted with her during 20 minutes. We also questioned children about their representations of the demonstrator using the same pre and post-test questionnaire.

We analysed children's scores in each experimental task, their answers in questionnaires and the time they spent looking at the demonstrator. Our main hypothesis was that a social robot such as the NAO robot is as effective as a human demonstrator to give instructions and to be understood by 3 to 6 years old children. We expected an effect of age in these different tasks: we hypothesis that younger children will obtain lower scores than older ones, in both conditions (robot and human demonstrator).

2. Materials and methods

2.1. Participants

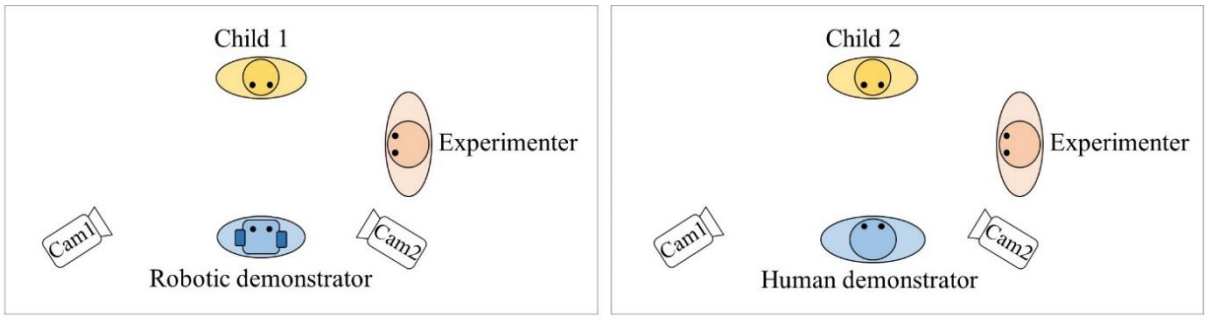
Data was collected between October 2020 and May 2021. Seventy-three children aged 3 years and 3 months to 6 years and 1 month were tested. All children were French native

speakers and came from 4 different schools. We obtained parental consent for all children to take part in the study and to be recorded.

2.2. Experimental procedure

2.2.1. General setup

Experimental sessions took place at school, in a quiet room in the presence of the experimenter and the demonstrator (Figure 1). Four different experimenters, all adult females, one per school, were recruited and trained to follow the experimental set up. The experimenter asked the child to sit on the floor in front of the demonstrator and introduced the demonstrator to him/her. The demonstrator was a female adult for half of the children and the NAO robot for the other half. She gave instructions to the child. Due to the Covid19, the experimenter and the human demonstrator wore face masks during the whole experiments. Each individual session lasted approximately 20 min. Two Sony HDR-CX410VE cameras placed on tripods from either side of the demonstrator were used to record the child's behaviours.



(A) Robot group (B) Human group

Figure 1. Top view of the experimental setup for the two groups: (A) with the robotic demonstrator and (B) with the human demonstrator (inspired by Kennedy et al., 2015). The child and the demonstrator sat in front of each other and the experimenter was on the right side of the demonstrator.

2.2.2. Human group: exposure to an adult demonstrator

The human demonstrator was always the same adult female and remained as neutral as possible during the experiment: she gave the instructions to the child but she did not encourage him/her or give him/her feedbacks.

2.2.3. Robot group: exposure to the NAO robot demonstrator

Created by Aldebaran Robotics, the NAO robot version 6 is 58 cm high and bipedal. It has 25 degrees of freedom which allows a lot of possibilities of movements. It can manipulate small objects with its three fingers - hands. As presented in Figure 2, NAO has two video cameras on its forehead and its mouth. It has vocal capacities both for recognition and synthesis: it is equipped with a stereo broadcast system made up of 2 loudspeakers, on its ears and 4 omnidirectional microphones (2 on the top of its head and 2 on the back of its head). It also has two ultrasonic sensors (or sonars) which allows it to estimate the distance to obstacles in its environment. Moreover, it has contact and tactile sensors: tactile head and hands, chest button and feet bumpers.

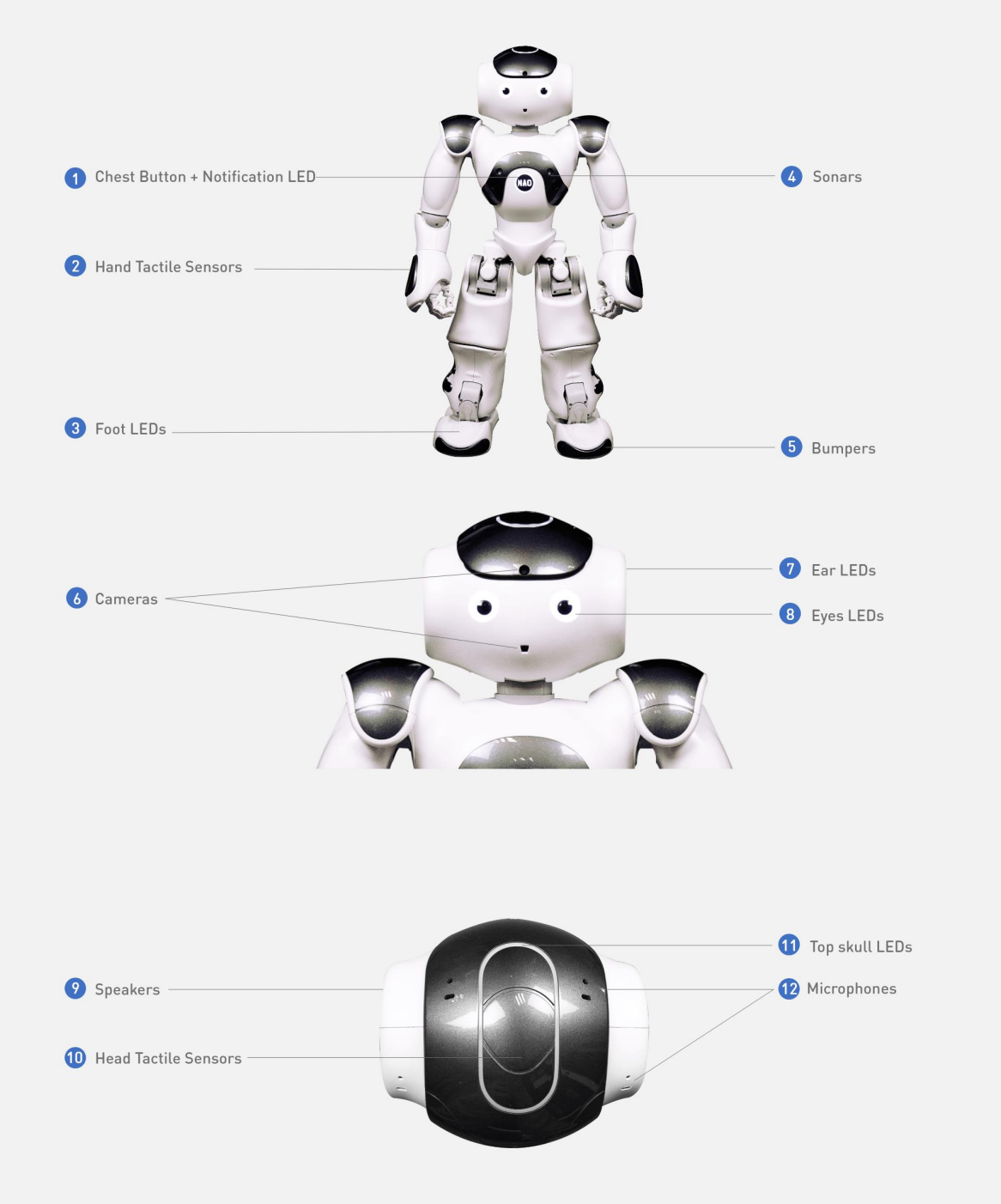


Figure 2. Overview of the humanoid NAO6 robot from SoftBank Robotics' website (source: https://www.softbankrobotics.com/).

The NAO robot can be programmed to carry out autonomously a set of tasks. However, for our experiment, the robot was fully tele-operated to enable the robot to act contingently

(Kennedy et al., 2017): the experimenter controlled the robot during the session, via a touch-screen tablet, an iPad mini 4 (20.32 x 13.48 x 0.61 cm). As shown in Figure 3, the setup consisted on connecting the tablet on the wifi access point of the robot. The robot ran a web server that can be accessed with any browser on the tablet. An HTML page displays the interface and executed the sequence of the experiment in javascript (« vue.js » for the graphical interface, « LibQi » for controlling the robot).

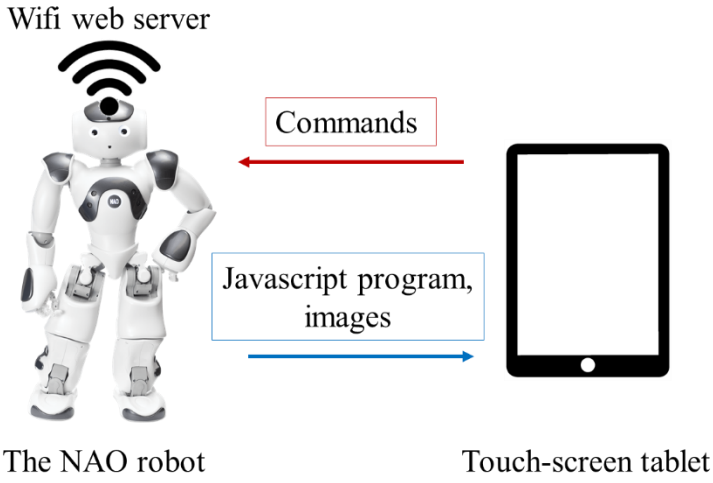


Figure 3. The robot is controlled by a touch-screen tablet.

As the human demonstrator was a female, we attributed a feminine gender to the NAO robot and a feminine name (“Naomie”). We will use the pronoun “her” in the following text.

To ensure that the two demonstrators pronounced the same sentences in both conditions, vocal recordings of the female adult were played back by the robot. Sentences were recorded while she was alone in a silent room and as she was addressing instructions to a child.

On 73 children initially recruited, 9 asked to stop before the experiment was over, whilst in the two other cases, there were technical issues with the camera or the tablet. In total, 62 child-demonstrator interactions were completed in a between-subject procedure (school A: 7 children, school B: 10 children, school C: 17 children, school D: 28 children). Children were separated into three groups of age: Group 1, “G1”: from 3 years old and one month to 4 years

old and one month (mean ± SE = 3.73 ± 0.07); Group 2, “G2”: from 4 yo and two months to 5 yo and 1 month (mean ± SE = 4.71 ± 0.06) ; Group 3, “G3”: from 5 yo and and two months to 6 yo and one month (mean ± SE = 5.65 ± 0.059). As much as possible, we balanced experimental groups concerning the age and gender of the children as presented in Table 1.

	Robot group	Human group	Total number
Group 1: 3 yo 1 month to 4 yo 1 month			
Girls	4	4	*8*
Boys	2	4	*6*
Group 2: 4 yo 2 months to 5 yo 1 month			
Girls	7	6	*13*
Boys	6	7	*13*
Group 3: 5 yo 2 months to 6 yo 1 month			
Girls	6	8	*14*
Boys	5	3	*8*
Total number	*30*	*32*	*62*

Table 1. Number of children in each experimental group regarding their age and gender.

2.2.4. Experimental tasks and scores

To help the reader, we wrote the script of the experiment in Supplementary Information (Appendix 1). The experiment started with a questionnaire and a warm-up phase.

2.2.4.1. Questionnaires about child’s representations of the demonstrator

At the beginning of the session, the experimenter asked the child about his/her representations of the demonstrator, either the human or the robot (see Supplementary

Information, Appendix 1). At the end of the experiment, the same questions were asked to evaluate whether the interactions with the demonstrator had an influence on the child's answers.

2.2.4.2. Warm-up phase

Before starting the experimental tasks, there was a warm-up phase to allow the child to see how the demonstrator, robotic or human depending on the group, spoke and moved before the beginning of the experiment. The demonstrator said her name to the child and asked the child to introduce herself/himself. She sang a popular French rhyme with hand gestures and encouraged the child to follow it.

Then, each child was tested in three different situations: 1) a vocabulary task, 2) a body task including three parts: 2.a) comprehension of body parts labels on their own body and on the demonstrator's body task, 2.b) an imitation task and 2)c. an emotions' identification and expression task, and 3) a spontaneous helping situation and 3) a spontaneous helping situation. Tasks 1 and 2 were randomly presented across children. The experiment always finished with task 3 as we supposed that previous interactions with the demonstrator influenced the propensity of the child to help the demonstrator.

2.2.4.3. Vocabulary task

The demonstrator was seated in front of the child, on the ground. The experimenter placed the touch-screen tablet between the child and the demonstrator in which appeared images of imaginary animals (see Supplementary Information, Figure 4). We created these four imaginary animals to ensure that children had not seen them before. They were designed and hand drew by a painter and then reproduced on Adobe Photoshop CC2014, version 15.0.

Presentation phase:

One at the time and for 10 sec each, images of the four imaginary animals appeared in a random order on the tablet.

Naming phase:

Then, the demonstrator warned the child that she/he was going to learn their names. On the four images, only two were named (called "target images"). Imaginary animals were randomly associated to two non-words: "dax" as in the article of Kory Westlund and al. (2017) and "grid" inspired from examples in the article of Havy and colleagues (2013).

The animal appeared and the demonstrator said "*Look at this animal. It is a grid. It is a grid*" then asked the child "*What is the name of this animal?*". If the child answered "*grid*", he/she obtained 1 point. If he/she did not answer, he had 0 point.

A second animal appeared and the demonstrator said "*Look at this animal. It is a dax. It is a dax. What is the name of this animal?*". If the child answered "*dax*", he/she obtained 1 point. If he/she did not answer, he had 0 point.

Test phases:

For the two test phases, images were presented by pair side-by-side on the tablet. First, a target image and a non-target image appeared on the screen. The demonstrator said "*Touch the dax*". When the child touched the screen of the tablet, the two images disappeared and the screen went blank. Then, two other images appeared, the other target image and a non-target image and the demonstrator said "*Touch the grid*".

Finally, after another blank screen, the two target images were presented side-by-side and the demonstrator asked the child to "*Touch the grid*". When the child touched the screen, the pair disappeared, then the pair of target images appeared again and the demonstrator asked "*Touch the dax*".

The child obtained 1 point for each target image correctly identified (4 points in total).

Response times, i.e. latencies between the apparition of the images on the screen and the touch of the child, were also recorded by the tablet and were accessible on a specific web page using the screen touch tablet.

At the end of the task, the experimenter took back the tablet and told the child that he/she will participate to another game.

2.2.4.4. Body task

Comprehension of body parts labels:

We evaluated the comprehension of eleven body parts labels: face, eye, nose, mouth, ear, shoulder, elbow, hand, belly, knee, foot. Each body part was randomly chosen in this list. First, the demonstrator asked the child "*Show me your [body part]*" and repeated it for the eleven possibilities. Then she asked "*Show me my [body part]*" and also repeated it for the eleven body parts.

The child obtained 1 point for each body part correctly shown (in total, 11 points for identification on his body and 11 points for identification of the demonstrator's body).

Imitation of body parts' sequences:

In a second phase, we evaluated body parts labels production: the demonstrator randomly showed one of her body part, for example, her eye and asked "*What is the name of this body part?*" and a second one, for example her belly, and said "*And this one?*". We scored 1 point if the child correctly identified the body part showed by the demonstrator.

As it was complicated for the NAO robot to point her nose and its ear, we decided to keep 9 body parts for this sub-task (face, eye, mouth, shoulder, elbow, hand, belly, knee and foot in Supplementary Information, Figure 5).

Then, the demonstrator placed her hands on her knees and asked the child to repeat the body parts sequence. We gave 1 point if the child showed the correct body part and respected the order of the body part in the sequence.

The demonstrator performed two sequences with two body parts ("sequence#1", "sequence#2") and a third sequence with three body parts ("sequence#3"). As the majority of

children had difficulties to understand the instructions in this task, the first two-body parts sequence (sequence#1) was used by the experimenter to explain the instructions again in order to help the child to understand it.

Emotions task:

Four images of a child expressing an emotion (joy, sadness, anger, fear. Supplementary Information, Figure 6) were placed in front of the child. The demonstrator asked "*Do you know what an emotion is? Show me the image in which the child feels [emotion]*". If the child showed the correct image, he/she obtained 1 point (4 points in total).

Then, the demonstrator and the child stood up. The demonstrator asked "*Show me how do you express [emotion] with your whole body"*. She waited for the child to express the emotion. The instruction was repeated with the four emotions. If the child expressed the emotion with his/her whole body, he/she obtained 1 point (4 points in total). If he/she expressed the emotion only with his/her face, he/she obtained 0.5 point. If he/she did not move or described verbally his/her behaviour when he/she felt the emotion, he/she obtained 0 point.

As the NAO robot does not have facial expressions, we investigated emotions recognition when emotions were expressed with the body. We selected the four key poses exhibited by the NAO robot that were the most successfully identified by adults and children in previous studies (Beck et al., 2010; Beck et al., 2013): joy, fear, anger and sadness. The demonstrator struck each of these four key poses and between each pose, she went back to a neutral pose (Supplementary Information, Figure 7).

For this task, the demonstrator said "*Let me show you what I am doing when I am feeling an emotion*". The demonstrator reproduced one of the four emotional poses and the experimenter asked the child "*Which emotion do you think it is?*". After the child replied, the demonstrator returned to the neutral pose. She repeated this sequence for the three other body postures. If the child identified the correct posture, he/she obtained 1 point (4 points in total).

2.2.4.5. Spontaneous helping situation

The demonstrator and the child were seated in front of each other. The experimenter gave a blue stick to the demonstrator while saying "*Look, I have something for you*". The demonstrator grabbed it, looked at it and let it fall on the ground, expressing her surprise and saying "*Oh, it fell!*" Then, she tried to reach the stick with her hand. We measured the latency between the stick falling and the first time the child touched the stick to give it back to the demonstrator. If the child did not touched the stick after 10 s, he/she obtained the score of 11 (corresponding to 11 s).

At the end the session, the demonstrator thanked the child for his/her participation and told him/her goodbye.

2.2.5. Adaptations of the experimental protocol

As 7 children from G1 (aged from 3 yo and one month to 4 yo and one month) asked to stop the experiment and were excluded from the final sample, we decided to adapt the experimental protocol in order to reduce the duration of the experiment (Table 2).

	Experiment with children	
Experimental tasks	Groups 3 and 2	Group 1
Pre-test questionnaire	17 questions	**7 questions**
Warm-up phase	Yes	Yes
Vocabulary task	Yes	**No**
Body task	Yes	Yes
Spontaneous helping situation	Yes	Yes
Post-test questionnaire	17 questions	**7 questions**

Table 2. Comparison of experimental protocols regarding the age groups of the children.

2.3. Behavioural analysis

2.3.1. Video recordings and analysis

All sessions were video recorded. Cameras allowed to have a global view of the setup which encompassed the face and the body of the child. We coded the direction of child's gazes during the whole experiment.

2.3.2. Ethogram

We analysed child's behaviours when the demonstrator was talking to him/her using the software BORIS v.7.9 (Supplementary Information, Table 3).

2.4. Statistical analysis

We conducted models trying to predict each variable with the demonstrator type (robot or human), the child's age, the child's age in interaction with the demonstrator's type and the child's gender as predictors. We started each model including school and experimenter as random effects of a GLME. We adapted the link and distributions to increase the model goodness-of-fit using AIC and % of variance explained (squared R) as indicators. We removed high p-values factors when it increased the goodness-of-fit, and switched to a GLM (without random effects), following the same rationale (Table 4).

		Model	Distribution	Link	Predictors	Random effects
Vocabulary task	*Scores*	GLM	Gamma	Log	Demonstrator type Child's age Child's age x demonstrator type Child's gender	No
	Latencies of response	GLME		Reciprocal		School Experimenter
Body task	*Comprehension BPL**	GLM		Log		No
	*Production BPL**	GLM		Identity		No
	Imitation of sequences	GLME		Reciprocal		School Experimenter
	Respect of order in sequences	GLME		Identity		School Experimenter
Emotions task	*Identification on images*	GLM		Log		No
	Expressions of emotions	GLME		Reciprocal		School Experimenter
	Citation of emotions	GLM		Identity		No
	Identification emotional key poses	GLME		Reciprocal		School Experimenter
Helping situation	*Latencies of helping*	GLME		Reciprocal		School Experimenter
Other behaviours	*Gaze times during the whole experiment*	GLME	Normal	Identity		School Experimenter
	Gaze times emotional key poses task	GLM	Gamma	Reciprocal		No
	Number of tactile contacts	GLME	Gamma	Reciprocal		School Experimenter

Table 4. Models used for each task. (*BPL refers to body parts labels).

We detail below data in which we conducted specific analysis.

2.4.1. Imitation of body parts sequences

We decided to remove sequence#1 from the statistical analysis as it was necessary to use it to explain again instructions to children. We kept the second two-body parts sequence (sequence#2) and the three-body parts sequence only (sequence#3). As results in sequence 2 and 3 followed the same patterns, we calculated a number of correct answers adding answers from sequence 2 and 3 and we conducted GLM on these data.

2.4.2.Gaze time task by task

We conducted Mann-Whitney tests to test whereas there was a difference between the human and the robot group concerning the time spent looking at the demonstrator task by task. We found a significant difference in time spent looking at the demonstrator in the task of identification of the demonstrator's key poses. Mean percentage of time spent looking at the demonstrator was higher in the robot group (78.56% ± 2.12 (mean ± SE)) than in the human group (64.93% ± 2.07, mean ± SE, Mann-Whitney test: $W = 298$, $p = 0.01$).

Then, we conducted a GLME to predict if the gaze time in the key poses sub task was influenced by the demonstrator, the age of the child as fixed effects and the experimenter as a random effect.

2.4.3. Pre and post-test questionnaires

The number of correct answers summed among all items (for G2 and G3 only) was not normally distributed (Shapiro-Wilk test: $W = 0.87$, $p < 0.001$). Inverting the data (number of errors + 1) gave an appropriate gamma distribution. The data were not different between pre- and post-tests (Wilcoxon signed-rank test: $W = 444$, $p = 0.27$). Therefore, we averaged the data from pre- and post- together. We conducted a GLM implementing a gamma distribution and a reciprocal link, with demonstrator type (robot or human), the child's age and their interaction as predictors. The number of correct answers summed among the 7 items presented to G1 only

(for all age groups) was not normally distributed (Shapiro-Wilk test: W = 0.92, $p < 0.001$). Inverting the data (number of errors + 1) gave an appropriate gamma distribution. We implemented a GLM with a gamma distribution and a reciprocal link, using the child's age, demonstrator type and their interaction as predictors.

We also compared the number of "yes" answers, item by item, to compare children's representations in the human and robot groups. Data were not normally distributed (Shapiro-Wilk tests: $p < 0.05$). We used exact Fisher tests to test whether there was a difference between the two groups.

3. Results

3.1. Experimental tasks and scores

For each task, we presented means and standard errors for G1, G2 and G3 and for human and robot group in Supplementary Information, Tables 5 to 10.

3.1.1. Questionnaire about child's representations of the demonstrator

3.1.1.1. Mean number (average between pre and post-tests) of correct answers in the questionnaire

Correct answers in the questionnaire are presented in Supplementary Information, Table 11. We found that categorisation was clearer for the human demonstrator than for the robot demonstrator: standard errors were higher in the robot group which means that answers were more homogeneous for the human group (Figure 8). We also observed higher inter individual variability for younger children (G1).

As shown in Figure 8, in G2 and G3, for the human group, items "*can hurt someone*", "*know what you think*", "*made by someone*" presented the highest standard errors. In the robot group items "*object*", "*eat*", "*grow*", "*happy*", "*toy*", "*think*", "*feel tickles*", "*can be broken*" presented the highest standard errors.

All children of G3 identified that the human demonstrator was a human, was happy and can see things and that she was not an object neither an animal, a machine or a toy. In the robot group, the only question that obtained 100% of correct answer was "*Is the robot an animal?*".

In G2, all children said that the human demonstrator is not a machine. In G1, all children said the robot is not an animal and that the human can grow.

Using a GLM, we found that the number of correct answers (G2 and G3) increased with age ($t(44) = 2.73$, $p < 0.001$) but not with the demonstrator type ($t(44) = 1.62$, $p = 0.11$) although there was a trend for the effect of the demonstrator to be stronger with increasing age (interaction $t(44) = -1.93$, $p = 0.06$).

When focusing on the 7 items presented to the children of the G1 (all groups included), the number of errors decreased with age ($t(58) = 3.01$, $p = 0.003$), and this effect of age was stronger when the demonstrator was a human rather than a robot (interaction age x demonstrator type; $t(58) = -2.08$, $p = 0.04$; demonstrator type: $t(48) = 1.67$, $p = 0.10$).

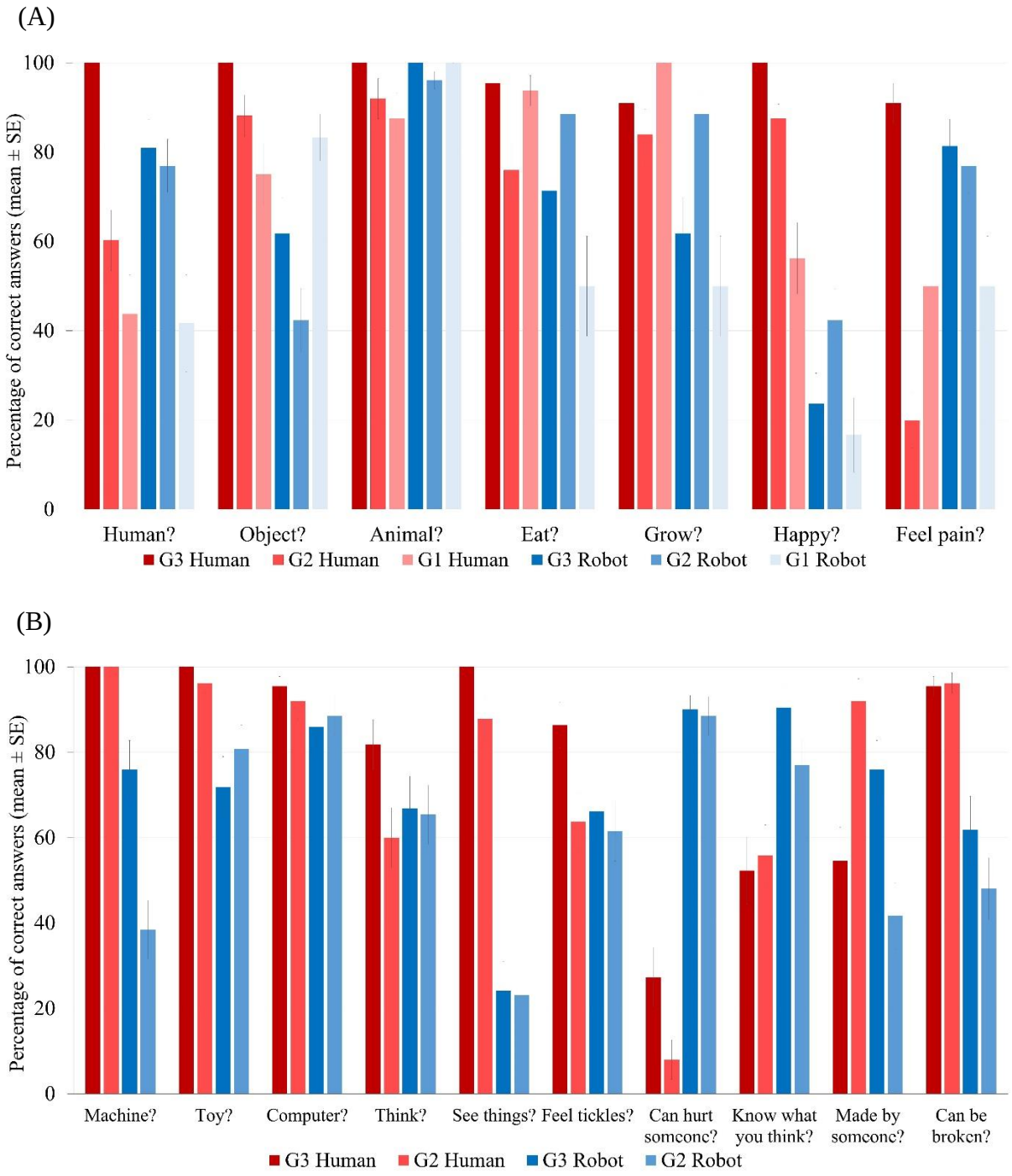


Figure 8. Percentage of correct answers (mean ± SE) for each item of the questionnaire, for human and robot groups and for each group of age: (A) questions common to G1, G2 and G3 and (B) questions specific to G2 and G3.

3.1.1.2. Number of “yes” answers in the questionnaire in pre-test

We used the number of responses “yes” for each item (for example “Is the demonstrator a human?) to compare children’s representations about human and robot demonstrators.

In groups G2 and G3, as presented in Table 12, we found no difference in the number of “yes” for items “animal”, “toy”, “computer”, “the demonstrator can feel pain”, “the demonstrator can hurt someone”.

Moreover, children considered the human demonstrator more as a “human” than the robot demonstrator. More children also considered that the human demonstrator can “eat”, “grow up”, ”think”, be “happy”, “see things”, “feel tickles” and that she knows what the child “think” compared to the robotic demonstrator.

On the contrary, children considered that the robot demonstrator was more like an “object” and like a “machine” than the human demonstrator. More children also considered that it was “made by someone” and it “can be broken” compared to the human demonstrator.

Representations of children in the G1 were slightly different and the demonstrator’s abilities seemed less clear. We found the same number of “yes” answers for both demonstrators, human and robot, concerning items “human”, “animal”, “can eat”, “can grow up”, “can be happy”, “can feel pain” whereas the robot demonstrator was more considered as an “object” than the human demonstrator (Table 13).

G2 and G3	*p* value in an exact Fisher test
Human?	**< 0.001**
Object?	**< 0.001**
Machine?	**< 0.001**
Animal?	1
Toy?	0.09
Computer?	0.35
Eat?	**< 0.001**
Grow?	**< 0.001**
Think?	**0.003**
Happy?	**0.02**
See things?	**< 0.001**
Feel tickles?	**< 0.001**
Feel pain?	0.08
Hurt someone?	0.35
Know what you think?	**< 0.001**
Made by someone?	**0.02**
Can be broken?	**< 0.001**

Table 12. Exact Fisher tests' *p* values when comparing the number of "yes" answers for each item of the questionnaire, for both human and robot group and for G2 and G3.

G1	*p* value in an exact Fisher test
Human?	1
Object?	**0.02**
Animal?	1
Eat?	0.26
Grow?	0.07
Happy?	1
Feel pain?	1

Table 13. Exact Fisher tests' *p* values when comparing the number of "yes" answers for each item of the questionnaire, for both human and robot group and for G1.

3.1.2. Vocabulary task

Means of correct answers were high in both groups (human group: 88.89% ± 2.17 and robot group: 90.28% ± 2.35 (mean ± SE), Figure 9). The number of errors significantly decreased with age ($t(43) = -4.35$, $p < 0.001$) whereas it did not vary with the type of demonstrator ($t(43) = 0.13$, $p = 0.89$), the gender (($t(43) = 1.31$, $p = 0.19$) or the interaction age x demonstrator: $t(43) = -0.13$, $p = 0.89$).

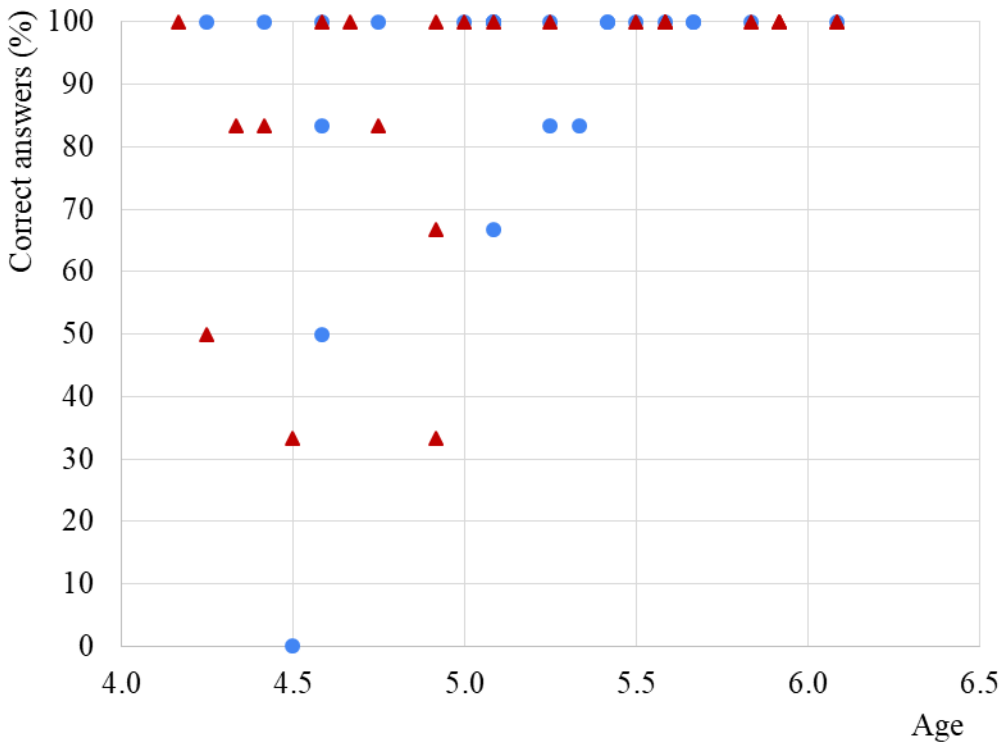


Figure 9. Distribution of mean percentage of correct answers in the vocabulary task depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

Mean reaction times to touch the tablet for the four answers were similar between the two groups: 4.46 s ± 0.21 (mean ± SE) for the human group and 4.22 ± 0.18 (mean ± SE) for the robot group (Figure 10). We attempted to predict latencies with a GLME and found no effect of age ($t(43) = 1.27$, $p = 0.87$), demonstrator type ($t(43) = 0.51$, $p = 0.61$), gender ($t(43) = -0.24$, $p = 0.81$) and no interaction between age and demonstrator type ($t(43) = -0.46$, $p = 0.65$).

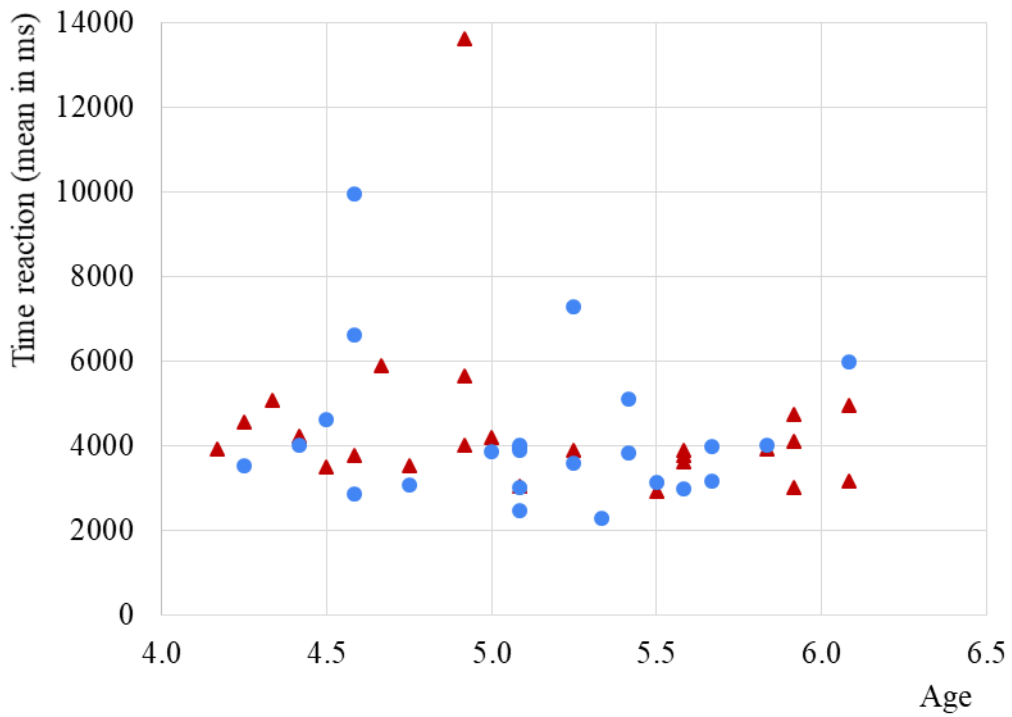


Figure 10. Distribution of mean reaction time in the vocabulary task depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

3.1.3. Body task

3.1.3.1. Comprehension of body parts labels

Identification of body parts on the child's body was very high in both groups (human group: 94.03% ± 0.63 and robot group: 91.52% ± 1.90, mean ± SE, Figure 11A) as identification on the demonstrator's body (human group: 87.78% ± 2.31 and robot group: 91.82% ± 1.80 (mean ± SE), Figure 11B).

The number of correct body identifications on the child's body did not vary with our predictors (age ($t(57) = -1.22$, $p = 0.22$), type of demonstrator ($t(57) = 1.79$, $p = 0.07$), gender (($t(57) = 0.01$, $p = 0.98$), interaction age x demonstrator: $t(57) = -1.75$, $p = 0.08$).

The number of correct body identifications on the demonstrator's body increased with age (($t(57) = -2.68$, $p = 0.001$) but it did not vary with other predictors (type of demonstrator

(t(57) = 0.37, p = 0.71), gender ((t(57) = 0.41, p = 0.68), interaction age x demonstrator: t(57) =-0.40, p = 0.69).

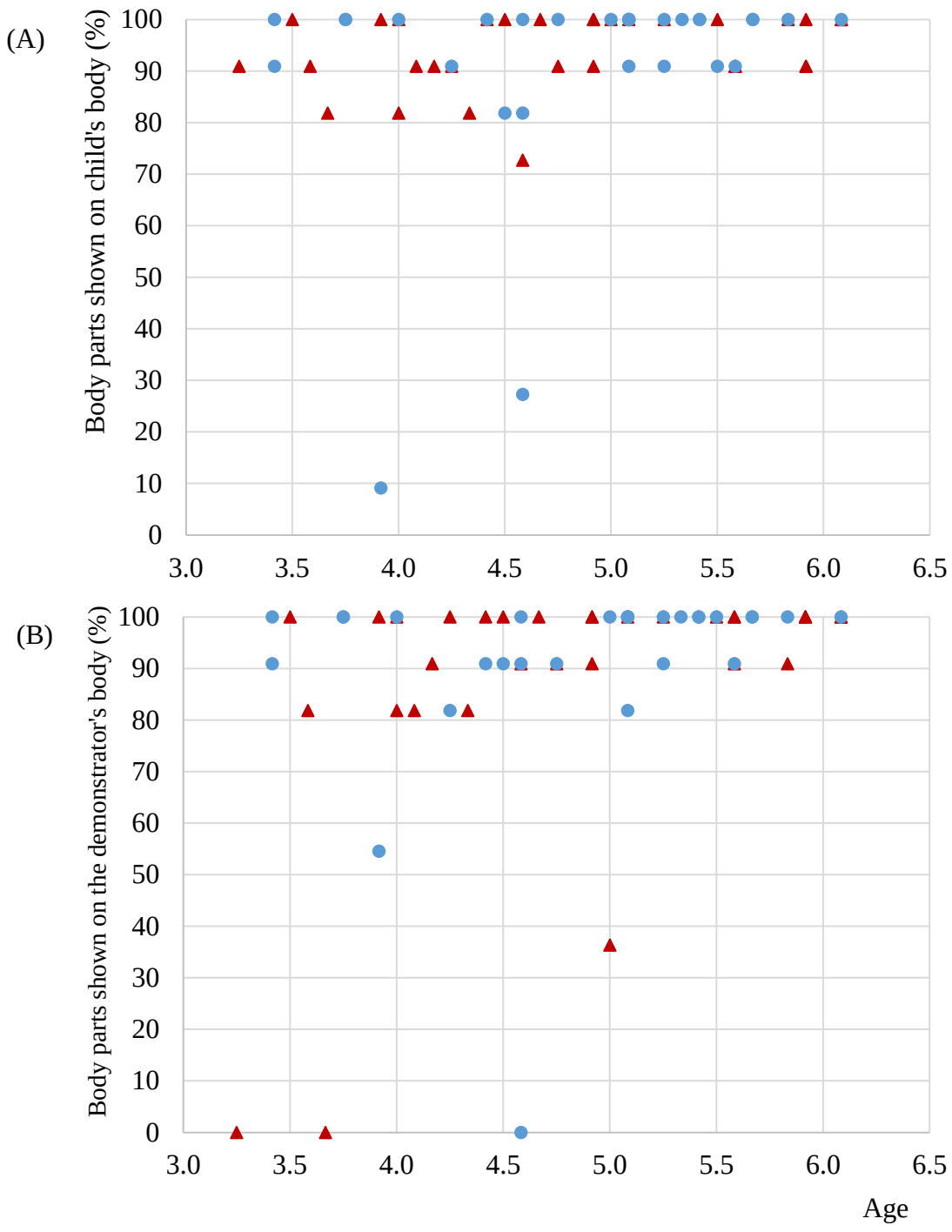


Figure 11. Distribution of correct answers (in %) in the body parts' task depending on the age of the child for the robot group (blue circle) and the human group (red triangle) on: (A) the child's body and (B) on the on the demonstrator's body.

Imitation of body parts sequences:

Results of the imitation task were separated into 3 parts (labels of body parts production, imitation, respect of the order).

The mean percentage of body parts shown by the demonstrator and correctly named by children, i.e. correctly labelled, is higher in the human group than in the robot group (human group: 83.69% ± 4.02 and robot group: 69.03% ± 3.68, mean ± SE, Figure 12).

The number of body parts correctly labelled, significantly increased with age ($t(57) = -3.13$, $p < 0.001$) whereas it did not vary with the type of demonstrator ($t(57) = 0.35$, $p = 0.72$), the gender ($t(57) = 1.69$, $p = 0.09$) or the interaction age x demonstrator ($t(57) = -0.04$, $p = 0.96$).

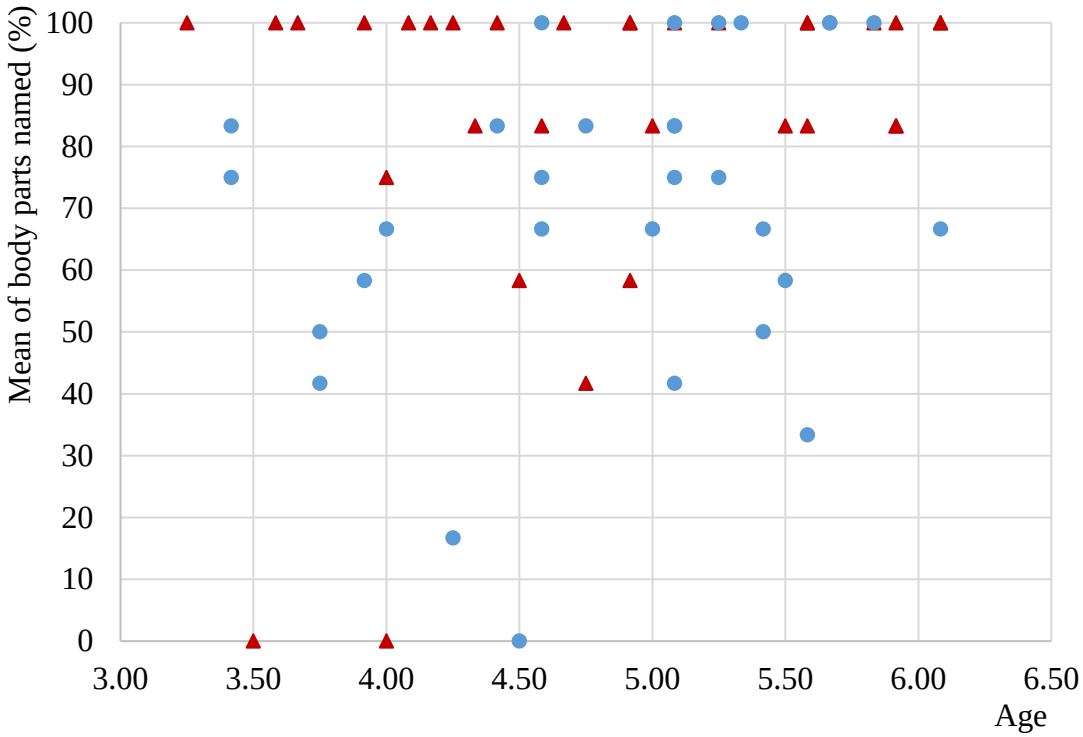


Figure 12. Distribution of mean correct answers (in %) in naming body parts depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

The mean percentage of body parts correctly imitated is higher in the human group than in the robot group (human group: 65.29% ± 4.42 and robot group: 46.20% ± 4.70, mean ± SE, Figure 13).

The number of body parts correctly imitated in sequences by children significantly increased with age ($t(57) = 3.78$, $p < 0.001$) whereas it did not vary with the type of demonstrator ($t(57) = 1.70$, $p = 0.09$), the gender (($t(57) = -0.41$, $p = 0.68$) but the effect of age was stronger for the human demonstrator than for the robot demonstrator (interaction $t(57) = -2.06$, $p = 0.04$).

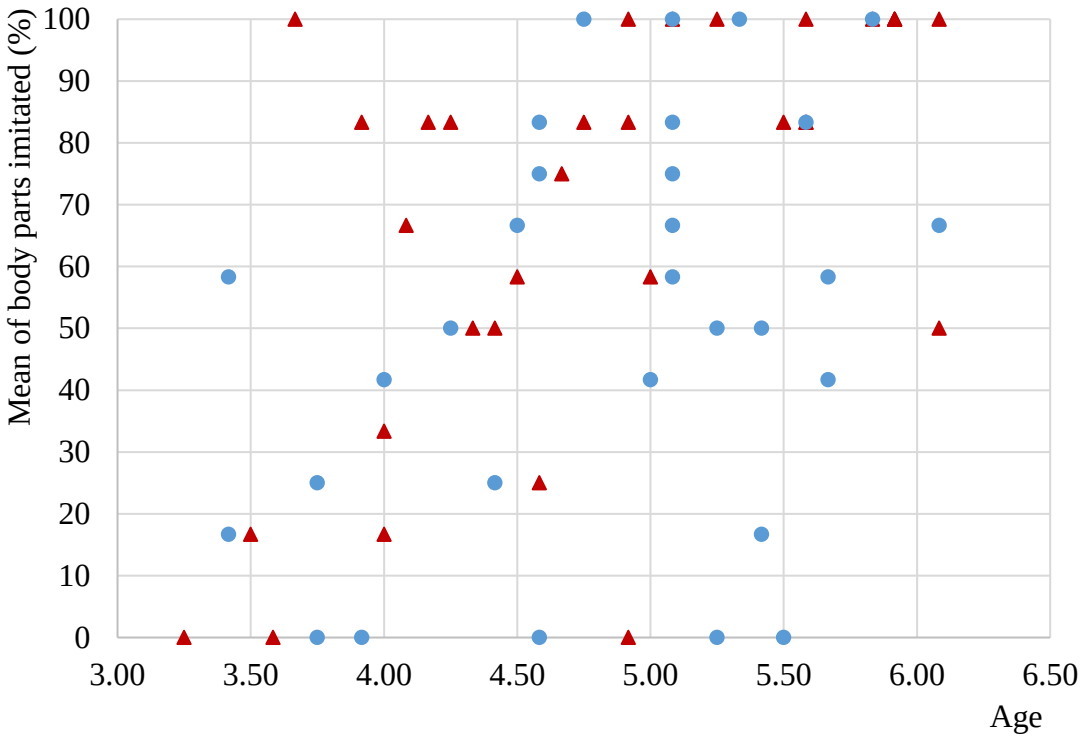


Figure 13. Distribution of mean correct answers (in %) in imitating body parts depending on the age of the child for the robot group (blue circle) and the human group (red triangle)

The mean percentage of body parts correctly imitated is also higher in the human group than in the robot group (human group: 49.60% ± 5.49 and robot group: 29.56% ± 3.80, mean ± SE, Figure 14).

The proportion of correct responses in the right order significantly increased with the child's age ($t(57) = 7.53$, $p < 0.01$), increased when demonstrator was human ($t(57) = -5.29$, $p < 0.01$), when the child's gender was male ($t(57) = -6.66$, , $p < 0.001$) and the increase with age was stronger when the demonstrator was a robot (interaction age x demonstrator type: $t(57) = 4.52$, $p < 0.01$).

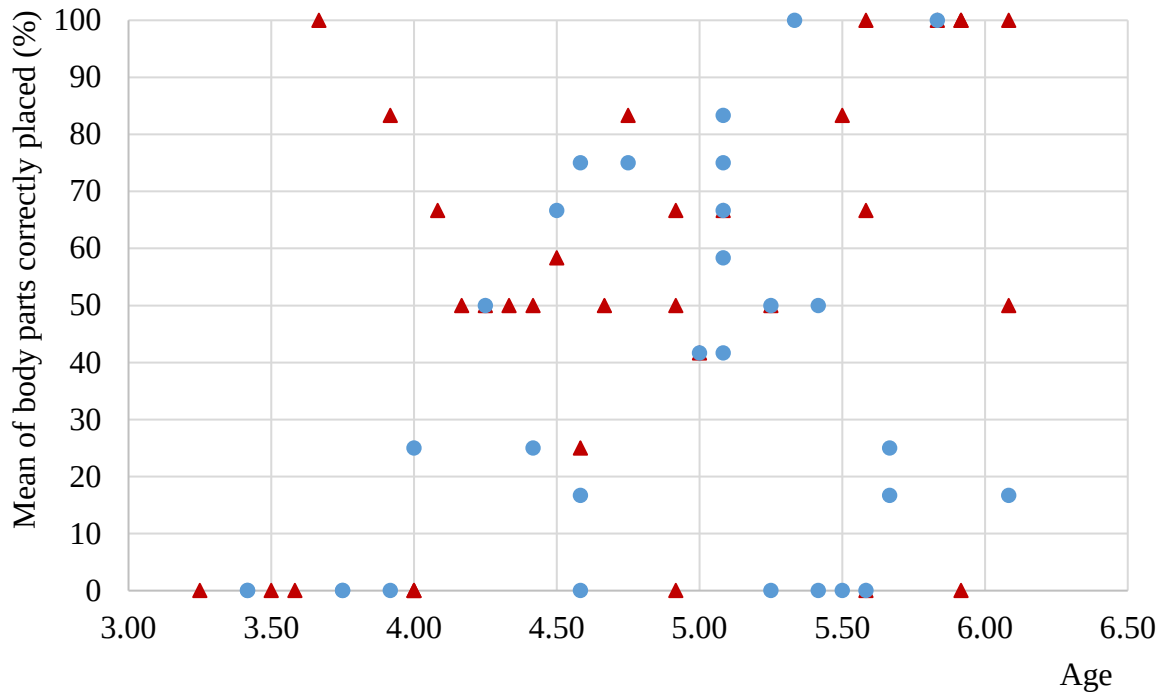


Figure 14. Distribution of mean body parts correctly placed in the sequence of imitation (in %) depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

3.1.3.2. Identification of emotions

On average, 94.53% ± 1.34 of images of emotions were correctly identified by children of the human group and 96.67% ± 0.99 of images of emotions by children of the robot group (Figure 15).

The number of correct emotions identified did not vary with our predictors (age (t(57) =-0.05, p = 0.95, type of demonstrator (t(57) = 1.38, p = 0.17), gender ((t(57) = -0.03, p = 0.97), interaction age x demonstrator: t(57) =-1.57, p = 0.12).

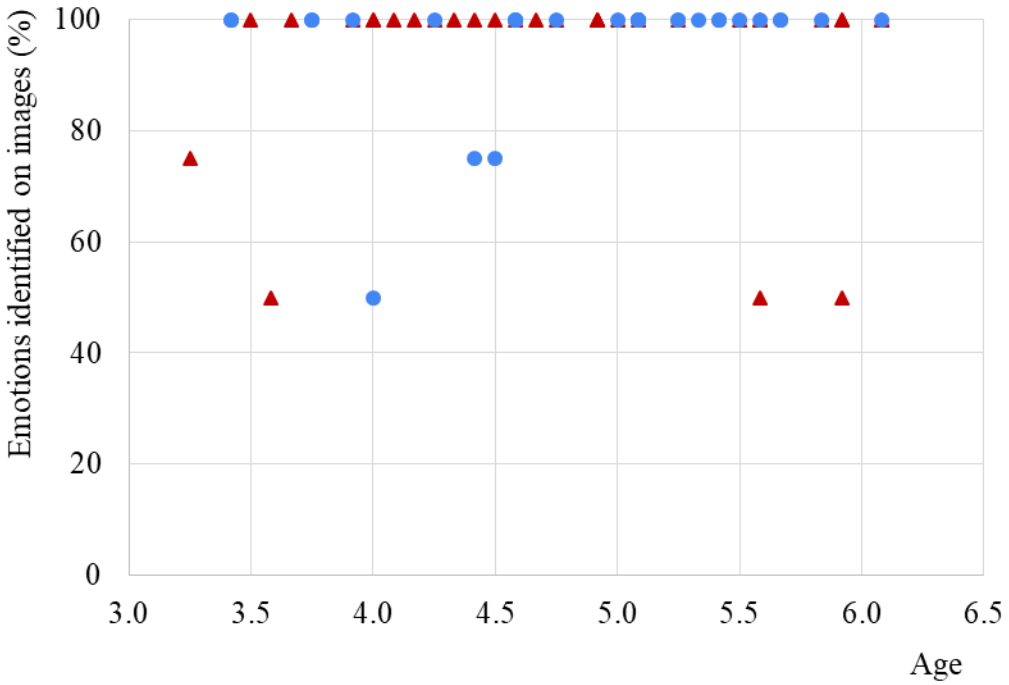


Figure 15. Distribution of the percentage of emotions correctly identified on images depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

50.39% ± 4.04 of emotions were expressed with the whole child's body in the human group and 48.74% ± 3.57 in the robot group (Figure 16). We observed a huge individual variability. The number of emotions expressed did not vary with our predictors (age ((t(57) = 1.63, p = 0.10, type of demonstrator (t(57) = 0.61, p = 0.54), gender ((t(57) = 1.22, p = 0.22), interaction age x demonstrator: t(57) =-0.64, p = 0.51).

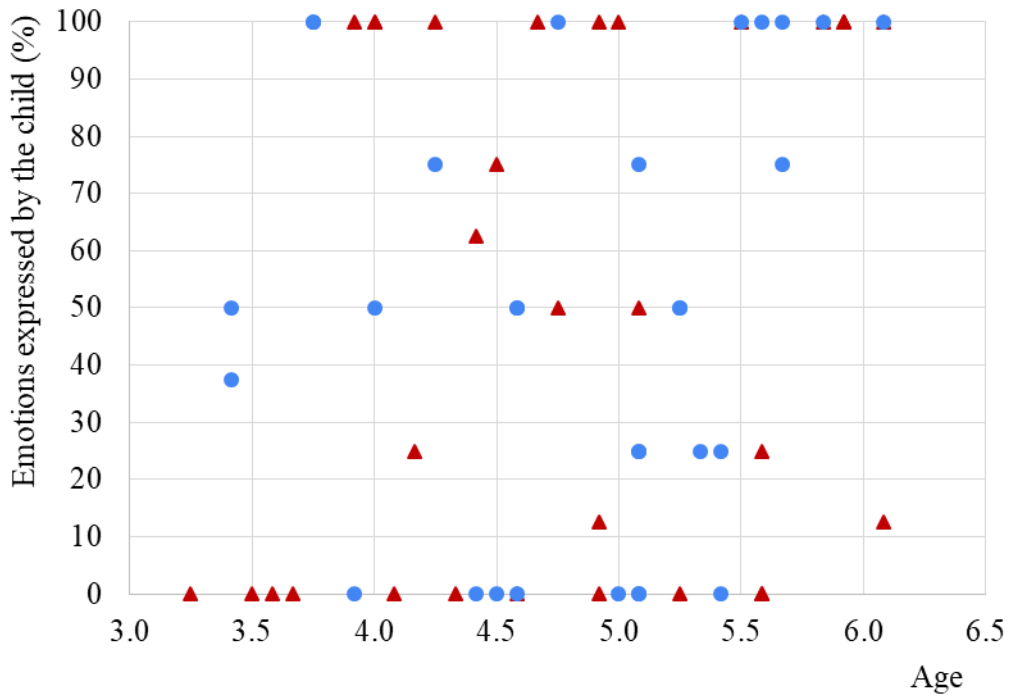


Figure 16. Distribution of the percentage of emotions expressed by the child depending on his age for the robot group (blue circle) and the human group (red triangle).

On average, 50.83% ± 4.26 (mean ± SE) emotions were correctly cited by children of the human group and 71.67% ± 4.05 by children of the robot group (Figure 17). We observed an important effect of children's ages: older children cited significantly more emotions than younger ones ((t(55) = -3.82, , $p < 0.001$). This score was not influenced by the type of demonstrator (t(55) = -1.55, $p = 0.12$), the gender ((t(55) = 1.30, $p = 0.19$) or the interaction age x demonstrator: t(55) =1.52, $p = 0.13$).

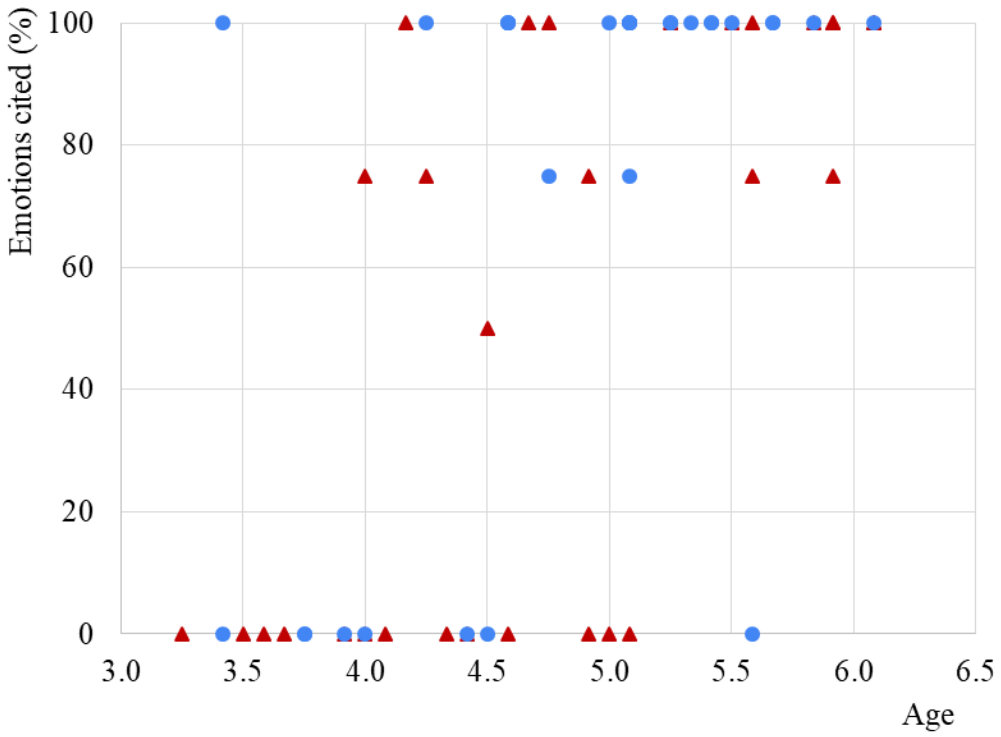


Figure 17. Distribution of the percentage of emotions correctly cited depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

On average, 35.94% ± 2.44 (mean ± SE) of emotional key poses were correctly identified by children of the human group and 38.33% ± 2.30 by children of the robot group (Figure 18). We observed an effect of children's ages: the mean percentage of emotional key poses recognized by younger children was lower compared to older ones ((t(57) = -2.76, p = 0.007) but not with other predictors as type of demonstrator (t(57) = -1.07, p = 0.28), gender ((t(57) = 0.36, p = 0.17), interaction age x demonstrator: t(57) =1.02, p = 0.31).

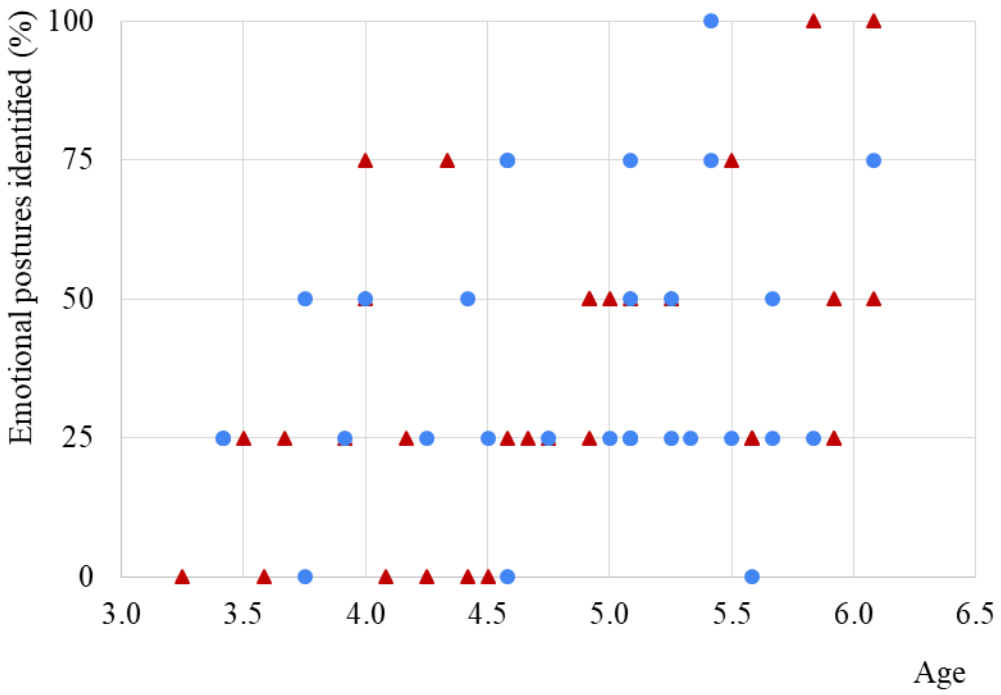


Figure 18. Distribution of the percentage of emotional key poses correctly identified depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

3.1.4. Spontaneous helping situation

On average, children reacted faster to give back the stick to the robot demonstrator than to the human demonstrator (robot group: 5.30 s ± 0.40 (mean ± SE), human group: 7.43 s ± 0.41 (mean ± SE). On the 30 children of the robot group, 15 touched the stick less than 5 s after it felt while 9 did not touch the stick after 10 s. In the 32 children of the human group, 11 touched the stick less than 5 s after it felt while 19 did not touch the stick after 10 s (Figure 19).

The help time did not vary with our predictors (age ((t(57) = 0.51, p = 0.60), type of demonstrator (t(57) = 0.56, p = 0.57), gender ((t(57) = 0.24, p = 0.80), interaction age x demonstrator: t(57) = -0.29, p = 0.77).

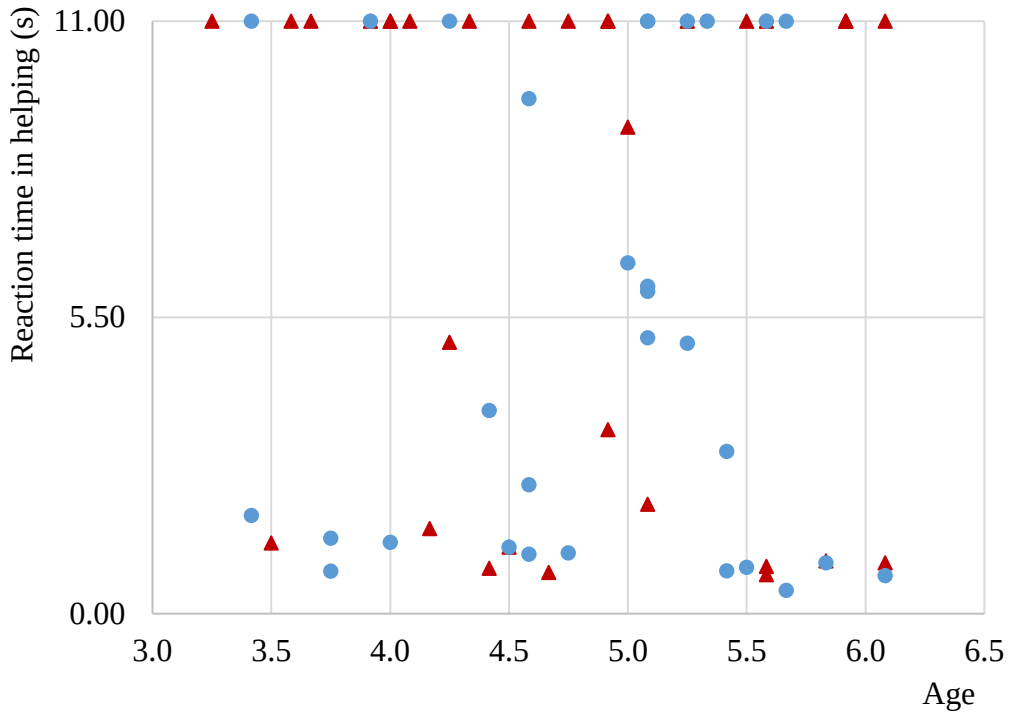


Figure 19. Distribution of the reaction times in helping (in second) depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

3.2. Gaze times and other behaviours

3.2.1. Time spent looking at the demonstrator during the whole session

We calculated the mean percentage of time spent looking at the demonstrator for the whole session, for each child. Using a GLME, the average time spent gazing at the demonstrator was not predicted by age ($t(55) = 1.72$, $p = 0.09$), the demonstrator type ($t(55) = 0.687$, $p = 0.50$), gender ($t(55) = -0.67$, $p = 0.50$) or the interaction age x demonstrator ($t(55) = -0.43$, $p = 0.67$) (Figure 20).

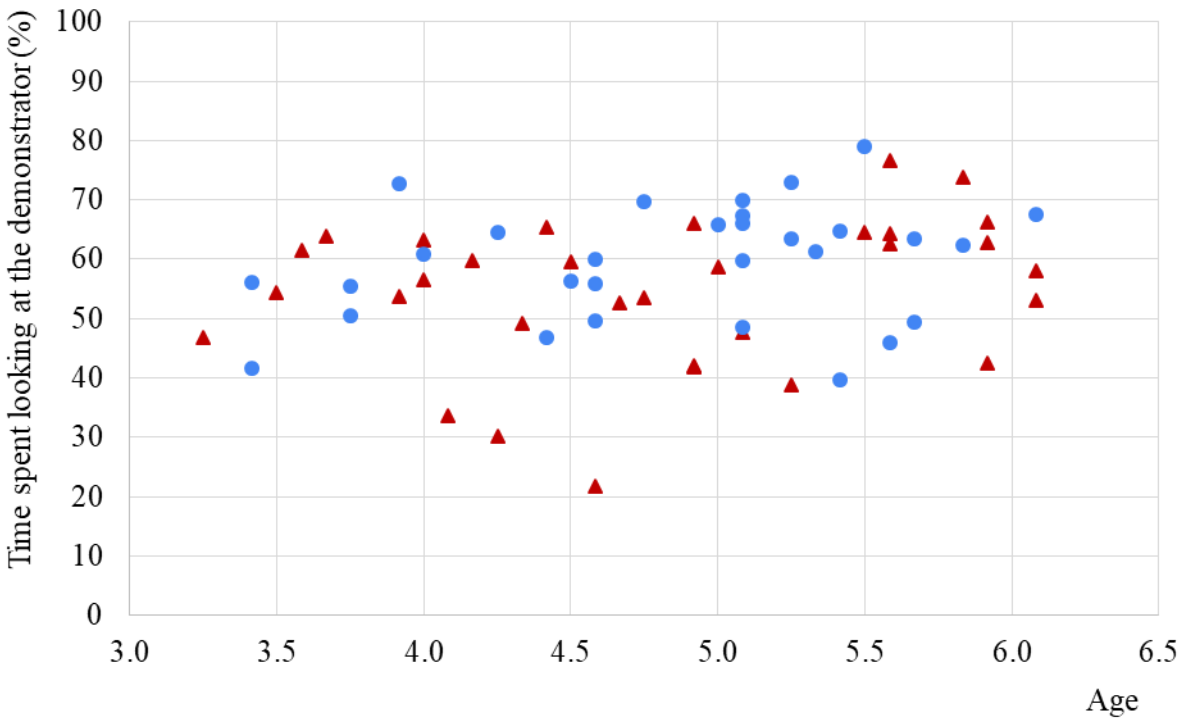


Figure 20. Distribution of the percentage of time spent looking the demonstrator depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

3.2.2. Time spent looking at the demonstrator during the key poses sub task

We found a significant difference in time spent looking at the demonstrator in the sub task of identification of the demonstrator's key poses. Mean percentage of time spent looking at the demonstrator was higher in the robot group (Mann-Whitney test: W = 298, p = 0.01).

Using a GLME, the mean percentage of time spent gazing at the demonstrator for that task increased significantly with age (t(59) = 2.29, p = 0.02) and was greater with the robot demonstrator than the human demonstrator (t(59) = 2.15, p = 0.03).

3.2.3. Direction of gaze when the demonstrator questioned the child

When children answered questions asked by the demonstrator, they looked more at the experimenter in the robot group than in the human group than in the human group (Mann-Whitney test: W = 259.5, $p < 0.001$).

3.2.4. Number of times the child touched the demonstrator

The number of times the child touched the demonstrator varied following an age x demonstrator type interaction (t(59) = -4.13, , $p < 0.001$): the number of tactile contacts increased more with age for the robot demonstrator group than for the human demonstrator group. Gender had no significant influence (t(59) = -1.24, $p = 0.22$) (Figure 21).

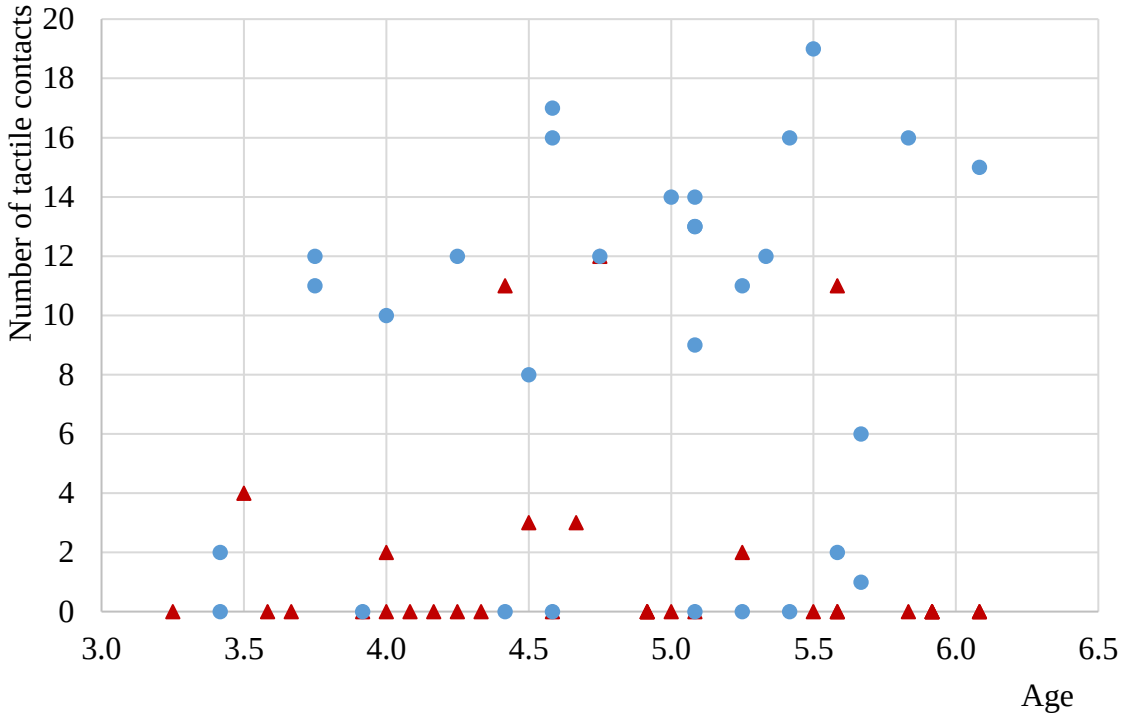


Figure 21. Distribution of the percentage of the number of times the child touched the demonstrator depending on the age of the child for the robot group (blue circle) and the human group (red triangle).

4. Discussion

The originality of our study was to compare, at different ages, children's reactions to instructions given the demonstrator type. We evidenced that children from 3 to 6 years old answered similarly to a NAO robot as they did to an unfamiliar adult. Indeed, we did not find any significant influence of the demonstrator on mean scores of correct answers. As we expected, the number of correct answers increased with age for most of the tasks we performed: in the vocabulary task, in the identification of body parts on the demonstrator's body, in all sub tasks of imitation (the number of body parts labels named by the child, the number of body part imitated and the respect of order). In the identification of the emotion task, the subtask consisting of emotions citation and the identification of emotional key poses were best succeeded by older children compared to younger ones. Finally, children had better representations of the human and the robot demonstrators with age.

The vocabulary task was completed by children from G2 (4 years old and 2 months to 5 years old and 1 month) and G3 (5 years old and 2 months to 6 years old and 1 month) only. We observed a ceiling effect for group 3 were most of the children correctly named and recognised target animals. The mean latency response in the vocabulary task was not influenced by any predictors. As in Kory Westlund and colleagues' study (2017), we found that human and robot are as efficient tutors to elicit short term memorisation of novel words in 3 to 6 years old children. Further research should be conducted to evaluate the possibility of long-term memorisation as in the study of de Wit and colleagues (2018): they found that 4 to 6 years old children remembered names of animals in a second language and taught by the NAO robot one week after the experiment.

Identification of body parts labels on the child's body was not influenced by any of the predictors and we observed a ceiling effect. Camões-Costa and colleagues (2011) found that comprehensive repertoires of children from 2,2 years old to 3,4 years old contained between 15 and 33 body parts (median = 27) on the 50 stimuli presented. The most correctly located body parts were: tongue, foot, head, back, tummy, mouth, belly, eye, teeth, hand, leg, fingers, arm, ear, bottom, hair, neck, knee, thumb, cheek, shoulder, toes, lips, eyebrow, elbow. There was a positive correlation between children's comprehension and their age. Then, we can make the hypothesis that by the age of 3 years old, body parts evaluated in our experiment were already known by most of the children, which explains the ceiling effect in our study.

However, we found that identification of body parts labels on the demonstrator's body increased with children's age. Pointing to body parts of another person is different from pointing body parts on his own body. This is in line with studies which evidenced that patients with autotopoagnosia (impairment at pointing to parts of their own body named by an experimenter) do not necessarily present heterotopoagnosia (impairment at pointing to parts of their own body named by an experimenter) (Pisella et al., 2019). Indeed, spatial cognition theories differentiate two reference frames: representation used by the brain to convey the layout of points in space (Klatzky, 1998) related to two distincts neural circuits. While the egocentric frame conveys positions of points using the body as the centre of the surrounding space, the allocentric frame concerns the positions of points using external landmarks. If both frames of reference are present in 5 years old children (Newcombe & Huttenlocher, 2003), further studies showed a gradual development from an egocentric to a more stable map-like representation between 5 and 10 years of age (Bullens et al., 2010). It can explain the improvement with age of the ability to locate body parts on the demonstrators' body we found in our study.

In the imitation task, we evaluated the ability to produce labels of body parts. We found that older children gave more correct answers than younger ones. It is concordant with previous studies highlighting that this ability develops later in development compared to the comprehension of body parts labels (Camões-Costa et al., 2011). Concerning the imitation of body parts sequences, we also found an effect of age and this effect was stronger for the human demonstrator than for the robot demonstrator. We hypothesis that imitation of pointing body parts was easier as the human demonstrator had a more familiar body for the children than the robot body. Finally, the proportion of correct responses in the right order significantly increased with the child's age increased when the demonstrator was human, when the child's gender was male and the increase with age was stronger when the demonstrator was a robot.

Regarding the identification of emotions in the body postures task, we first assessed emotions labelling on images. We observed a ceiling effect: scores were very high and not influenced by the predictors. The ceiling effect is concordant with the study of Widen & Russel (2003) which demonstrated that facial expressions are correctly associated with labels ("happiness", "sadness", "anger", "fear", "surprise" and "disgust", the six primary emotions described by Ekman & Friesen, 1976) by 3 years old children. In another study (Visser et al., 2008), the ability to recognize sadness, anger, joy and fear on very simple drawings of human faces is acquire between 4 and 5 years old.

Concerning emotions expressed by the child, we found no effect of the predictors and a huge interindividual variability. We hypothesis that more than the demonstrator type (human or robot), scores were influenced by the emotional expressiveness of children – the capacity to accurately communicate feelings nonverbally - and the tendency to feel comfortable in the presence of unfamiliar adults, variables that we did not measured. Indeed, Riggio & Riggio (2002) showed in their meta-analysis that more emotional expressive people also have a more extraverted personality (extraverts are outgoing, talkative, impulsive and uninhibited and have

many social contacts (Eysenck & Eysenck, 1968)). Then, the recall task where children had to cite the 4 emotions was largely influenced by age.

Finally, we assessed labelling of emotions in the demonstrator key poses. We found no effect of the demonstrator type but an improvement with children's ages. Correct labels of the demonstrator emotional key poses were lower than in the study of Beck and colleagues in adults (2010) and children from 11 to 13 years old (2013). We made two assumptions about the effect of children's age. First, Izard (1971) pointed out that emotion labelling implies a more complex cognitive process than emotion recognition, which can be fully mature later in the child's development. Secondly, even if Beck and colleagues (2010, 2013) concluded that emotional key poses can be accurately recognised without any context, Li & Chignell (2011) demonstrated that knowledge about the social and environmental context in which the key poses of a teddy bear robot are expressed improve the success of identification. We hypothesize that the context might be more important for younger children than for older children and adults as they are still learning to recognise emotions.

In the spontaneous helping situation, we found no effect of predictors but a tendency to reach the stick faster in the robot group than in the human group. Children might have understood that the robot had limited motor capacities and that she could not reach the stick whereas children knew that the human demonstrator could easily reach out her hand to grab the stick.

We also analyzed other behaviours such as gaze times, tactile contacts and we evaluated children representations of the two demonstrators.

Gaze times during the whole experiment did not vary with the predictors but in the emotional key poses task, children spent more time looking at the robot demonstrator, as if they needed more time to identify emotions expressed by the robot.

When children answered questions asked by the demonstrator, they looked more at the experimenter (the woman present during the experiment in both conditions) in the robot group than in the human group which supports the idea that children looked for validation from the experimenter whilst answering to the robot demonstrator more than to the human demonstrator.

The number of tactile contacts varied with the interaction between children's age and the demonstrator type. The number of tactile contacts increased more with age for the robot demonstrator group than for the human demonstrator group. Touching the robot might be more socially acceptable than touching an unfamiliar adult. Moreover, older children could be less shy and feel more comfortable to touch the robot than younger ones.

Regarding pre and post test questionnaires, we found that categorisation was clearer for the human demonstrator than for the robot demonstrator: standard errors were higher in the robot group which translated that answers were more homogeneous for the human group. The number of correct categorisations in group 2 and group 3 increased with age. We found a similar result when comparing group 1, group 2 and group 3 on the seven common items: the number of errors decreased with age and this effect of age was stronger when the demonstrator was a human rather than a robot. It can be linked to the improvement of children knowledge about human and robot capacities.

Taken together, our study provides encouraging results concerning the use of the social robot NAO with children between 3 and 6 years old to evaluate children knowledge in several domains, from vocabulary learning to prosocial behaviour.

Acknowledgments

We would like to thank all the children who participated as well as schools' directors who coordinated the study. We also thanked Louisane Araguas, Camille Nozières, Aïcha Bekkaoui, Oumaima Doumeksa and Pauline Bernard for their help in collecting and analysing data.

References

Aksoy, P., & Baran, G. (2010). Review of studies aimed at bringing social skills for children in preschool period. Procedia - Social and Behavioral Sciences, 9, 663 669. https://doi.org/10.1016/j.sbspro.2010.12.214

Bartneck, C., & Forlizzi, J. (2004). *A design-centred framework for social human-robot interaction*. 591-594. https://doi.org/10.1109/ROMAN.2004.1374827

Beck, A., Cañamero, L., & Bard, K. (2010). *Towards an Affect Space for robots to display emotional body language*. 464-469. https://doi.org/10.1109/ROMAN.2010.5598649

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of Emotional Body Language Displayed by a Humanoid Robot : A Case Study with Children. *International Journal of Social Robotics*, *5*(3), 325-334. https://doi.org/10.1007/s12369-013-0193-z

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education : A review. *Science Robotics*, *3*(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Bergès, J., & Lézine, I. (1963). *Test of imitation of gestures: Methods for plotting body image and movements of children from 3 to 6 years of age*. Masson & Cie.

Bullens, J., Iglói, K., Berthoz, A., Postma, A., & Rondi-Reig, L. (2010). Developmental time course of the acquisition of sequential egocentric and allocentric navigation strategies. *Journal of Experimental Child Psychology*, *107*(3), 337-350. https://doi.org/10.1016/j.jecp.2010.05.010

Camões-Costa, V., Erjavec, M., & Horne, P. J. (2011). The impact of body-part-naming training on the accuracy of imitative performances in 2- to 3-year-old children. *Journal of the Experimental Analysis of Behavior*, *96*(3), 291-315. https://doi.org/10.1901/jeab.2011.96-291

Costa, S., Lehmann, H., Dautenhahn, K., Robins, B., & Soares, F. (2014). Using a Humanoid Robot to Elicit Body Awareness and Appropriate Physical Interaction in Children with Autism. *International Journal of Social Robotics*, *7*, 1-14. https://doi.org/10.1007/s12369-014-0250-2

de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E., & Vogt, P. (2018). *The Effect of a Robot's Gestures and Adaptive Tutoring on Children's Acquisition of Second Language Vocabularies*. 50-58. https://doi.org/10.1145/3171221.3171277

Dunfield, K. A., Best, L. J., Kelley, E. A., & Kuhlmeier, V. A. (2019). Motivating Moral Behavior : Helping, Sharing, and Comforting in Young Children With Autism Spectrum Disorder. *Frontiers in Psychology*, *10*, 25. https://doi.org/10.3389/fpsyg.2019.00025

Dunfield, K. A., & Kuhlmeier, V. A. (2013). Classifying prosocial behavior : Children's responses to instrumental need, emotional distress, and material desire. *Child Development*, *84*(5), 1766-1776. https://doi.org/10.1111/cdev.12075

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56-75. https://doi.org/10.1007/BF01115465

Eysenck, S. B. G., & Eysenck, H. J. (1968). The Measurement of Psychoticism : A Study of Factor Stability and Reliability. *British Journal of Social and Clinical Psychology*, *7*(4), 286-294. https://doi.org/10.1111/j.2044-8260.1968.tb00571.x

Havy, M., Nazzi, T., & Bertoncini, J. (2013). Phonetic processing during the acquisition of new words in 3-to-6-year-old French-speaking deaf children with cochlear implants. *Journal of Communication Disorders*, *46*(2), 181-192. https://doi.org/10.1016/j.jcomdis.2012.12.002

Izard, C. E. (1971). The face of emotion. New York: Appleton Century-Crofts Educational Division, Meredith.

Kennedy, J., Baxter, P., & Belpaeme, T. (2015). Comparing Robot Embodiments in a Guided Discovery Learning Interaction with Children. *International Journal of Social Robotics*, *7*(2), 293-308. https://doi.org/10.1007/s12369-014-0277-4

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., Senft, E., & Belpaeme, T. (2017). Child Speech Recognition in Human-Robot Interaction : Evaluations and Recommendations. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 82-90. https://doi.org/10.1145/2909824.3020229

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12577-12580. https://doi.org/10.1073/pnas.0705345104

Klatzky, R. L. (1998). Allocentric and Egocentric Spatial Representations : Definitions, Distinctions, and Interconnections. In C. Freksa, C. Habel, & K. F. Wender (Éds.), *Spatial Cognition : An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (p. 1-17). Springer. https://doi.org/10.1007/3-540-69342-4_1

Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., & Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction*, *13*, 1-9. https://doi.org/10.1016/j.ijcci.2017.04.001

Li, J., & Chignell, M. (2011). Communication of Emotion in Social Robots through Simple Head and Arm Movements. *International Journal of Social Robotics*, *3*(2), 125-142. https://doi.org/10.1007/s12369-010-0071-x

Newcombe, N., & Huttenlocher, J. (2003). *Making Space : The Development of Spatial Representation and Reasoning*. MIT Press.

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*(1), 222-231. https://doi.org/10.1016/j.cognition.2007.12.003

Perloff, R. M. (1982). Gender constancy and same-sex imitation : A developmental study. *The Journal of Psychology: Interdisciplinary and Applied*, *111*(1), 81-86. https://doi.org/10.1080/00223980.1982.9923516

Pisella, L., Havé, L., & Rossetti, Y. (2019). Body awareness disorders : Dissociations between body-related visual and somatosensory information. *Brain : a journal of neurology*, *142*, 2170-2173. https://doi.org/10.1093/brain/awz187

Riggio, H. R., & Riggio, R. E. (2002). Emotional Expressiveness, Extraversion, and Neuroticism : A Meta-Analysis. *Journal of Nonverbal Behavior*, *26*(4), 195-218. https://doi.org/10.1023/A:1022117500440

Suzuki, R., & Lee, J. (2016). Robot-play therapy for improving prosocial behaviours in children with Autism Spectrum Disorders. *2016 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*. https://doi.org/10.1109/MHS.2016.7824238

Suzuki, R., Lee, J., & Rudovic, O. (2017). *NAO-Dance Therapy for Children with ASD*. 295-296. https://doi.org/10.1145/3029798.3038354

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S., & Leseman, P. (2019). Social Robots for Language Learning : A Review. *Review of Educational Research, 89*(2), 259-295. https://doi.org/10.3102/0034654318821286

Visser, N., Alant, E., & Harty, M. (2008). Which Graphic Symbols do 4-Year-Old Children Choose to Represent Each of the Four Basic Emotions? *Augmentative and Alternative Communication, 24*(4), 302-312. https://doi.org/10.1080/07434610802467339

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science (New York, N.Y.), 311*(5765), 1301-1303. https://doi.org/10.1126/science.1121448

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, *39*(1), 114-128. https://doi.org/10.1037//0012-1649.39.1.114

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, *106*(4), 208-220. https://doi.org/10.1016/j.jecp.2010.03.002

Supplementary Information

Appendix 1. Script of the experiment

The experimenter installs the child in front of the demonstrator. Alice is the name of the human demonstrator, Naomie is the name given to the robotic demonstrator. Then she asked the following questions:

Pre test questions for G2 and G3:

- Is Alice/Naomie a human? An object? A machine? An animal? A toy? A computer?
- Does Alice/Naomie eat? Grow up? Think? Is happy? See things?
- If I tickle Alice/Naomie, will she feel it?
- Does Alice/Naomie can feel pain?
- Does Alice/Naomie can hurt someone?
- Does Alice/Naomie know what you think?
- Did someone make Alice/Naomie?
- Does Alice/Naomie can be broken?

Pre test questions kept for G1:

- Is Alice/Naomie a human? An object? An animal?
- Does Alice/Naomie eat? Grow up? Is happy?
- Does Alice/Naomie can feel pain?

The demonstrator remains silent while the experiment asked the questionnaire to the child.

Warm-up phase

- Hello, my name is Naomie (/Alice) and you, what is your name? **(Question 1)**
- Do you know the song “Petit moulin”? **(Question 2)**
- I will sing it and if you want, you can sing it with me.

“Tourne, tourne, petit moulin. Tapent, tapent, petites mains. Vole, vole, petit oiseau. Nage, nage, petit poisson. Petit moulin a bien tourné, petites mains ont bien tapé, petit oiseau a bien volé, petit poisson a bien nagé”.

Vocabulary task

- I will show you images of imaginary animals. Are you ready? Let’s go!
- Now, you will learn their name. Remember their name because I will ask you later.
- Look at this animal. It is a dax. It is a dax.
- What is the name of this animal? **(Question 3)**
- Look at this animal. It is a grid. It is a grid.
- What is the name of this animal? **(Question 4)**
- Touch the dax.
- Touch the grid.
- Touch the grid.
- Touch the dax.

Body task

Comprehension of body parts labels:

- Now, you have to show me one of your body part. To start, show me your [body part] (repeated for face, eye, nose, mouth, ear, shoulder, elbow, hand, belly, knee, foot).
- Now, show me my [body part](repeated for face, eye, nose, mouth, ear, shoulder, elbow, hand, belly, knee, foot).

Imitation of body parts sequences:

- I will show you several body parts on my body. Look carefully because you will have to do the same thing after.

- The demonstrator randomly showed one of his body part, for example, his eye and asked “What is the name of this body part?” **(Questions 5, 7, 9)** and a second one, for example his belly, and said “And this one?” **(Questions 6, 8, 10 and 11)**. Then, the demonstrator placed his hands on his knees and said “This is you turn”.

Identification of emotions:

- Do you know what an emotion is? **(Question 12)**
- We will speak about 4 emotions: anger, joy, sadness and fear. Show me the image in which the child feels [emotion].
- I stand up for the rest of the exercise. Please, stand up too.
- Then, the demonstrator and he child stood up. The demonstrator asked “Show me how do you express [emotion] with your whole body”.
- Can you tell me what are the 4 emotions we talked about? **(Question 13)**
- Let me show you what I am doing when I am feeling an emotion.

The demonstrator reproduced one of the four emotional poses and the experimenter asked the children “Which emotion do you think it is?” After the child replied, the demonstrator returned to the neutral pose. The demonstrator repeated this sequence for the three other body postures.

Spontaneous helping situation

- The demonstrator and the child were seated in front of each other. The experimenter gave a blue stick to the demonstrator while saying “Look, I have something for you”. The demonstrator grabbed it, looked at it and let it fall on the ground while saying “Oh, it fell!”. She tried to reach the stick with his hand.
- At the end of the experiment, the demonstrator said “Thank you, goodbye!”

The demonstrator remains silent while the experiment asked the post-test questionnaire to the child.

Post test questions for G2 and G3:

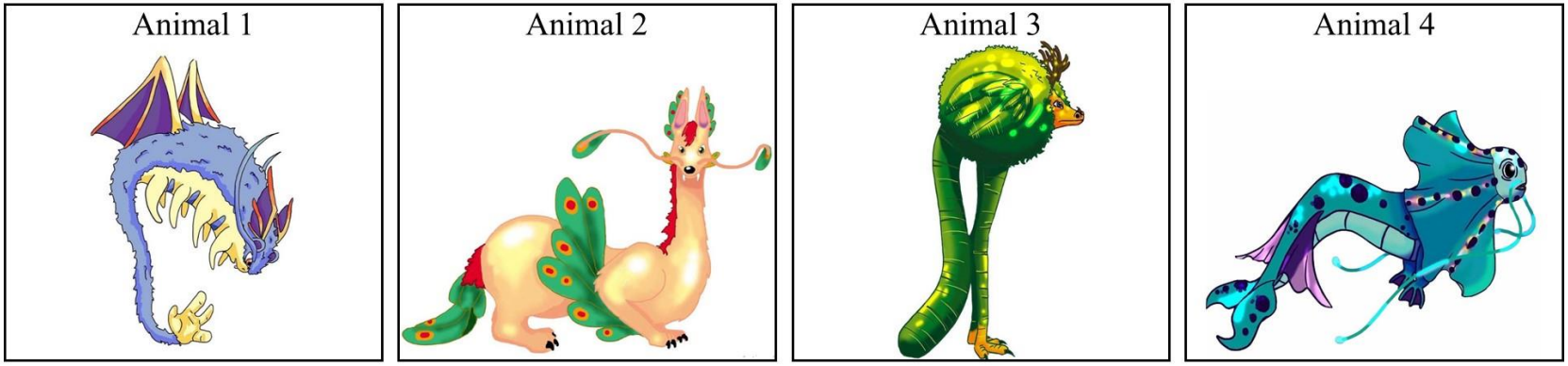
- Is Alice/Naomie a human? An object? A machine? An animal? A toy? A computer?
- Does Alice/Naomie eat? Grow up? Think? Is happy? See things?
- If I tickle Alice/Naomie, will she feel it?
- Does Alice/Naomie can feel pain?
- Does Alice/Naomie can hurt someone?
- Does Alice/Naomie know what you think?
- Did someone make Alice/Naomie?
- Does Alice/Naomie can be broken?

Post test questions kept for G1:

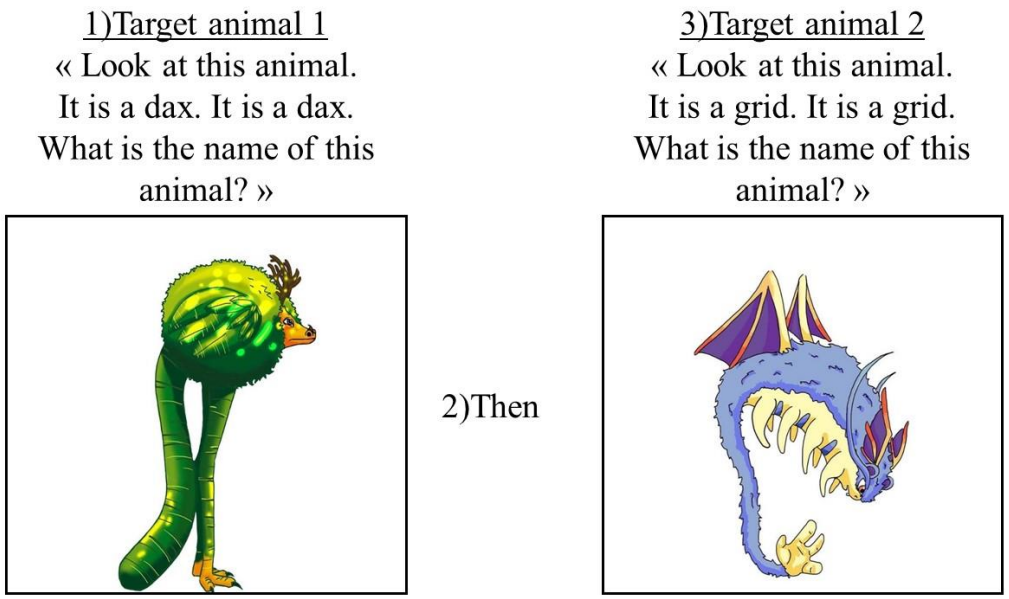
- Is Alice/Naomie a human? An object? An animal?
- Does Alice/Naomie eat? Grow up? Is happy?
- Does Alice/Naomie can feel pain?

Figure 4. Description of the vocabulary task. Images of imaginary animals were created by Louisane Araguas.

1. Presentation phase: each animal was presented individually on the screen-touch tablet during 10s.



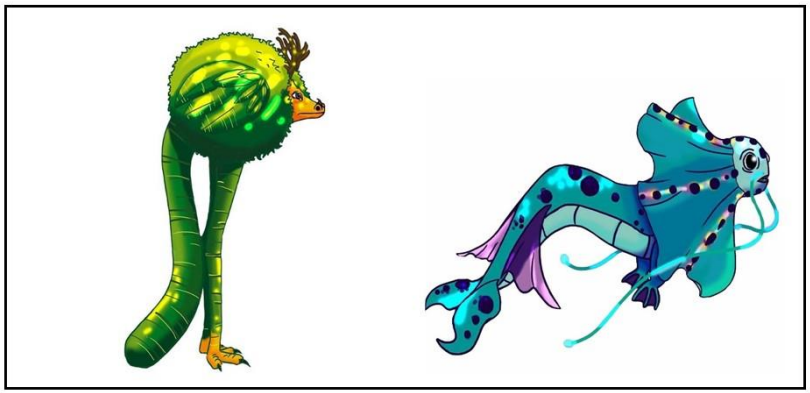
2. Naming phase: the two target animals were named one after the other.



3. Test phase

Target versus non-target animals:

1) « Touch the dax »



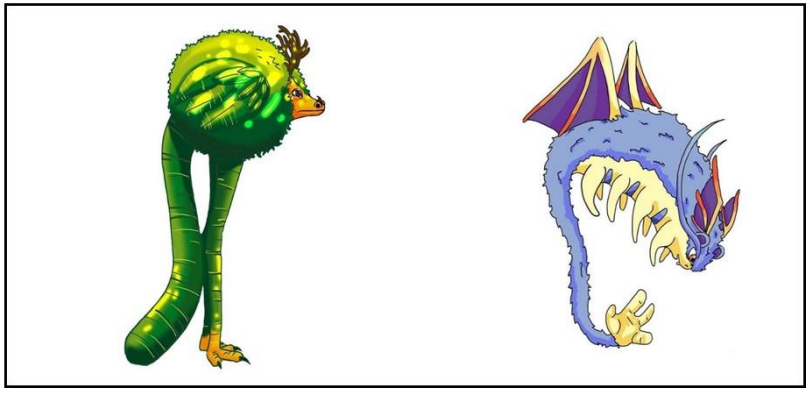
2)Then

3) « Touch the grid »



Target versus target animals:

1) « Touch the grid »



2)Then

3) « Touch the dax »



Figure 5. The 9 body parts pointed by the demonstrator in the imitation task: (A) human and (B) robotic.

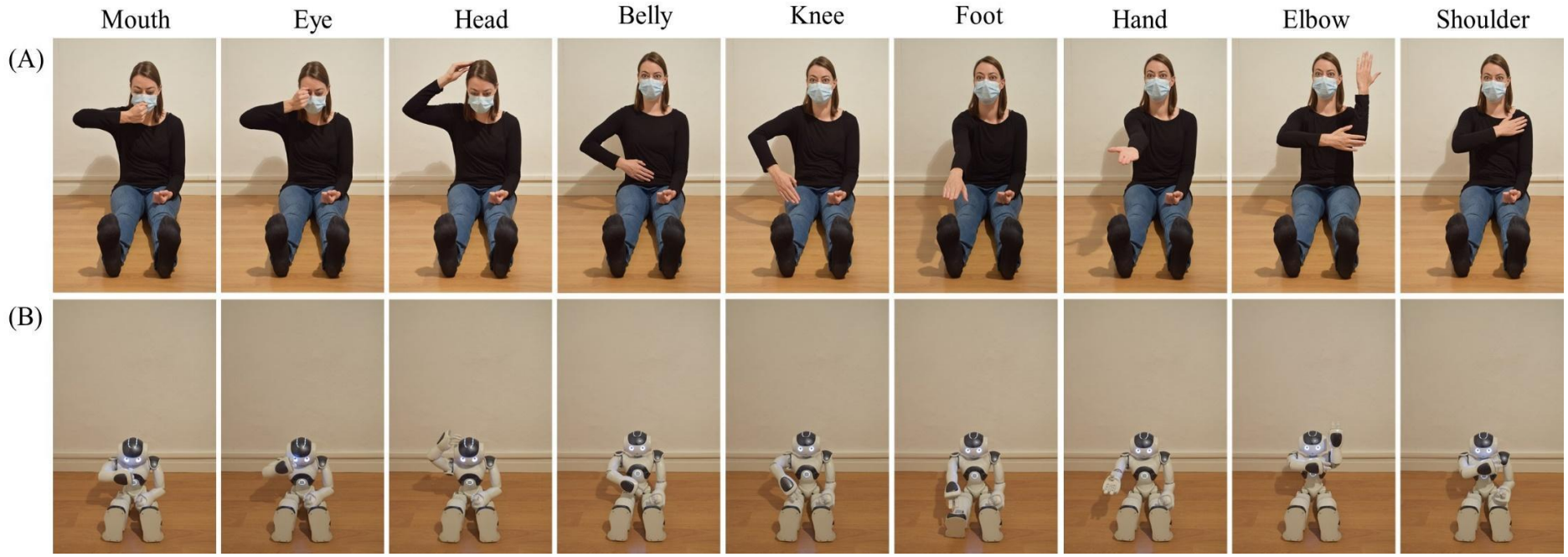


Figure 6. Images used in the task of emotions' identification (source: https://fr.freepik.com/)



Figure 7. The neutral and emotional postures exhibited by the demonstrator: (A) human and (B) robotic.

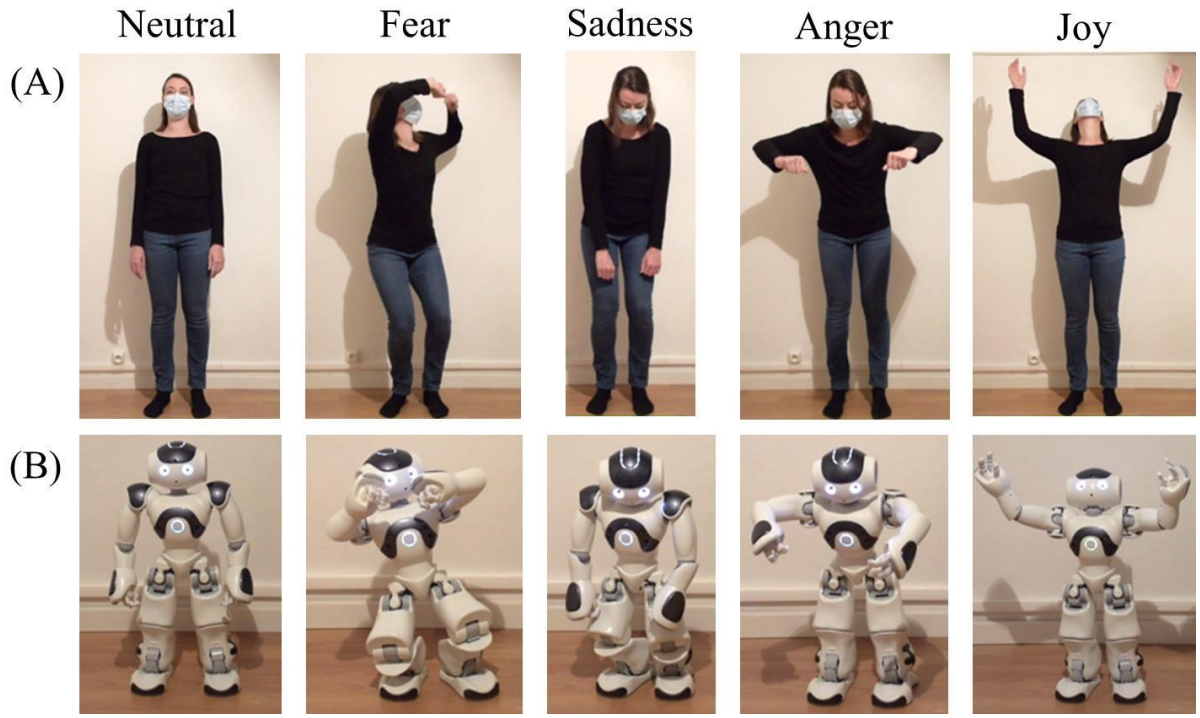


Table 3. Ethogram used on BORIS software to analyse child's behaviours. Events with a duration inferior to 1s were excluded.

Code	Behaviour	Description
a	The demonstrator talks	Moments where the demonstrators is talking to the child, in s
Where does the child look when the demonstrator is talking?		
q	at the demonstrator	Child gaze is orientated towards the demonstrator, in s
s	at the experimenter	Child gaze is orientated towards the experimenter, in s
e	at the touch-pad screen, the images or the blue stick	Child gaze is orientated towards the objects, in s
d	anywhere else	Child gaze is orientated anywhere else, in s
During the experiment, the demonstrator asks direct questions to the child. What is the child's reaction?		
f	No answer	The child does not respond to the demonstrator question
g	Answering while looking at the demonstrator	The child answers to the demonstrator while looking at her
h	Answering while looking at the experimenter	The child answers to the demonstrator while looking at the experimenter
j	Answering while looking in another direction	The child answers to the demonstrator while looking in another direction
t	The child touches the demonstrator	Number of times the child touches the demonstrator

Table 5. Percentage of correct answers and latencies (mean ± SE) of responses in the vocabulary task for G2 and G3 and for the human and the robot groups.

	Human		**Robot**	
	Correct answers vocabulary task	Latencies of responses vocabulary task	Correct answers vocabulary task	Latencies of responses vocabulary task
G2	79.49% ± 3.55	5 s ± 0.38	84.62% ± 4.16	4.30 s ± 0.34
G3	100%	3.82 s ± 0.15	96.97 ± 1.02	4.13 s ± 0.22

Table 6. Percentage of correct body parts labelled (mean ± SE) for G1, G2 and G3 and for the human and the robot groups.

	Human		**Robot**	
	On the child's body	On the demonstrator's body	On the child's body	On the demonstrator's body
G1	92.05% ± 1.34	68.18% ± 7.59	83.33% ± 7.46	90.91% ± 3.71
G2	93.71% ± 1.19	90.91% ± 2.41	90.21% ± 2.79	86.71% ± 3.73
G3	95.87% ± 0.72	97.52% ± 0.64	97.52% ± 0.64	98.35% ± 0.55

Table 7. Percentage of correct answers in the imitation subtask (mean ± SE) for G1, G2 and G3 and for the human and the robot groups.

	Human			**Robot**		
	Identification of body parts	Imitation of body parts sequences	Respect of order in the sequence	Identification of body parts	Imitation of body parts sequences	Respect of order in the sequence
G1	71.88% ± 7.99	39.58% ± 6.85	31.25% ± 7.79	62.50% ± 3.18	23.61% ± 4.74	4.17% ± 2.08
G2	85.26% ± 2.80	65.38% ± 4.05	49.36% ± 2.81	67.31% ± 4.21	63.46% ± 3.99	51.92% ± 3.57
G3	93.94% ± 2.46	90.91% ± 4.63	68.18% ± 11.77	77.27% ± 3.64	51.52% ± 5.35	32.58% ± 5.74

Table 8. Percentage of correct answers in the emotion subtask (mean ± SE) for G1, G2 and G3 and for the human and the robot groups.

	Human				**Robot**			
	Identification on images	Emotions expressed	Citation of emotions	Identification of key poses	Identification on images	Emotions expressed	Citation of emotions	Identification of emotional key poses
G1	90.63% ± 3.29	37.50% ± 9.15	9.38% ± 4.69	25% ± 4.72	91.67% ± 4.17	56.25% ± 7.86	16.67% ± 8.33	29.17% ± 3.84
G2	100%	51.92% ± 5.70	41.67% ± 6.59	30.77% ± 3.21	96.15% ± 1.30	30.77% ± 4.93	80.77% ± 5.13	38.46% ± 3.35
G3	90.91% ± 3.05	57.95% ± 7.36	95% ± 1.67	50% ± 4.46	100%	65.91% ± 5.66	90.91% ± 4.55	43.18% ± 4.49

Table 9. Percentage of questions answered while looking at the experimenter (mean ± SE) for G1, G2 and G3 and for the human and the robot groups.

	Human	**Robot**
G1	7.95% ± 2	22.73% ± 3.47
G2	2.42% ± 0.52	17.12% ± 3.38
G3	3.50% ± 0.80	17.60% ± 3.81

Table 10. Number of tactile contacts (mean ± SE) toward the demonstrator for G1, G2 and G3 and for the human and the robot groups.

	Human	**Robot**
G1	0.75 ± 0.26	5.83 ± 1.17
G2	2.23 ± 0.59	9.85 ± 0.85
G3	1.18 ± 0.50	8.91 ± 1.10

Table 11. Correct answers in the questionnaire for the human and the robot group

	Correct answers	
	Human group	**Robot group**
Human?	Yes	No
Object?	No	Yes
Animal?	No	No
Eat?	Yes	No
Grow?	Yes	No
Happy?	Yes	No
Feel pain?	Yes	No
Machine?	No	Yes
Toy?	No	No
Computer?	No	No
Think?	Yes	No
See things?	Yes	No
Feel tickles?	Yes	No
Hurt someone?	Yes	No
Know what you think?	No	No
Made by someone?	No	Yes
Can be broken?	No	Yes

CHAPITRE II

Body awareness and prosocial behaviour in Autism Spectrum Disorder children interacting either with a human or a robot



Le robot NAO déguisé par les enfants du SESSAD lors des présentations en groupe.

Body awareness and prosocial behaviour in Autism Spectrum Disorder children interacting either with a human or a robot

Alice Araguas [1*], Arnaud Blanchard [2], Adrien Chopin[3], Sébastien Derégnaucourt [1, 4] and Bahia Guellaï[1,4]

[1]*Laboratoire Ethologie Cognition Développement, Université Paris Nanterre, 200 Avenue de la République, F92001 Nanterre Cedex, France*

[2] *Laboratoire Equipe Traitement de l'Information et Systèmes, Université Cergy-Pontoise, F95000 Pontoise, France*

[3]*INSERM, CNRS, Institut de la Vision, Sorbonne Université, F75012, Paris, France*

[4]*Institut Universitaire de France*

* Correspondence: Alice Araguas, Laboratoire Ethologie Cognition Développement, UPL, Univ Paris Nanterre, F92000 Nanterre, France. E-mail address: alice.araguas@parisnanterre.fr

Abstract

Past years, multiple studies showed promising results concerning the use of social robots with ASD children. In our study, we compared behaviours of six children with mild Autism Spectrum Disorder from 4 to 8 years old regarding it was an unfamiliar adult or the humanoid NAO robot who gave them instructions in body awareness tasks (comprehension and production of body parts labels, imitation, identification of emotions on boy postures) and in a pro-social situation. We observed an important inter-individual variability in children's responses and knowledges. We did not find any difference in scores between the robot and the human conditions. However, we found that children looked more at the robotic demonstrator than at the human one during experimental sessions. Taken together, our results suggest that the NAO robot can be successfully used to evaluate body parts knowledge, emotions identification and the tendency of ASD children to act pro-socially.

Keywords: child-robot interaction, ASD children, body awareness, emotions, emotional body postures, prosocial behaviour

1. Introduction

According to DSM-5, Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder in which individuals express highly repetitive and restrictive interests and behaviours and persistent deficits in social communication and interaction across multiple contexts (American Psychiatric Association, 2013). Socio-communicative deficits have important consequences on their daily life (White et al., 2007): ASD individuals exhibit difficulties with socialisation range from basic aspects of social interaction, such as joint attention and emotion perception, to impairments in more complicated social tasks, like establishing and maintaining relationships (Kjellmer et al., 2012; Mashal & Kasirer, 2012; Matson & Wilkins, 2007; Romanczyk et al., 2005 cited in Cervantes et al., 2013). As profiles of ASD children are widely heterogeneous, many authors and clinicians emphasize the necessity to develop various therapeutic and educative tools (Pennisi et al., 2016).

In the last decade, numerous scientific projects focused on the use of social robots as educational robots which convey information in a tutor-student relationship and as rehabilitation robots which provide structured physical therapy (Scassellati et al., 2018). Following the conclusions of two recent reviews (Damianidou et al., 2020; Saleh et al., 2020), robotic sessions seem to be beneficial to individuals with ASD especially for social communication and interaction skills. In another review, Pennisi and colleagues (2016) compared performances of participants with ASD in a robot and a human conditions. In 12 out of 16 studies, they found better performances in the robot condition than in the human condition (Bird et al., 2007; Damm et al., 2013; Duquette et al., 2008; Kim et al., 2013; Lee et al., 2012; Lee & Obinata, 2013; Michaud et al., 2007; Pierno et al., 2008; Shamsuddin et al., 2012, 2013; Wainer et al., 2014a,b).

Humanoid robots are the most frequently used with ASD individuals: their cute design (i.e. with big eyes, curved shape, highly expressive body language and facial expressions) make

them appealing and help avoiding fear reactions especially among children (Raptopoulou et al., 2021). On the 163 articles reviewed, Saleh and colleagues (2020) found that the NAO robot is by far the most used (in 30.5% of studies) followed by KASPAR robot (8.5%) and ZENO robot (5%) as presented in Figure 1.



Figure 1. Pictures of the three most frequently humanoid robots used in research with ASD individuals: (A) the NAO robot[1] , (B) the KASPAR robot[2] and (C) the ZENO robot[3].

With ASD children, social robots are mainly used for teaching and skills practicing (28.8% of studies) but also evaluating abilities (19.6%), providing feedback or encouragement (12.3%), joint attention training (10.4%), eliciting social behaviours, emotion recognition and expression (6.1%), imitation (6.1%), vocalization (4.3%), turn-taking (2.5%) and sometimes diagnostic evaluation (1.2%) (Saleh et al., 2020).

In the present study, we focused our analysis on 4 to 8 years old children with mild ASD. Mild ASD is considered as the least severe form of ASD (Level 1) compared to the two

[1] Source: https://www.softbankrobotics.com/emea/fr/nao/
[2] Source: https://robots.ieee.org/robots/kaspar/
[3] Source: https://www.hansonrobotics.com/zeno/

other levels of severity. Mild ASD individuals require minimal support, i.e. minimal environmental modifications, for daily functioning (American Psychiatric Association, 2013).

We compared ASD children's behaviours when facing either a NAO robot, to which we attributed a feminine gender (robot condition), or an unfamiliar adult (human condition). Each child was tested across 4 sessions, two per condition. Each session included three tasks: a body parts task including the comprehension and the production of body parts labels then, the imitation of body parts' sequences, an emotions task and a spontaneous helping situation.

First of all, having full repertoire of movement behaviours allows to communicate accurately and to understand others' movements during social interactions (Palestra et al., 2014). Posture – body parts positions – can be considered as the first step in improving ASD children's abilities because of its importance for both gross and fine motor functions (Srinivasan and al., 2014 in Palestra et al., 2014). Moreover, authors suggest that ASD individuals have difficulties in using their own bodily information to put themselves in the place of someone else (Conson et al., 2015). We found that body knowledge had been mostly tested through imitation in previous studies, using KASPAR robot (Costa et al., 2015) and NAO robot (Suzuki & Lee, 2016; Suzuki et al., 2017). This is the reason why the first aim of our study was to evaluate separately knowledge of body parts labels on the child's body and on the demonstrator's body following the procedure described by Bergès & Lézine (1963). In a second part, we evaluated body parts labels production and imitation of designated body parts sequences.

Individuals with ASD also have difficulties in using their body to convey information which conduct to less efficient social communication. They have deficits in emotions recognition from the voice intonation, facial expressions and body language (Hobson, 1986a, b; Baron-Cohen et al., 2000, 2001 cited in Baron-Cohen, 2009). In their review, Saleh and colleagues (2020) cited nine out of ten studies in which social robots had been successfully used

to improve emotions recognition and expression for ASD children. In the present study, we evaluated the identification of emotions in different supports (images and the demonstrator body postures). In two studies, Beck and colleagues (2010; 2013) explored the identification of the NAO robot's emotional key poses - defined as static body postures - by adults and 11 to 13 years old children with a typical development. They demonstrated that adults and children can recognize emotional body key poses even without facial expressions. We replicated this study with younger and ASD children using the four most recognised NAO key poses (anger, sadness, fear, happiness). Moreover, we evaluated emotions expression by children.

In the last part of our study, we measured prosocial behaviour as ASD decreased spontaneous (i.e., without very explicit cues) prosocial skills (Bacon et al., 1998; Hale & Tager-Flusberg, 2005). Prosocial behaviours refer to the general ability to support others, which encompasses crucial skills in the development of friendships and reciprocal relationships such as helping, sharing, comforting, co-operating, and volunteering (Jameel et al., 2014). It is important for children to notice and interpret social cues, to develop empathy toward others, to recognize when help is needed and to decide to help or not by performing an appropriate behaviour (Marion, 2003). In our study, we focused on instrumental need which, for Dunfield and colleagues (2019), occurs when an individual is unable to complete a goal directed behaviour. As proposed by these authors, we used the term of "spontaneous helping" to refer to "other-oriented acts aimed at alleviating another's instrumental need" (Dunfield et al., 2019). We evaluated whether ASD children will help the demonstrator after having interacted with her during 15 minutes.

In the present study, we analysed children's scores in each task and the time they spent looking at the demonstrator. We hypothesized that scores would be better in the robot condition than in the human condition and that children will spend more time looking at the demonstrator in the robot condition than in the human condition.

2. Materials and methods

2.1. Participants

Six children (one girl, five boys) aged between 4 years old 9 months and 7 years old 11 months (mean ± SE: 6 years 7 months ± 2 months) and with a diagnosis of a mild Autism Spectrum Disorder (ASD) participated to the study (Table 1). Diagnoses were performed by medical experts. Full permission to take part and to be video and audio recorded was acquired for all participants. All children were French native speakers and came from a specialized institution for children with ASD. Data were collected in May 2021 in an Educational and Care Service.

Child	**Gender**	**Age**
E1	Boy	4 years 9 months
E2	Boy	6 years 2 months
E3	Boy	6 years 6 months
E4	Girl	6 years 7 months
E5	Boy	7 years 6 months
E5	Boy	7 years 11 months

Table 1. Children's profile (gender, age).

2.2. Experimental procedure

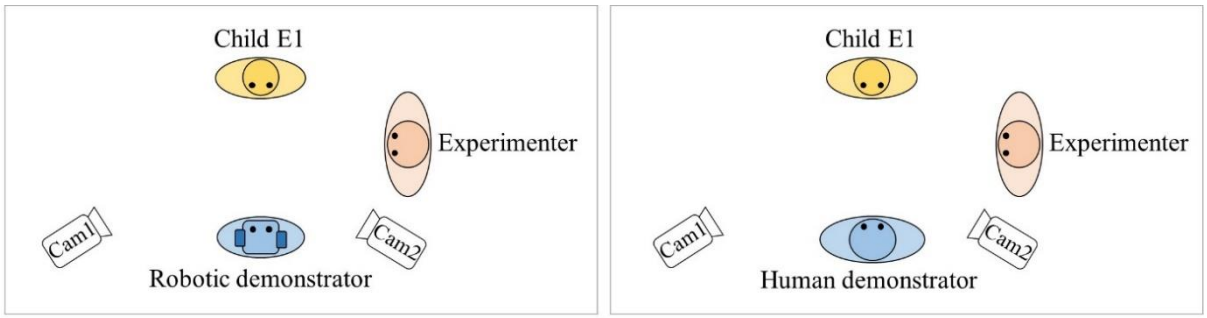
The experimental procedure is an adaptation of the one used in our previous study with typical developing children (see Chapter I).

2.2.1 General setup

The experiment took place in a specialized institution, in a quiet room in presence of two adult females (Figure 2). The experimenter asked the child to sit on the floor in front of the

demonstrator and introduced the demonstrator to him/her. To decrease anxiety of novelty, the experimenter was the child's person of reference in the institution. Two experimenters, one was the referent for five children, the other one was the referent for one child, were trained to follow the experimental setup.

The demonstrator was an unfamiliar woman for the human condition and the NAO robot



(A) Robot condition (B) Human condition

for the robot condition. Due to the Covid19, the experimenter and the human demonstrator wore face masks during the whole experiment. Two Sony HDR-CX410VE cameras placed on tripods from either side of the demonstrator allow us to record the child's behaviours (including gazes and vocalizations).

Figure 2. Top view of the experimental setup for the two conditions: (A) robot condition, (B) human condition (inspired by Kennedy et al., 2015). The child and the demonstrator sat in front of each other and the experimenter was on the right side of the demonstrator.

As we made the hypothesis that children with ASD will improve their performances with the repetition of sessions, we decided to organise four sessions per child, following an intra-subject procedure. Each child was tested during 2 blocks of 30 min separated by 7 days.

Each block included the human condition and the robot condition. Half of the children started with the human condition, the other half with the robot condition (see Table 2).

2.2.2. Human condition: exposure to an adult demonstrator

The human demonstrator was always the same woman and remained as neutral as possible during the experiment: she gave the instructions to the child but she did not encourage him/her or give him/her feedbacks.

2.2.3. Robot condition: exposure to the NAO robot demonstrator

Created by Aldebaran Robotics, the NAO robot version 6 is a bipedal humanoid robot 58 cm high and with 25 degrees of freedom. It can perform various movements and catch small objects with its three fingers - hands. As presented in Figure 3, NAO is equipped with two video cameras on its mouth and its forehead. A stereo broadcast system made up of 2 loudspeakers, on its ears and 4 omnidirectional microphones (2 on the top of its head and 2 on the back of its head) allows it to have vocal capacities (recognition and synthesis). It can estimate its distance to obstacles in the surrounding using two ultrasonic sensors (or sonars). Moreover, it has contact and several tactile sensors: tactile head and hands, chest button and feet bumpers.

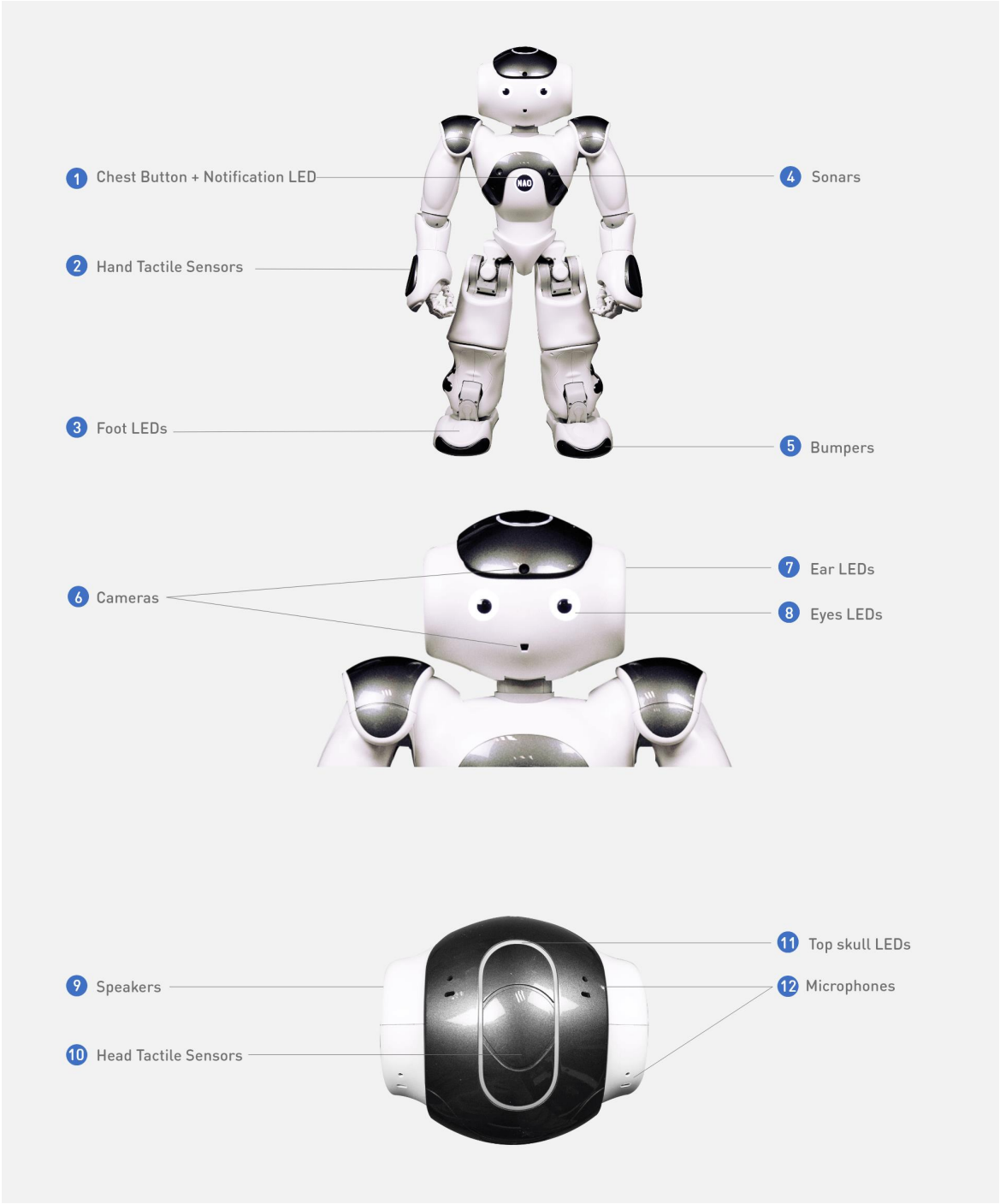


Figure 3. Overview of the humanoid NAO6 robot from SoftBank Robotics' website (source: https://www.softbankrobotics.com/).

It can be programmed to carry out autonomously a set of tasks. For our experiment, we created a program containing experimental tasks hosted on a website that allowed the experimenter to control the robot via the touch-screen tablet during the test.

The NAO robot can be programmed to carry out autonomously a set of tasks. However, for our experiment, the robot was fully tele-operated to enable the robot to act contingently (Kennedy et al., 2017): the experimenter controls the robot via a touch-screen tablet, an iPad mini 4 (20.32 x 13.48 x 0.61 cm), during the session. As shown in Figure 4, the setup consisted on connecting the tablet on the wifi access point of the robot. The robot ran a web server that can be accessed with any browser on the tablet. An HTML page displays the interface and executed the sequence of the experiment in javascript (« vue.js » for the graphical interface, « LibQi » for controlling the robot).

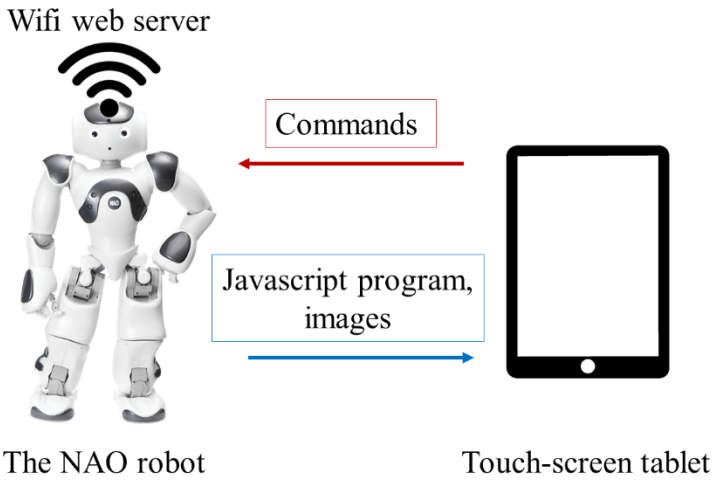


Figure 4. The robot is controlled by a touch-screen tablet.

As the human demonstrator was a female, we attributed a feminine gender to the NAO robot and a feminine name (“Naomie”). We will use the pronoun “her” in the following text.

To ensure that the two demonstrators pronounced the same sentences in both conditions, vocal recordings of the female adult were played back by the robot. Sentences were recorded while she was alone in a silent room and as she was addressing instructions to a child.

2.2.4. Experimental tasks and scores

Experimental tasks were organised as described in Table 2.

	Session	Child E1	Child E2
Week 1	1	Human condition: warm-up phase, body parts task, emotions task	Robot condition: warm-up phase, body parts task, emotions task
	2	Robot condition: warm-up phase, body parts task, emotions task, **spontaneous helping situation**	Human condition: warm-up phase, body parts task, emotions task, **spontaneous helping situation**
Week 2	3	Robot condition: warm-up phase, body parts task, emotions task	Human condition: warm-up phase, body parts task, emotions task
	4	Human condition: warm-up phase, body parts task, emotions task, **spontaneous helping situation**	Robot condition: warm-up phase, body parts task, emotions task, **spontaneous helping situation**

Table 2. Experimental tasks depending on the organisation of experimental sessions.

2.2.4.1. Warm-up phase

Before starting the experimental tasks, there was a warm-up phase to allow the child to see how the demonstrator (robotic or human depending on the session) spoke and moved before the beginning of the experiment. The demonstrator said her name to the child and asked the child to introduce himself/herself. The demonstrator sang a popular French rhyme (see

Appendix 1 in Supplementary Information) with hand gestures and encouraged the child to follow it.

2.2.4.2. Body parts task

Comprehension of body parts labels:

We evaluated the comprehension of eleven body parts labels: face, eye, nose, mouth, ear, shoulder, elbow, hand, belly, knee, foot. Each body part was randomly chosen in this list. First, the demonstrator asked the child "*Show me your [body part]*" and repeated it for the eleven possibilities. Then she asked "*Show me my [body part]*" and also repeated it for the eleven body parts.

The child obtained 1 point for each body part correctly shown (in total, 11 points for identification on his/her body and 11 points for identification of the demonstrator's body).

Imitation of body parts' sequences:

In a second phase, we evaluated body parts labels production: the demonstrator randomly showed one of her body part, for example, her eye and asked "*What is the name of this body part?*" and a second one, for example her belly, and said "*And this one?*". We scored one point if the child correctly named the body part showed by the demonstrator.

As it was complicated for the NAO robot to point its nose and its ear, we decided to keep 9 body parts for this sub-task (face, eye, mouth, shoulder, elbow, hand, belly, knee and foot, Figure 5 in Supplementary Information).

Then, the demonstrator placed her hands on her knees and asked the child to repeat the body parts sequence. We counted one point if the child showed the correct body part and respected the order of the body part in the sequence.

The demonstrator performed two sequences with two body parts ("sequence#1", "sequence#2") and a third sequence with three body parts ("sequence#3"). As the majority of children had difficulties to understand the instructions in this task, the first two-body parts

sequence (sequence#1) was used by the experimenter to explain the instructions again in order to help the child to understand it.

2.2.4.3. Emotions task

Identification on images:

Four images of a child expressing an emotion (joy, sadness, anger, fear. Supplementary Information, Figure 6) were placed in front of the child. The demonstrator asked "*Do you know what an emotion is? Show me the image in which the child feels [emotion]*". If the child showed the correct image, he/she obtained 1 point (4 points in total).

Emotions expressed by the child:

Then, the demonstrator and the child stood up. The demonstrator asked "*Show me how do you express [emotion] with your whole body"*. She waited for the child to express the emotion. The instruction was repeated with the four emotions. If the child expressed the emotion with his/her whole body, he/she obtained 1 point (4 points in total). If he/she expressed the emotion only with his face, he/she obtained 0.5 point. If he/she did not move or described verbally his/her behaviour when he/she felt the emotion, he/she obtained 0 point.

Identification of emotions expressed by the demonstrator's key poses:

As the NAO robot does not have facial expressions, we investigated emotions recognition when emotions were expressed with the body. We selected the four key poses exhibited by the NAO robot that were the most successfully identified by adults and children in previous studies (Beck et al., 2010; Beck et al., 2013): joy, fear, anger and sadness. The demonstrator struck each of these four key poses and between each pose, she went back to a neutral pose (Supplementary Information, Figure 7).

For this task, the demonstrator said "*Let me show you what I am doing when I am feeling an emotion*". The demonstrator reproduced one of the four emotional poses and the experimenter asked the child "*Which emotion do you think it is?*" After the child replied, the

demonstrator returned to the neutral pose. She repeated this sequence for the three other body postures. If the child identified the correct posture, he/she obtained 1 point (4 points in total).

2.2.4.4. Spontaneous helping situation

As presented in Table 2, this task ended each block of two consecutive sessions. The demonstrator and the child were seated in front of each other. The experimenter gave a blue stick to the demonstrator while saying "*Look, I have something for you*". The demonstrator grabbed it, looked at it and let it fall on the ground, expressing her surprise and saying "*Oh, it fell!*". Then, she tried to reach the stick with her hand. We calculated the latency between the stick falling and the first time the child touched the stick to give it back to the demonstrator. If the child did not touched the stick after 10 s, he/she obtained the score of 11 s.

At the end the session, the demonstrator thanked the child for his/her participation and told him/her goodbye.

2.2.5. Ethogram

We analysed child's behaviours when the demonstrator was talking to the child using the software BORIS v.7.9 (Supplementary Table 3 in Supplementary Information). The script in Supplementary Information details all sentences pronounced by the demonstrator.

2.3. Statistical analysis

Concerning scores, some data did not follow a normal distribution (Shapiro-Wilk tests: $p < 0.05$). We compared results obtained for the first and the second sessions for each condition using Wilcoxon signed-rank tests for paired data when data were not normally distributed and paired-sample Student t-tests when they were. We found no difference between session 1 and session 2 with the robot demonstrator and between session 1 and 2 with the human demonstrator (Wilcoxon signed-rank tests and Student t-tests: $p > 0.05$). We then merged results from sessions 1 and 2 with the robot demonstrator and from sessions 1 and 2 with the human to test the difference between the two conditions using Wilcoxon tests for paired data.

Moreover, we removed sequence#1 of the imitation subtask from the statistical analysis as we to use it to explain instructions again to children. We kept the second two-body parts sequence (sequence#2) and the three-body parts sequence only (sequence#3). As for the same condition, we found no difference concerning the success of imitation between sequence#2 and sequence#3 (Wilcoxon test for paired data: robot condition: V = 30, p = 0.10, human condition: V = 13, p = 0.66). We merged results from sequence#2 and sequence#3 for analysis. We followed the same procedure for gaze times.

3. Results

3.1. Scores

3.1.1. Body parts task

3.1.1.1. Comprehension of body parts labels

Identification of body parts on the child's body was high for all children in both conditions: 94.70% ± 0.68 (mean ± SE) for the robot condition, 93.18% ± 1.27 (mean ± SE) for the human condition. We found no difference between the conditions (Wilcoxon test for paired data: V = 4.5, p = 0.58). Five out of 6 children had difficulties, for at least one session, to show their elbow and they showed their arm, their shoulder, their armpit or their neck instead. The child E4 obtained 100% of correct answers for the two sessions.

Identifications of body parts on the demonstrator's body were also high for all the children in both conditions: 95.45% ± 1.31 (mean ± SE) for the robot condition, 87.12% ± 4.09 (mean ± SE) for the human condition. We found no difference between the conditions (Wilcoxon test for paired data: V = 6, p = 0.18). Results were widely influenced by the behaviour of the child E3 during the first session of the human condition. He did not obtain any point because he did not follow the instruction and showed body parts on his own body even if the experimenter repeated it. Elbow was also the body part that was the most difficult to identify but scores were better than on the child's body: 3 children showed their knee or their neck

instead of their elbow. Children E1 and E4 obtained 100% of correct answers for the two sessions.

3.1.1.2. Imitation of body parts' sequences

Body parts shown by the robotic demonstrator were significantly more complicated to label than body parts shown by the human demonstrator (robot condition: 52.78% ± 4.21 (mean ± SE), human condition: 84.72% ± 3.43, Wilcoxon test for paired data: V = 64.5, p = 0.006).

In the robot condition, head and hand were labelled with the highest rate (head: 75% ± 30.62; hand: 72.73% ± 26.86 (mean ± SE)). Shoulder was never identified and belly and eye were labelled with the lowest rate (belly: 14.29% ± 26.46; eye: 28.57% ± 34.15).

In the human condition, mouth, head, hand and foot were named by all children. Shoulder, eye, knee and belly were recognized between 60 and 90% (shoulder: 87.50% ± 23.39; eye: 85.71% ± 26.46; knee: 66.67% ± 54.43; belly: 60% ± 43.82, mean ± SE). Elbow is the body part labelled with the lowest rate (50% ± 40.82).

Imitation of body parts was lower for the robot condition (robot condition: 48.61% ± 5.42, mean ± SE, human condition: 56.94% ± 5.50). However, we found no statistical difference in imitation between human and robot conditions (Wilcoxon test for paired data: V = 31.5, p = 0.31).

The respect of body parts order in sequences was significantly higher in the human condition rather than in the robot condition (robot condition: 35.42% ± 5.26, mean ± SE, human condition: 55.56% ± 5.63, Wilcoxon test for paired data: V = 44, p = 0.01).

3.1.2. Emotions task

3.1.2.1.Identification on images

All children recognized correctly the four emotions on images, regardless of the condition.

3.1.2.2.Expression of emotions by the child

We found no difference between the two conditions concerning expression of emotions (robot condition: 70.83% ± 5.24, (mean ± SE), human condition: 66.67% ± 5.12, Wilcoxon test for paired data: V = 22.5, $p = 0.57$). We observed a high inter-individual variability. Child E4 and E2 obtained the highest scores, they expressed all emotions, regardless the condition, with their whole body and face. Children E5 and E6 obtained the lowest scores as they had difficulties expressing emotions with their body and they described their reactions or gave examples of situations instead. For example, the child E5 said “When I am sad, I use my eyes to cry” and “When I am sad, I use my legs to walk slowly”. The child E6 said “I am sad when someone hits me” and “I am sad when I feel lonely”.

3.1.2.3. Identification of emotional key poses on the demonstrator

On average, less than half of the demonstrator’s emotional postures were correctly identified (robot condition: 45.83% ± 5.29 (mean ± SE). human condition: 47.92% ± 3.59). We found no difference between the two conditions (Wilcoxon test for paired data: V = 22, $p = 1$). As shown in Figure 8A, joy was recognized in 58.33% ± 28.46 (mean ± SE) of the robot condition sessions and in 41.67% ± 28.46 (mean ± SE) of the human condition sessions. Sadness was recognized in 25% ± 25 of the robot sessions and in 58.33% ± 28.46 of the human sessions (Figure 8B). Anger was equally recognized in both conditions (50% ± 28.87, Figure 8C). Fear was the less recognized one as children identified it in 33.33% ± 27.22 (mean ± SE) of the robot sessions and only in 25% ± 25 (mean ±SE) of the human sessions (Figure 8D).

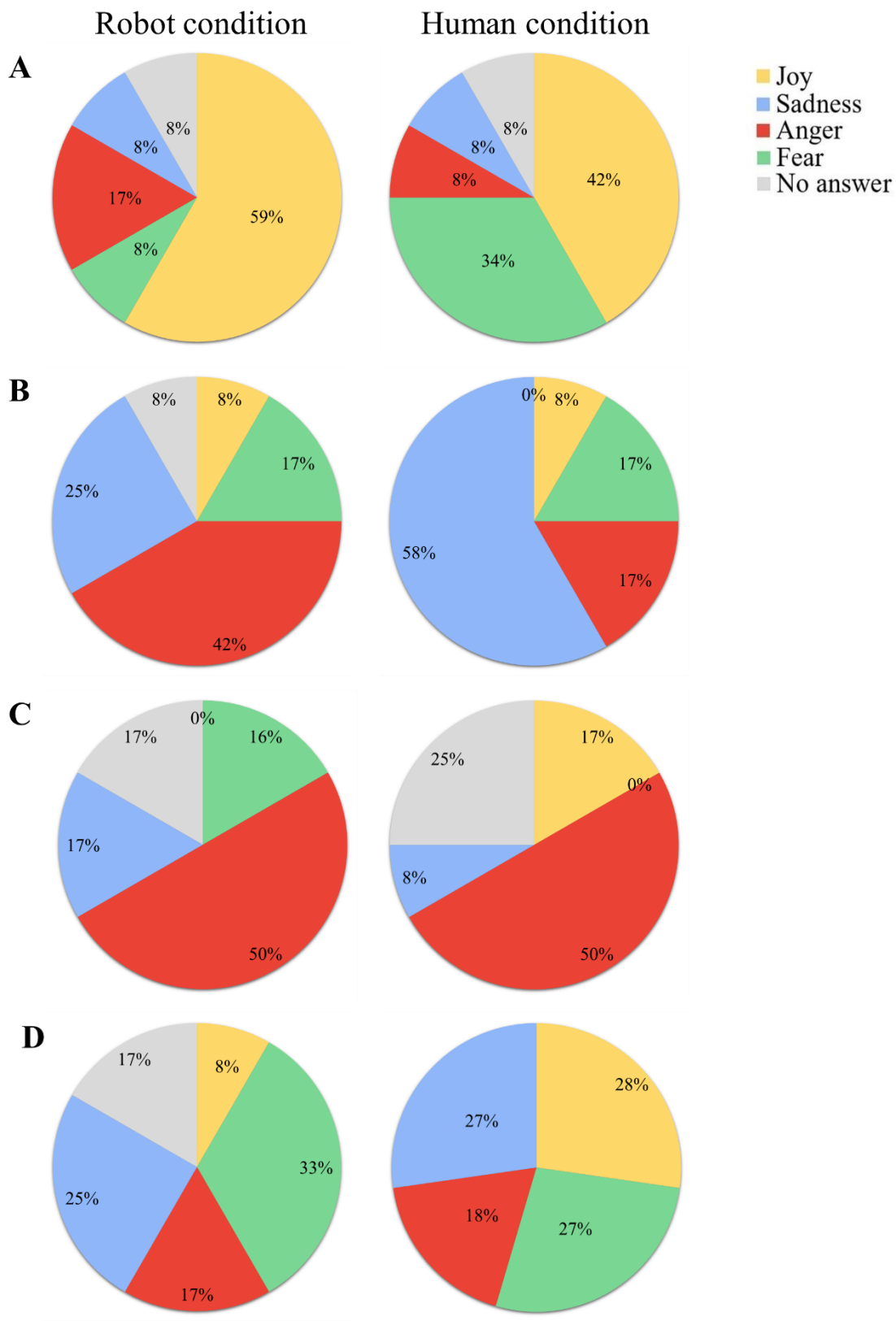


Figure 8. Identification of emotional body postures in the robot condition and in the human condition: (A) Joy, (B) Sadness, (C) Anger and (D) Fear.

3.1.3 Spontaneous helping situation

Reaction times to give back the stick to the demonstrator were less than 5 s in both conditions (robot condition: 3.81 s ± 0.82 (mean ± SE), human condition: 2.71 s ± 0.84) except for the child E3 (Figure 9). He was the only one who did not touch the stick during the 10 s. The child E6 touched the stick less than 1 s after it fell (robot condition: 0.99 s, human condition: 0.53 s). We found no difference between the two conditions (human vs. robot, Wilcoxon test for paired data: V = 14, p = 0.56).

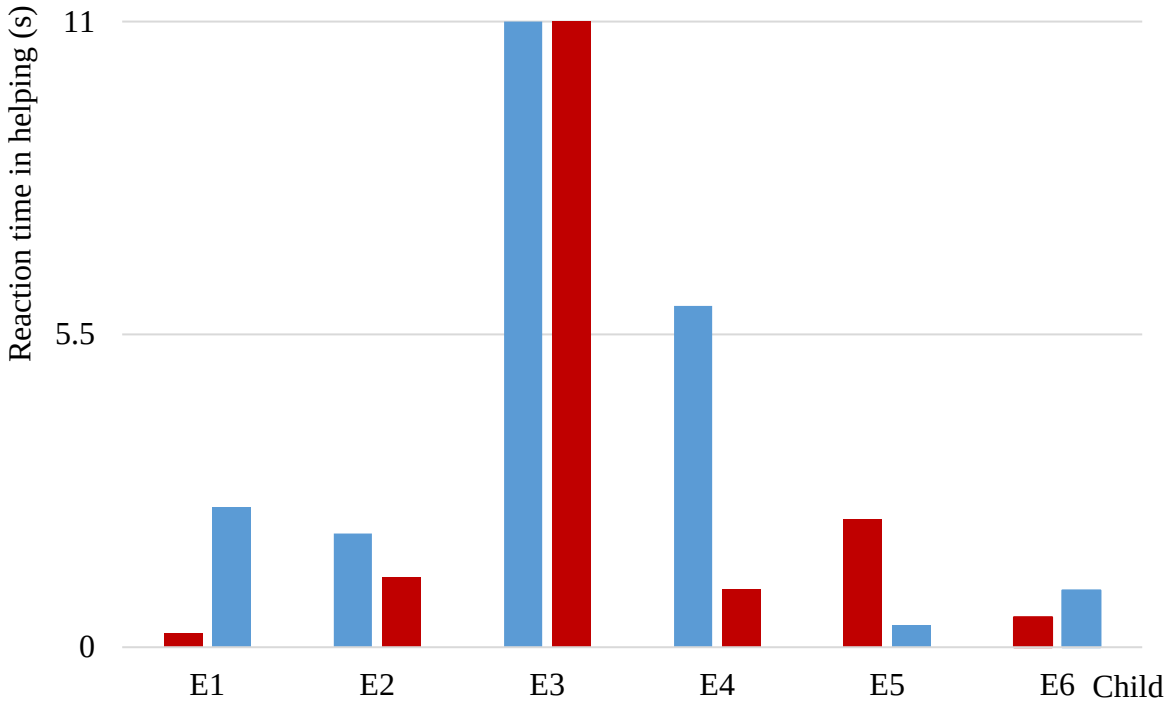


Figure 9. Reaction time of the 6 ASD-children (left to right: from the youngest to the oldest) in helping situations for the robot condition (blue) and the human condition (red) in s.

3.2. Eye gaze

In both conditions, children looked at the demonstrator more than 50% of the duration of the session. They looked significantly more at the demonstrator when it was the robot than

when it was the human (robot condition: 71.51% ± 1.24 (mean ± SE), human condition: 59.16% ± 1.41; Wilcoxon test for paired data: V = 68, p = 0.02).

We compared children's gaze times directed to the demonstrator task by task. We found that they looked significantly more at the demonstrator when it was the robot than when it was the human during the warm-up phase (robot condition: 91.69% ± 1.04 (mean ± SE), human condition: 76.48% ± 2.09; Wilcoxon test for paired data: V = 75, p = 0.002), during the imitation task (robot condition: 88.38% ± 1.19 (mean ± SE), human condition: 75.77% ± 2.82; Student t-test: t = -2.37, df = 11, p = 0.037) and for identification of emotional postures (robot condition: 81.60% ± 2.14 (mean ± SE), human condition: 62.95% ± 2.26; Student t-test: t = 3.44, df = 11, p = 0.005).

We found no difference in times spent looking at the demonstrator in the robot and in the human conditions for the body part task (Wilcoxon tests for paired data: on the child's body: V = 47, p = 0.23, on the demonstrator's body: V = 56, p = 0.20) and for the expression of emotions by the child (Wilcoxon test for paired data: V = 60, p = 0.11).

3.3. Other behaviours

3.3.1 Direction of gaze when the demonstrator questioned the child

During the experiment, the demonstrator asked 11 questions to the child (see Appendix 1 and Table 3 in Supplementary Information). The number of demonstrator's questions answered by the child while looking at the demonstrator was similar in both conditions (robot condition: 5 questions ± 0.26, human condition: 5 questions ± 0.06 (mean ± SE), Wilcoxon test for paired data: V = 7, p = 1). We found similar results for the number of demonstrator's questions answered while looking at the experimenter (robot condition: 0.50 question ± 0.16, human condition: 0.17 question ± 0.05 (mean ± SE), Wilcoxon test for paired data: V = 2, p = 0.35).

3.3.2 Number of tactile contacts

On average, children touched the human demonstrator 6.92 times ± 1.34 (mean ± SE) and the robot demonstrator 12.17 times ± 1.14 (mean ± SE) but we found no statistical difference between the two conditions (human vs. robot, Wilcoxon test for paired data: V = 1, p = 0.10, Figure 10).

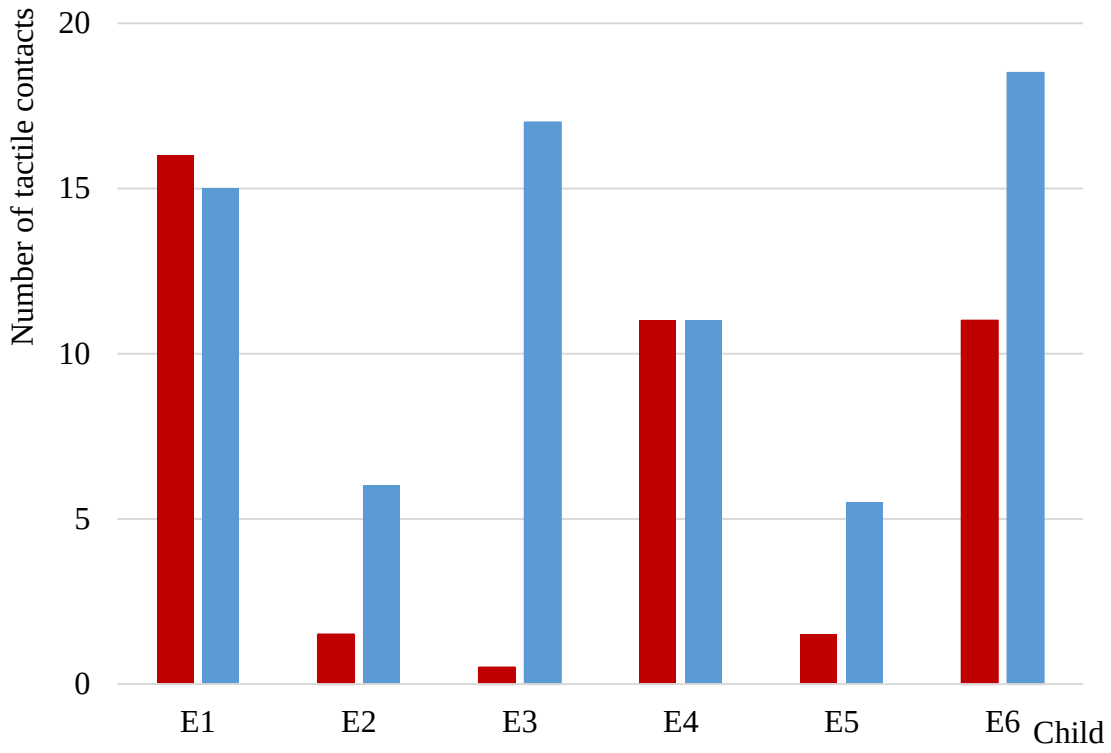


Figure 10. Number of times the child touched the demonstrator (left to right: from the youngest to the oldest) for the robot condition (blue) and the human condition (red).

4. Discussion

The purpose of this study was to compare ASD children reactions in three tasks: 1) body parts knowledge, 2) recognition and expression of emotions and 3) helping situation whether the instructor was either an unfamiliar adult or the NAO robot. Each child participated in four sessions, 2 in the human condition and 2 in the robot condition, in a randomly assigned order.

As already mentioned, since we did not find any difference between the two sessions for each condition, we merged results to compare reactions in the human condition to those in the robot condition. We then compared the two conditions and found no effect of the demonstrator on children's performances except for two subtasks: the production of body parts labels and the respect of order in the sequence of body parts.

Children showed high level of knowledge about their body structural representation and the demonstrator's – human or robotic – one as they correctly identified around 90% of body parts. Elbow was the less identified body part on children's body and on the demonstrator's body, regardless the condition.

In the first part of the imitation task, the percentage of correct body parts labels was, on average, lower than in the body parts task. Children had significantly more difficulties to name body parts shown by the robotic demonstrator rather than by the human demonstrator. In the robot condition, head and hand were named with the highest rate whereas belly and eye with the lowest rate and shoulder was never identified. We hypothesis that the NAO robot cannot efficiently show its body parts compared to the human demonstrator because its hands are big compared to its other body parts. In the human condition, mouth, head, hand and foot were recognized by all children. Elbow is the body part recognized with the lowest rate.

Taking together, these results are consistent with the study of Camões-Costa and colleagues (2011) on typical developing children under 3 and a half years old. Authors demonstrated that head and hand are part of the ten - out of 50 - most recognised body parts and that joints - as elbow - are the least well-known body parts.

We found no difference concerning the percentage of correct body parts imitated between the two conditions. Children respected more the order of body parts in the sequence when they imitated in the human condition than in the robot condition. We hypothesis that it was easier to name body parts and to memorize the order of body parts in the sequence because

the body of the human demonstrator is more similar to child's body compared to the body of the robotic demonstrator. The absence of difference concerning imitation is coherent with the study of Schleihauf and colleagues (2021) in children from 5 to 6 years old. Authors compared over imitation, the tendency to imitate irrelevant actions to achieve a precise goal (in this study, retrieving marbles from a puzzle box), when the model was either an unfamiliar adult or the NAO robot. They found that children equally over imitated both demonstrators.

Concerning the emotions task, all children correctly pointed at the four emotions (joy, fear, anger, sadness) on images regardless of the condition. Widen & Russel (2003) demonstrated that 3 years old typically developing children can successfully identify and name happiness, sadness, anger, fear, surprise and disgust (the six primary emotions described by Ekman & Friesen, 1976) on pictures. As emotion recognition is often evaluated based on facial expressions, many studies used the Ekman 60 faces test from the FEEST (Facial expressions of emotion: Stimuli and tests) stimulus battery (Young et al., 2002). However, in the present study, this subtask served as a baseline for the following subtask of emotions recognition on the demonstrator's body postures. To our knowledge, there is no standardized set of images showing children's emotions expressed both by the face and the body posture which leads us to use pictures presented in Supplementary Information (Figure 6).

Regarding the expression of emotions by children, we observed a high inter-individual variability and we found no difference between the two conditions (robot vs human). The meta-analysis of Riggio & Riggio (2002) demonstrated a strong relationship between emotional expressiveness – the ability to express emotions – and extraversion personality traits (Riggio & Riggio, 2002). We make the assumption that children's scores in this task were related to their personality and their capacity to feel comfortable (which we did not evaluated during the experiment) more than the demonstrator type.

The identification of emotional key poses was equivalent between the two conditions and less than half of the emotional postures were correctly identified on the demonstrator. Key poses for joy and sadness were the most correctly identified, followed by the key pose for anger and then the key pose for fear. These results contrast with higher rates found the Beck and colleagues (2013) in their study about 11 to 13 years old typically developed children. We hypothesis that young ASD children may need cues and context to interpret efficiently the demonstrator's body postures.

Finally, we found no difference regarding reaction times in the last task (i.e., to give back the stick to the demonstrator) between the two conditions. And children seem to understand that both the robot and the adult can't reach the stick. This result is coherent with the study of Dunfield and colleagues (2019) in which 4 years old ASD children helped a human demonstrator by giving him back a piece of puzzle out of reach.

In our study, we considered that both demonstrators looked at children most of the time. This was particularly controlled in the robot condition as NAO's gaze was fixed: it did not blink and its head was mostly orientated in the child's direction. As children with ASD often presented a deficit in social gaze (Groom and al., 2017; Scassellati et al., 2012), we compared time spent looking at the demonstrator in both conditions and in each sub task. In both conditions, children looked at the demonstrator more than 50% of the duration of the session and they looked significantly more at the demonstrator when it was the robot than when it was the human during the whole experiment and during the warm-up phase, the imitation task and the identification of emotional key poses. This is similar to the study reported by Tapus and colleagues (2012) who found more eye gazes and shared attention during children-NAO robot interactions compared to children-human interactions.

Taken together, our study provides evidences that our experimental design can be successful used with mild ASD children as they all interacted with both demonstrators from the

first session and they remained engaged across the four sessions. Our experimental protocol has the characteristics of most frequently used protocols in educational research with social robots: a one-to-one interaction where the NAO robot gave instructions to the child (Johal et al., 2020).

However, it also presents several limitations as frequently highlighted in this field: in their literature review, Begum and colleagues (2016) observed that only minimal progress had been made in the development of robots as clinical tools for ASD intervention and that clinicians are generally not convinced about the usefulness of social robots. They argued that the majority of human-robot interactions studies do not follow standards of clinical research for ASD intervention (Begum et al., 2016). To overcome this problematic situation and improve future research in the field of ASD child-robot interaction, Raptopoulou and colleagues (2021) suggest to follow several guidelines: 1) before starting experiments, assessing ASD children social skills using a methodology shared by several researchers in order to facilitate comparison between studies; 2) developing a standard experimental protocol based on previous studies; 3) ensuring a minimal sample size of 12 children and the presence of a control group of typically developing children; 4) anticipating conducting a minimum of 10 sessions per child and 5) considering conducting follow up sessions 1 to 3 months after the intervention and then a year after the first intervention. The necessity to conduct longitudinal studies to ensure that results are valid is also crucial for Damianidou and colleagues (2020).

Following Raptopoulou and colleagues' framework (2021), we will further replicate this study in a longer period of time to evaluate notably whether the demonstrator can help children to learn body parts and emotions they did not know at the beginning of the experiment. We will also include more ASD children with different profiles. Moreover, all tasks proposed in our study were also evaluated in a group of children with typical development, aged from 3 to 6 years old, which allows us future comparisons.

Acknowledgements

We would like to thank all the children who participated as well as the educative team of the Educational and Care Service.

References

American Psychiatric Association, DSM-5 Task Force. (2013). Diagnostic and statistical manual of mental disorders:DSM-5™. American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425596

Bacon, A. L., Fein, D., Morris, R., Waterhouse, L., & Allen, D. (1998). The responses of autistic children to the distress of others. *Journal of Autism and Developmental Disorders*, *28*(2), 129-142. https://doi.org/10.1023/a:1026040615628

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*, 3567-3574. https://doi.org/10.1098/rstb.2009.0191

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The « Reading the Mind in the Eyes » Test revised version : A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 241-251.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2000). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning Disorders*, *5*.

Beck, A., Cañamero, L., & Bard, K. (2010). *Towards an Affect Space for robots to display emotional body language*. 464-469. https://doi.org/10.1109/ROMAN.2010.5598649

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of Emotional Body Language Displayed by a Humanoid Robot : A Case Study with Children. *International Journal of Social Robotics*, *5*(3), 325-334. https://doi.org/10.1007/s12369-013-0193-z

Begum, M., Serna, R. W., & Yanco, H. A. (2016). Are Robots Ready to Deliver Autism Interventions? A Comprehensive Review. *International Journal of Social Robotics*, *8*(2), 157-181. https://doi.org/10.1007/s12369-016-0346-y

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education : A review. *Science Robotics*, *3*(21). https://doi.org/10.1126/scirobotics.aat5954

Bergès, J., & Lézine, I. (1963). *Test of imitation of gestures: Methods for plotting body image and movements of children from 3 to 6 years of age*. Masson & Cie.

Bird, G., Leighton, J., Press, C., & Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1628), 3027-3031. https://doi.org/10.1098/rspb.2007.1019

Cervantes, P. E., Matson, J. L., Adams, H. L., Williams, L. W., Goldin, R. L., & Jang, J. (2013). Comparing social skill profiles of children with autism spectrum disorders versus children with attention deficit hyperactivity disorder : Where the deficits lie. *Research in Autism Spectrum Disorders*, *7*(9), 1104-1110. https://doi.org/10.1016/j.rasd.2013.05.008

Camões-Costa, V., Erjavec, M., & Horne, P. J. (2011). The impact of body-part-naming training on the accuracy of imitative performances in 2- to 3-year-old children. *Journal of the Experimental Analysis of Behavior*, *96*(3), 291-315. https://doi.org/10.1901/jeab.2011.96-291

Conson, M., Mazzarella, E., Esposito, D., Grossi, D., Marino, N., Massagli, A., & Frolli, A. (2015). « Put myself into your place » : Embodied simulation and perspective taking in autism spectrum disorders. *Autism Research*, *8*(4), 454-466. https://doi.org/10.1002/aur.1460

Costa, S., Lehmann, H., Dautenhahn, K., Robins, B., & Soares, F. (2015). Using a Humanoid Robot to Elicit Body Awareness and Appropriate Physical Interaction in Children with Autism. *International Journal of Social Robotics*, *7*, 1-14. https://doi.org/10.1007/s12369-014-0250-2

Damianidou, D., Arthur-Kelly, M., Foggett, J., Beh, E., & Eidels, A. (2020). Associating Cognitive Functions with Technology Features Used to Support Employment for People with Intellectual and Developmental Disability. *Advances in Neurodevelopmental Disorders*, *4*. https://doi.org/10.1007/s41252-020-00164-9

Damm, O., Malchus, K., Jaecks, P., Krach, S., Paulus, F., Naber, M., Jansen, A., Kamp-Becker, I., Stenneken, P., & Wrede, B. (2013). Different gaze behavior in human-robot interaction in Asperger's syndrome : An eye-tracking study. *International Symposium on Robot and Human Interactive Communication*. https://pub.uni-bielefeld.de/record/2614705

Dunfield, K. A., Best, L. J., Kelley, E. A., & Kuhlmeier, V. A. (2019). Motivating Moral Behavior : Helping, Sharing, and Comforting in Young Children With Autism Spectrum Disorder. *Frontiers in Psychology*, *10*, 25. https://doi.org/10.3389/fpsyg.2019.00025

Duquette, A., Michaud, F., & Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, *24*(2), 147-157. https://doi.org/10.1007/s10514-007-9056-5

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56-75. https://doi.org/10.1007/BF01115465

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, *43*(5), 1038-1049. https://doi.org/10.1007/s10803-012-1645-2

Groom, M. J., Kochhar, P., Hamilton, A., Liddle, E. B., Simeou, M., & Hollis, C. (2017). Atypical Processing of Gaze Cues and Faces Explains Comorbidity between Autism Spectrum Disorder (ASD) and Attention Deficit/Hyperactivity Disorder (ADHD). *Journal of Autism and Developmental Disorders*, *47*(5), 1496-1509. https://doi.org/10.1007/s10803-017-3078-4

Hale, C. M., & Tager-Flusberg, H. (2005). Social communication in children with autism : The relationship between theory of mind and discourse development. *Autism: The International Journal of Research and Practice*, *9*(2), 157-178. https://doi.org/10.1177/1362361305051395

Hobson, R. P. (1986a). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *27*(3), 321-342. https://doi.org/10.1111/j.1469-7610.1986.tb01836.x

Hobson, R. P. (1986b). The autistic child's appraisal of expressions of emotion : A further study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *27*(5), 671-680. https://doi.org/10.1111/j.1469-7610.1986.tb00191.x

Jameel, L., Vyas, K., Bellesi, G., Roberts, V., & Channon, S. (2014). Going 'above and beyond' : Are those high in autistic traits less pro-social? *Journal of Autism and Developmental Disorders*, *44*(8), 1846-1858. https://doi.org/10.1007/s10803-014-2056-3

Johal, W. (2020). Research Trends in Social Robots for Learning. *Current Robotics Reports*, *1*, 1-9. https://doi.org/10.1007/s43154-020-00008-3

Kennedy, J., Baxter, P., & Belpaeme, T. (2015). Comparing Robot Embodiments in a Guided Discovery Learning Interaction with Children. *International Journal of Social Robotics*, *7*(2), 293-308. https://doi.org/10.1007/s12369-014-0277-4

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., Senft, E., & Belpaeme, T. (2017). Child Speech Recognition in Human-Robot Interaction : Evaluations and Recommendations. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 82-90. https://doi.org/10.1145/2909824.3020229

Kjellmer, L., Hedvall, Å., Fernell, E., Gillberg, C., & Norrelgen, F. (2012). Language and communication skills in preschool children with autism spectrum disorders : Contribution of cognition, severity of autism symptoms, and adaptive functioning to the variability. *Research in Developmental Disabilities*, *33*(1), 172-180. https://doi.org/10.1016/j.ridd.2011.09.003

Lee, J., & Obinata, G. (2013). *Developing therapeutic robot for children with autism : A study on exploring colour feedback*. 173-174. https://doi.org/10.1109/HRI.2013.6483557

Lee, J., Takehashi, H., Nagai, C., & Obinata, G. (2012). Design of a therapeutic robot for interacting with autistic children through interpersonal touch. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/ROMAN.2012.6343835

Marion, M. (2003). Guidance of young children. Merrill Prentice Hall, Columbus.

Mashal, N., & Kasirer, A. (2012). Principal component analysis study of visual and verbal metaphoric comprehension in children with autism and learning disabilities. *Research in Developmental Disabilities*, *33*(1), 274-282. https://doi.org/10.1016/j.ridd.2011.09.010

Matson, J. L., & Wilkins, J. (2007). A critical review of assessment targets and methods for social skills excesses and deficits for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *1*(1), 28-37. https://doi.org/10.1016/j.rasd.2006.07.003

Michaud, F., Salter, T., Duquette, A., Mercier, H., Lauria, M., Larouche, H., & Larose, F. (2007). *Assistive Technologies and Child-Robot Interaction*. https://doi.org/10.13140/RG.2.1.3751.3765

Palestra, G., Bortone, I., Cazzato, D., Adamo, F., Argentiero, A., Agnello, N., & Distante, C. (2014). *Social Robots in Postural Education : A New Approach to Address Body Consciousness in ASD Children*. 290-299. https://doi.org/10.1007/978-3-319-11973-1_30

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics : A systematic review. *Autism Research: Official Journal of the International Society for Autism Research*, *9*(2), 165-183. https://doi.org/10.1002/aur.1527

Pierno, A. C., Mari, M., Lusher, D., & Castiello, U. (2008). Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, *46*(2), 448-454. https://doi.org/10.1016/j.neuropsychologia.2007.08.020

Raptopoulou, A., Komnidis, A., Bamidis, P. D., & Astaras, A. (2021). Human-robot interaction for social skill development in children with ASD : A literature review. *Healthcare Technology Letters*, *8*(4), 90-96. https://doi.org/10.1049/htl2.12013

Riggio, H. R., & Riggio, R. E. (2002). Emotional Expressiveness, Extraversion, and Neuroticism : A Meta-Analysis. *Journal of Nonverbal Behavior*, *26*(4), 195-218. https://doi.org/10.1023/A:1022117500440

Romanczyk, R., White, S., & Gillis, J. (2005). *Social Skills versus Skilled Social Behavior : A Problematic Distinction in Autism Spectrum Disorders*. https://doi.org/10.1037/H0100312

Saleh, M. A., Hanapiah, F. A., & Hashim, H. (2020). Robot applications for autism : A comprehensive review. *Disability and Rehabilitation. Assistive Technology*, *16*(6), 580-602. https://doi.org/10.1080/17483107.2019.1685016

Scassellati, B., Admoni, H., & Matarić, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, *14*, 275-294. https://doi.org/10.1146/annurev-bioeng-071811-150036

Scassellati, B., Boccanfuso, L., Huang, C.-M., Mademtzi, M., Qin, M., Salomons, N., Ventola, P., & Shic, F. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, *3*(21), eaat7544. https://doi.org/10.1126/scirobotics.aat7544

Schleihauf, H., Hoehl, S., Tsvetkova, N., König, A., Mombaur, K., & Pauen, S. (2021). Preschoolers' Motivation to Over-Imitate Humans and Robots. *Child Development*, *92*(1), 222-238. https://doi.org/10.1111/cdev.13403

Shamsuddin, S., Yusoff, H., Ismail, L., Mohamed, S., Hanapiah, F., & Zahari, N. (2012). Humanoid Robot NAO Interacting with Autistic Children of Moderately Impaired Intelligence to Augment Communication Skills. *Procedia Engineering 2012;41:1533-1538*, *2012*, 1533-1538. https://doi.org/10.1016/j.proeng.2012.07.346

Shamsuddin, S., Yussof, H., Mohamed, S., Hanapiah, F. A., & Ismail, L. I. (2013). Stereotyped behavior of autistic children with lower IQ level in HRI with a humanoid robot. *2013 IEEE Workshop on Advanced Robotics and its Social Impacts*. https://doi.org/10.1109/ARSO.2013.6705525

Sirigu, A., Grafman, J., Bressler, K., & Sunderland, T. (1991). Multiple representations contribute to body knowledge processing. Evidence from a case of autotopagnosia. *Brain: A Journal of Neurology*, *114 (Pt 1B)*, 629-642. https://doi.org/10.1093/brain/114.1.629

Srinivasan, S. M., Pescatello, L. S., & Bhat, A. N. (2014). Current perspectives on physical activity and exercise recommendations for children and adolescents with autism spectrum disorders. *Physical Therapy*, *94*(6), 875-889. https://doi.org/10.2522/ptj.20130157

Suzuki, R., & Lee, J. (2016). Robot-play therapy for improving prosocial behaviours in children with Autism Spectrum Disorders. *2016 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*. https://doi.org/10.1109/MHS.2016.7824238

Suzuki, R., Lee, J., & Rudovic, O. (2017). *NAO-Dance Therapy for Children with ASD*. 295-296. https://doi.org/10.1145/3029798.3038354

Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., Rusu, A. S., & David, D. O. (2012). Children with autism social engagement in interaction with Nao, an imitative robot : A series of single case experiments. *Interaction Studies*, *13*(3), 315-347. https://doi.org/10.1075/is.13.3.01tap

Wainer, J., Dautenhahn, K., Robins, B., & Amirabdollahian, F. (2014a). A Pilot Study with a Novel Setup for Collaborative Play of the Humanoid Robot KASPAR with Children with Autism. *International Journal of Social Robotics*, *6*, 45-65. https://doi.org/10.1007/s12369-013-0195-x

Wainer, J., Robins, B., Amirabdollahian, F., & Dautenhahn, K. (2014b). Using the Humanoid Robot KASPAR to Autonomously Play Triadic Games and Facilitate Collaborative Play Among Children With Autism. *IEEE Transactions on Autonomous Mental Development*. https://doi.org/10.1109/TAMD.2014.2303116

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, *39*(1), 114-128. https://doi.org/10.1037//0012-1649.39.1.114

White, S.W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders : A review of the intervention research. *Journal of Autism and Developmental Disorders*, *37*(10), 1858-1868. https://doi.org/10.1007/s10803-006-0320-x

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). Facial expressions of emotion : Stimuli and tests (FEEST). *Thames Valley Test Company (TVTC)*.

Supplementary Information

Appendix 1. Script of the experiment

The experimenter installs the child in front of the demonstrator (Alice is the name of the human demonstrator, Naomie is the name given to the robotic demonstrator).

Warm-up phase

- Hello, my name is Naomie (/Alice) and you, what is your name? **(Question 1)**
- Do you know the song "Petit moulin"? **(Question 2)**
- I will sing it and if you want, you can sing it with me.
- "Tourne, tourne, petit moulin. Tapent, tapent, petites mains. Vole, vole, petit oiseau. Nage, nage, petit poisson. Petit moulin a bien tourné, petites mains ont bien tapé, petit oiseau a bien volé, petit poisson a bien nagé".

Body task

Comprehension of body parts labels:

- Now, you have to show me one of your body part. To start, show me your [body part] (repeated for face, eye, nose, mouth, ear, shoulder, elbow, hand, belly, knee, foot)
- Now, show me my [body part](repeated for face, eye, nose, mouth, ear, shoulder, elbow, hand, belly, knee, foot)

Imitation of the sequence of body parts:

- I will show you several body parts on my body. Look carefully because you will have to do the same thing after.
- The demonstrator randomly showed one of his body part, for example, his eye and asked "What is the name of this body part?" **(Questions 3, 5, 7)** and a second one, for example his belly, and said "And this one?" **(Questions 4, 6, 8 and 9)**. Then, the demonstrator placed his hands on his knees and said "This is you turn".
-

Identification of emotions:

- Do you know what an emotion is? **(Question 10)**
- We will speak about 4 emotions: anger, joy, sadness and fear. Show me the image in which the child feels [emotion].
- I stand up for the rest of the exercise. Please, stand up too.
- Then, the demonstrator and he child stood up. The demonstrator asked “Show me how do you express [emotion] with your whole body”.
- What are the 4 emotions we talked about? **(Question 11)**
- Let me show you what I am doing when I am feeling an emotion.

The demonstrator reproduced one of the four emotional poses and the experimenter asked the child “Which emotion do you think it is?” After the child replied, the demonstrator returned to the neutral pose. The demonstrator repeated this sequence for the three other body postures.

Spontaneous helping situation

- The demonstrator and the child were seated in front of each other. The experimenter gave a blue stick to the demonstrator while saying “Look, I have something for you”. The demonstrator grabbed it, looked at it and let it fall on the ground while saying “Oh, it fell!”. She tried to reach the stick with his hand.
- At the end of the experiment, the demonstrator said “Thank you, goodbye!

Figure 5. The 9 body parts pointed by the demonstrator in the imitation task: (A) human and (B) robotic.

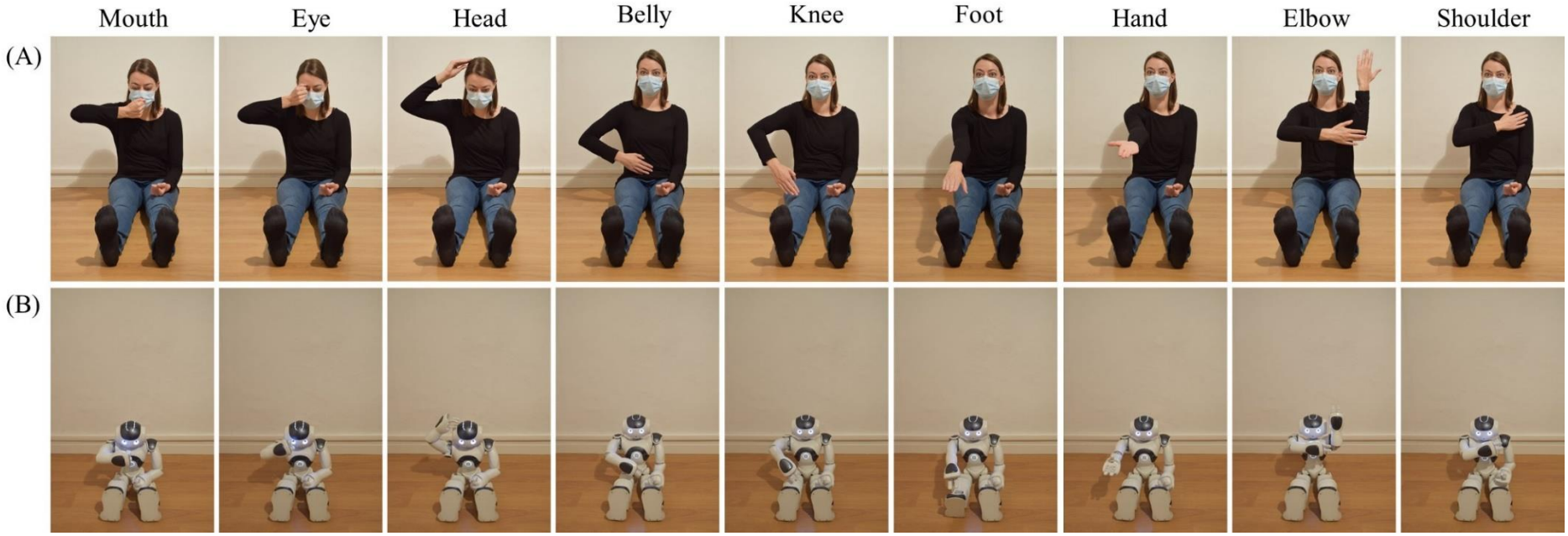


Figure 6. Images used in the task of emotions' identification (source: https://fr.freepik.com/).



Figure 7. The neutral and emotional postures exhibited by the demonstrator: (A) human and (B) robotic.

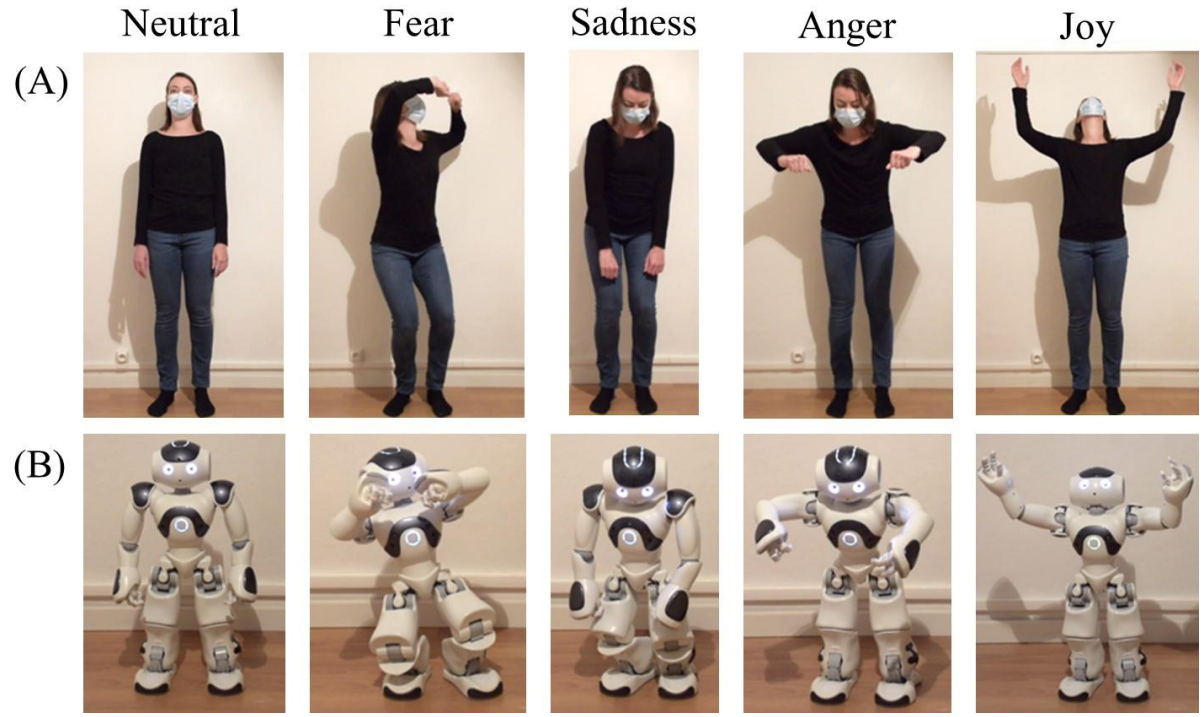


Table 3. Ethogram used on BORIS software to analyse child's behaviours. Events with a duration inferior to 1s were excluded.

Code	Behaviour	Description
a	The demonstrator talks	Moments where the demonstrators is talking to the child, in s
Where does the child look when the demonstrator is talking?		
q	at the demonstrator	Child gaze is orientated towards the demonstrator, in s
s	at the experimenter	Child gaze is orientated towards the experimenter, in s
e	at the touch-pad screen, the images or the blue stick	Child gaze is orientated towards the objects, in s
d	anywhere else	Child gaze is orientated anywhere else, in s
During the experiment, the demonstrator asks 11 questions to the child. What is the child's reaction?		
f	No answer	The child does not respond to the demonstrator question
g	Answering while looking at the demonstrator	The child answers to the demonstrator while looking at her
h	Answering while looking at the experimenter	The child answers to the demonstrator while looking at the experimenter
j	Answering while looking in another direction	The child answers to the demonstrator while looking in another direction
t	The child touches the demonstrator	Number of times the child touches the demonstrator

CHAPITRE III

Design of a robotic zebra finch for experimental studies on developmental song learning



© Louisane Araguas

Design of a robotic zebra finch for experimental studies on developmental song learning

Alice Araguas [1*], Bahia Guellaï[1], Philippe Gauthier [2¥], Florian Richer[2¥], Guglielmo Montone[3] and Sébastien Derégnaucourt [1, 4]

[1] *Laboratoire Ethologie Cognition Développement, Université Paris Nanterre, 200 Avenue de la République, F92001 Nanterre Cedex, France*

[2] *Institut des Systèmes Intelligents et de Robotique, CNRS UMR 7222, Sorbonne Université, France*

[3]*Laboratoire Psychologie de la Perception, CNRS UMR 8242, Université Paris Descartes, France*

[4] *Institut Universitaire de France*

* Correspondence: Sébastien Derégnaucourt, Laboratoire Ethologie Cognition Développement, UPL, Univ Paris Nanterre, F92000 Nanterre, France. E-mail address: sebastien.deregnaucourt@parisnanterre.fr

¥ Contributed equally to this work

Abstract

Birdsong learning has been consolidated as the model system of choice for exploring the biological substrates of vocal learning. In the Zebra Finch (*Taeniopygia guttata*), only males sing and they develop their song during a sensitive period in early life. Different experimental procedures have been used in the laboratory to train a young finch to learn a song. So far, the best method to get a faithful imitation is to keep a young bird singly with an adult male. Here we present the different characteristics of a robotic zebra finch that was developed with the goal to be used as a song tutor. The robot is morphologically similar to a real-size finch: it can produce movements and sounds contingently to the behaviours of a live bird. We present preliminary results regarding the acceptance of the robot by live birds and other possible applications beyond the scope of developmental song learning.

Keywords: animal behaviour, robotic bird, vocal contingency, multimodal communication, birdsong, developmental learning, *Taeniopygia guttata*

1. Introduction

Vocal production learning is defined as ‘instances where the signals themselves are modified in form as a result of experience with those of other individuals’ (Janik & Slater, 2000). Humans share this ability with other animal species from few taxonomic groups including oscine songbirds. Birdsong has been established as a model of choice to study the behavioural, cellular and molecular aspects of vocal production learning (Aamodt et al., 2020).

The Zebra Finch (*Taeniopygia guttata*) is considered as the "flying mouse" of birdsong research. This tiny bird is very easy to breed in captivity; only males sing and they develop their short and stereotyped song (1-1.5sec) during a sensitive period of early life (25-90 days after hatching). They learn to sing by memorizing and imitating the songs of surrounding conspecifics, mainly adult males (see Derégnaucourt, 2011 for a review; Derégnaucourt & Gahr, 2013). Since the pioneering works of Immelmann (1969), thousands of zebra finches were reared in controlled environments to examine how their experience affects the development of their song. Birds raised in social isolation during the sensitive phase of learning, develop abnormal songs that still contain sounds characteristics of the species (Price, 1979).

The importance of visual and physical contacts on song learning has also been evidenced; young birds separated from their tutor by an opaque partition or a grid, produce a worst imitation than birds that can physically interact with it (Derégnaucourt, 2011). Different experimental procedures were used in the laboratory to train young zebra finches to learn a song. Unlike other species of songbirds, passive exposure of zebra finch song recordings results in a poor imitation of song models (Derégnaucourt et al., 2013). Self-eliciting exposure to the song model, using an operant conditioning procedure, induces significant learning but with a high inter-individual variability: if some birds learn significantly (even some of them produce a very faithful copy of the song model broadcast), others show no sign of learning (Derégnaucourt et al., 2005; Derégnaucourt et al., 2013). Inter-individual variability may have

advantages. For example, we have shown a relationship between song learning and vocal plasticity during sleep: birds that showed stronger post-sleep deterioration during development achieved a better final imitation (Derégnaucourt et al., 2005). But it also has disadvantages. Some experimental procedures are complex (e.g. pinealectomy, Derégnaucourt et al., 2012; gene transfection: Haesler et al., 2007) and regulations regarding animal experimentation require to reduce sample sizes. In the Zebra Finch, the best way to get a close-to-perfect copy of the song is to raise a young male with an adult male (Derégnaucourt et al., 2013). This type of learning is of course exceptional; in natural conditions, young males learn by imitating their song mainly from their father, but also by copying parts of other singing males living in their close environment (Derégnaucourt & Gahr, 2013). This young-adult dyadic situation used to obtain a close-to-perfect imitation of the song is a reference for birdsong research (Derégnaucourt, 2011). However, this method does not allow precise control of the different variables involved in song learning (e.g. control of the behaviour of the tutor, number of songs heard, etc ...). If this method stresses the importance of social factors on song learning in the Zebra Finch, one cannot answer questions about the nature of these factors: are they visual, multimodal or is this result due to attentional phenomena?

One solution is to broadcast videos of singing adult males to young finches but so far, this use seems not reliable enough to obtain a good song copy in pupils (Deshpande et al., 2014, Ljubičić et al., 2016). Over the years, the use of robotic models increases in animal behaviour studies with promising results in several animal species including birds. For example, in the field, male Satin Bowerbirds (*Ptilonorhynchus violaceus*) adjusted their courtship displays in response to different postures of robotic females (Patricelli et al., 2002). In the laboratory, a robot was used to study social attachment in young quail chicks (Jolly et al., 2016). So far, these robots were used as pre-programmed devices without contingency with the behaviour of the interacting live birds. The importance of such contingency in the context of birdsong learning

has already been evidenced in oscine songbirds including zebra finches. In a recent study, Carouso-Peck & Goldstein (2019) demonstrated that song imitation is influenced by a non-vocal, visual feedback given by female finches and known as a response to attractive songs. When a contingent feedback was experimentally given to young males while they were singing (the broadcast of a video of a female producing the posture), they further develop better imitations of the tutor's song heard during their early life than their brother exposed to the same video non-contingently (Carouso-Peck & Goldstein, 2019). Such contingency might be necessary for the robot to be accepted as a valid tutor in the context of birdsong learning.

In order to investigate these aspects, we developed a robotic zebra finch that sings, moves and reacts to the behaviour of live zebra finches. We present also potential applications beyond the scope of birdsong learning.

2. Materials and methods

2.1. Design of the robot

The robot (named MANDABOT) was designed to allow for the following movements: rotation of the body around a vertical axis, rotation of the head, tail wagging and beak opening. We used taxidermic mounts of adult male zebra finches to design both the size and the shape of the different body parts (Figure 1).



Figure 1. Blueprint of the robotic zebra finch.

We printed them with Multi Jet Printing on a Projet 3510SD from 3DSystems which uses UV lights to cure parts from a liquid resin (Visijet M3 Black). MANDABOT was hand painted with mixes of Prince August paintings (France) to fit the colours of the plumage of an adult male zebra finch. The body is actuated by a DC motor with an encoder from Maxon. The coils and the body motor are controlled via a TB6612FNG Dual Motor Driver Carrier board from Pololu. The head is actuated by an AM1020 stepper motor from Faulhaber which is controlled via a TMC2130 SilentStepStick driver from Trinamic. Movements of the beak and the tail are controlled by coils and magnets obtained from commercial versions of DigiBirds (Silverlit Toys Menufactory, HongKong, China). A 3-axis accelerometer (ADXL362) permits to detect contacts between the bird and the robot. The robot is fixed on a grey plastic box (16 x 8.5 x 8 x cm). On each side of the robot, there is a perch (5 cm) connected to load cells (CZL639HD from Phidgets). The robot is equipped with a speaker (AS01808MR from PUI Audio) to broadcast pre-recorded zebra finch songs and calls. We designed the robot control using a client/server approach over Wifi.

We developed a custom server program running on a Beaglebone Black Wireless (physically close to the robot) and a custom client program running a laptop Dell Latitude 5300 computer running on Microsoft Windows 10. The server program handles the low-level control: move physically every actuators of robot, read positions and rotations, detect physical events, and play sounds and is listening for commands using HTTP protocol and JSON format. The client program connects to the server and handles the high-level control: ask for positions-rotations, perch states, react, send commands, record events, movements and sound. This approach allows more modularity and complex control over the robot (one laptop could control several robots), while using the more powerful CPU of the laptop than the Beaglebone embedded with the physical robot. We programmed the following behaviours: when the bird lands on the perch, the robot turns in his direction and randomly moves its head or broadcasts

a song; when the bird leaves the perch, the body of the robot turns back to the centre in a neutral position. Both closed-loop and open-loop can be launched for a precise period of time (e.g. 60 min). At the end of the loop, the computer generates a script containing all the sequences of movements and vocalisations produced by the robot. A script recorded in a closed-loop interaction context between the robot and a bird can then be run in a non-interactive mode with another bird. For the preliminary experiments presented below, the robot was placed in an experimental cage (46 x 23 x 27 cm) in a sound-proof chamber (103 x 56.5 x 65cm).

2.2. Interactive vocal loops between the bird and the robot

We developed a closed-loop program to mimic the vocal interactions between two finches. Such bird-machine vocal interactions have already been successfully applied in oscine songbirds including zebra finches (Lerch et al., 2011; Benichov et al., 2016). Sounds from the experimental cage are continuously recorded using a Behringer C-2 microphone connected to a PreSonus AudioBox 1818VSL (16 bits, 44.1 kHz). The audio stream is stored in a buffer of 180ms which corresponds roughly to the maximum duration of a contact call produced by a zebra finch. In parallel, signal amplitude is calculated. If it exceeds the detection threshold, features of the input signal are computed. The value of this threshold must be set experimentally because it depends on the acoustic environment of the cage and also to the loudness of the vocalizing bird interacting with the robot. Since contact calls produced by finches are very brief (on average 110ms, Ter Maat et al., 2014), the descriptor was therefore built on time windows of 32 samples (0.7ms). Every 0.7 s, the signal received is convolved to a Hamming window, before calculating the logarithm of the corresponding spectrum as following:

$$feature(t,:) = \log|fft(signal)|$$

Before being introduced into the network, these features must be standardized. To do this, we resize the features at the minimum duration of the database calls (83 ms) and we

normalized each feature (115x17 matrix) between 0 and 1. The 2D image is then reorganized into a column vector (1x1955). To train the artificial neural network, we used 390 *.wav audio files, containing different calls of four different birds and noises produced by birds during movements (e.g. cage noises, wing flaps).

Following Rassak et al. (2016), the neural network used is a multi-layer perceptron with one hidden layer composed of 300 neurons with 2 neurons in the output layer corresponding to one of two possible categories: bird call or cage noise. If the sound belongs to the first category, then the system broadcast a pre-recorded call from the speaker placed inside the robot. In a preliminary experiment based on 397 sounds, we obtained a correct recognition score of 0.9747 for bird calls. When a vocal sound produced by the bird (either a call or a song syllable) is detected by the system, a pre-recorded zebra finch call is broadcast through the speaker. The vocal response of the robot occurs with a minimum latency of 280ms: 180ms of features extraction followed by 100ms of computation by the neural network. The duration of this latency, similar to those used by other research group (e.g. Benichov et al., 2016) could also be experimentally increased if needed. The recorded sound (either a cage noise or a bird call) is saved as a *wav. file in a corresponding folder for eventual offline analysis or later use. An open-loop control is also possible: the system can send a call chosen randomly in the call database, either at a regular or at a random interval.

3. Results and discussion

In an ongoing experiment, 12 young finches were exposed singly during 1h, 5 days a week during 5 weeks, from day post hatch 35, to a robot programmed to exhibit contingent behaviour (movements, calls and songs). Based on preliminary results, finches spent about 61.73 ± 6.02 [0.18-77.57] % (median ± SE [min-max], individual means across 10 sessions of exposure) of the time on its close vicinity (less than 10cm). They also exhibited a high interest to the robot by pecking different parts of its body (Table 1, Figure 2).

Robot parts	Median ± SE	[min-max]
Beak	20.4 ± 3.38	[0-33.6]
Head	8.9 ± 2.43	[0-29.3]
Body	22.3 ± 4.38	[0-46.5]
Tail	13.4 ± 3.98	[0-43.5]

Table 1. Number of pecks to the different parts of the robot. Analysis based on individual means across 10 sessions. N= 12 birds.



Figure 2. Video screenshots of young zebra finches interacting with the robot.

Other potential experiments could investigate the importance of multimodal aspects on developmental song learning. For example, young finches could be exposed to a motionless robot, or to a robot with its beak opening synchronously or asynchronously with sound production.

We could also evaluate the importance of behavioural and vocal contingencies by increasing the latency of the robot's responses. In male zebra finches, song quality depends on the social context: for example, when they sing to females, songs are faster and more stereotyped (directed songs) that when they sing alone without any audience (undirected song) (Sossinka & Böhner, 1980). Painted as a female, the robotic zebra finch could be used in this courtship context. The robotic zebra finch could also be used to investigate some aspects of social cognition such as gaze following (Butler & Fernández-Juricic, 2014). Overall, these experiments could shed light on multimodal aspects of communication in birds.

Acknowledgments

We thank Philippe Groué, Emmanuelle Martin, Ophélie Bouillet, Marie Huet, Priscilla Roussel and Anaëlle Brunet for taking care of the birds. We thank Félix Bigand, Perrine Marsac, Philippine Prevost, Marie Soret and Lisa Jacquey for their help regarding the development of the vocal loop.

References

Aamodt, C. M., Farias-Virgens, M., & White, S. A. (2020). Birdsong as a window into language origins and evolutionary neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190060. https://doi.org/10.1098/rstb.2019.0060

Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A., & Tchernichovski, O. (2016). The Forebrain Song System Mediates Predictive Call Timing in Female and Male Zebra Finches. *Current biology : CB*, *26*(3), 309-318. https://doi.org/10.1016/j.cub.2015.12.037

Butler, S., & Fernández-Juricic, E. (2014). European starlings recognize the location of robotic conspecific attention. *Biology letters*, *10*. https://doi.org/10.1098/rsbl.2014.0665

Carouso-Peck, S., & Goldstein, M. H. (2019). Female Social Feedback Reveals Non-imitative Mechanisms of Vocal Learning in Zebra Finches. *Current Biology: CB*, *29*(4), 631-636.e3. https://doi.org/10.1016/j.cub.2018.12.026

Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the Zebra Finch (*Taeniopygia guttata*). *Interaction Studies*, *12*(2), 324–350. https://doi.org/10.1075/is.12.2.07der

Derégnaucourt, S., & Gahr, M. (2013). Horizontal transmission of the father's song in the zebra finch (Taeniopygia guttata). Biology Letters, 9(4), 20130247. https://doi.org/10.1098/rsbl.2013.0247

Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, *433*(7027), 710-716. https://doi.org/10.1038/nature03275

Derégnaucourt, S., Poirier, C., Kant, A. V. der, Linden, A. V. der, & Gahr, M. (2013). Comparisons of different methods to train a young zebra finch (Taeniopygia guttata) to learn a song. *Journal of Physiology, Paris*, *107*(3), 210-218. https://doi.org/10.1016/j.jphysparis.2012.08.003

Derégnaucourt, S., Saar, S., & Gahr, M. (2012). Melatonin affects the temporal pattern of vocal signatures in birds. *Journal of Pineal Research*, *53*(3), 245-258. https://doi.org/10.1111/j.1600-079X.2012.00993.x

Deshpande, M., Pirlepesov, F., & Lints, T. (2014). Rapid encoding of an internal model for imitative learning. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1781), 20132630. https://doi.org/10.1098/rspb.2013.2630

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology*, *5*(12), e321. https://doi.org/10.1371/journal.pbio.0050321

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. *Bird Vocalizations*, *61*.

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1-11. https://doi.org/10.1006/anbe.2000.1410

Jolly, L., Pittet, F., Caudal, J.-P., Mouret, J.-B., Houdelier, C., Lumineau, S., & Margerie, E. (2016). Animal-to-robot social attachment : Initial requisites in a gallinaceous bird. *Bioinspiration & biomimetics*, *11*, 016007. https://doi.org/10.1088/1748-3190/11/1/016007

Lerch, A., Roy, P., Pachet, F., & Nagle, L. (2011). Closed-loop bird-computer interactions : A new method to study the role of bird calls. *Animal cognition, 14,* 203-211. https://doi.org/10.1007/s10071-010-0353-6

Ljubičić, I., Hyland Bruno, J., & Tchernichovski, O. (2016). Social influences on song learning. *Current Opinion in Behavioral Sciences, 7,* 101-107. https://doi.org/10.1016/j.cobeha.2015.12.006

Patricelli, G. L., Uy, J. A. C., Walsh, G., & Borgia, G. (2002). Male displays adjusted to female's response. *Nature, 415*(6869), 279-280. https://doi.org/10.1038/415279a

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *Journal of Comparative and Physiological Psychology, 93*(2), 260-277. https://doi.org/10.1037/h0077553

Rassak, S., Nachamai, M., & Krishna, M. A. (2016). Survey study on the methods of bird vocalization classification. *2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC),* 1-8. https://doi.org/10.1109/ICCTAC.2016.7567337

Sossinka, R., & Böhner, J. (1980). Song Types in the Zebra Finch *Poephila guttata castanotis*. *Zeitschrift für Tierpsychologie, 53,* 123-132. https://doi.org/10.1111/j.1439-0310.1980.tb01044.x

Ter Maat, A., Trost, L., Sagunsky, H., Seltmann, S., & Gahr, M. (2014). Zebra finch mates use their forebrain song system in unlearned call communication. *PloS One, 9*(10), e109334. https://doi.org/10.1371/journal.pone.010933

CHAPITRE IV

Of birds and robots:

Influence of behavioural contingency on developmental song learning in the Zebra Finch (*Taeniopygia guttata*)



© Louisane Araguas

Of birds and robots: influence of behavioural contingency on developmental song learning in the Zebra Finch (*Taeniopygia guttata*)

Alice Araguas [1*] Bahia Guellaï[1,2], Adrien Chopin[3] and Sébastien Derégnaucourt [1, 2]

[1] *Laboratoire Ethologie Cognition Développement, Université Paris Nanterre, 200 Avenue de la République, F92001 Nanterre Cedex, France*

[2] *Institut Universitaire de France*

[3] *INSERM, CNRS, Institut de la Vision, Sorbonne Université, F75012 Paris, France*

* Correspondence: Alice Araguas, Laboratoire Ethologie Cognition Développement, UPL, Univ Paris Nanterre, F-92000 Nanterre, France. E-mail address: alice.araguas@parisnanterre.fr

Abstract

Social robots represent innovative and promising tools in the field of animal behaviour studies: they allow a high standardization of experimental procedure and they enable to control independently factors that can influence behaviours. However, studies on learning from robots are still scarce. Here we show that a robot can be used as a song tutor for young zebra finches (*Taeniopygia guttata*). In this species, song learning occurs early in the development and requires physical, visual and vocal interactions between pupils and tutors. We compared song learning in young zebra finches whether they were daily exposed to an unfamiliar adult male (live group) or an interactive zebra finch robot (robot group), during the sensitive period for song learning. We found that pupils produced good copies of tutor songs in both groups. In both groups, we also observed that song imitation was correlated to the number of songs heard by pupils in an inverted U way. In the robot group, song imitation linearly increased with the time spent on the robot's platform and was correlated with the number of pecking on the robot's beak and head in an inverted U way. Finally, our results suggest that young birds remained interested in the robot during the whole experiment. We provide the first evidence that a robot can successfully be used as a tutor to drive vocal learning in young animals.

Keywords: bird-robot interaction, song learning, sensitive phase of learning, behavioural study, *Taeniopygia guttata*

1. Introduction

Vocal learning is the ability to memorise and reproduce vocalisations. Humans share this ability with other animal species from few taxonomic groups including oscine songbirds It had been demonstrated that social contingency influences language acquisition in children (Roseberry et al., 2014) but also vocal learning in other species such as the Marmoset Monkey (*Callithrix jacchus*) (Takahashi et al., 2017) and bird species, as cowbirds (*Molothrus ater ater*) (West & King, 1988; King & West, 1989). In the Zebra Finch (*Taeniopygia guttata*), Carouso-Peck & Goldstein (2019) demonstrated that a visual and non-vocal feedback given by female finches and known as a response to attractive songs, influenced song learning. When a contingent feedback - the broadcast of a video of a female exhibiting this specific behaviour - was experimentally given to young males while they were singing, they further produced better copies of the song model heard early in life in comparison to their brother exposed to the same video non-contingently (Carouso-Peck & Goldstein, 2019).

Under natural conditions, male zebra finches learn to sing by imitating the songs of their conspecifics, mainly adult males (see Derégnaucourt, 2011 for a review; Derégnaucourt & Gahr, 2013). They learn their short and stereotyped song (1-1.5sec) during a sensitive period (25-90 days after hatching). As this small bird is well adapted to captivity and the development of his song is rapid, the zebra finch song has been established as a model of choice in neuroethology and allows to study neurobiological, hormonal and genetic basis of vocal learning (Hauber et al., 2021).

Many laboratory studies evaluated the influence of social factors on zebra finches' song development. For example, birds raised in social isolation during the sensitive phase of learning produce songs which present some of the characteristics of the species but are abnormal (Price, 1979). Physical and visual contacts are crucial during learning as young birds separated from their song tutor by an opaque partition or a grid, produce a worst copy than birds that can

physically interact with it (Derégnaucourt, 2011). Finally, in this species, young males produce the highest imitation of songs when they are raised with an adult male (Tchernichovski & Nottebohm, 1998; Derégnaucourt et al., 2013). This one-to-one situation constitute a reference for birdsong research (Derégnaucourt, 2011). However, this method does not allow precise control of the different variables involved in song learning (e.g. control of the behaviour of the tutor, number of songs heard, etc ...) which can be done by using a robotic tutor.

These last years, the number of social robots used in animal behaviour research increased with promising results in several species (Romano et al., 2019 for a review) including birds (Jolly et al, 2006; Patricelli et al., 2002; Patricelli & Krakauer, 2009). To our knowledge, no study has been yet conducted to check the influence of social contingency on learning in birds using a robotic device.

In the present study, we took advantage of the recent development of a robotic zebra finch (MANDABOT, Araguas et al., submitted) that sings, moves and reacts to the behaviour of a live zebra finch. We investigated whether this zebra finch robot could be used as a song tutor for a young zebra finch. In particular, we will measure the importance of behavioural contingency on song imitation.

To do so, we will compare song imitation in young finches exposed either to a live bird or to a robot finch. To check the role of behavioural contingency on song imitation, we will compare the learning outcome of birds raised either with a robot exhibiting contingent behaviour or a robot exhibiting non-contingent behaviour.

2. Materials and methods

2.1. Subjects

We used 36 male zebra finches (n= 12 adults, age ± SE: 6.42 years ± 0.15 [4.58-7.08], n= 24 juveniles) that hatched in the animal facilities of the University Paris Nanterre (France). After hatching, young zebra finches spent 14 days with their parents and other siblings in

housing cages (LD 14 : 10, T°=20-23°C). At 14 dph (days post hatch), mothers and offspring were transferred into an isolation room to avoid any exposure to adult male song. At 35 dph, young males were isolated in individual cages (46x23x27 cm) in sound-proof chambers (85x65x60 cm). Sound proof chambers were equipped with fans providing a low airflow and OSRAM DULUX lights on an automatic 14:10 light:dark schedule. Each cage contained three perches and we placed a round mirror (diameter: 10 cm) above one of them to reduce the impact of social isolation. Birds had access to water, food, sand and cuttle bones *ad libitum*. Once a week this diet was supplemented with vegetables. All procedures followed the European regulations on animal experimentation and were approved by the Darwin Ethic Comity of the French Ministry for National Education, Higher Education and Research (authorization #20641201905141523153 4).

2.1.1. Experimental procedure

After 2 days of habituation to the sound proof chamber, each young finch was transferred to a cage containing a song tutor during 1h, 5 days a week during 5 consecutive weeks. After exposure, the young finch was transferred to his housing cage. At day 1 of exposure, a grid placed in the middle of the cage separated the tutor from the pupil. Based on the results of previous experiments (Chen et al., 2016), we used this procedure to prevent eventual agonistic behaviours. At day 2, the grid was left only for the first 30 minutes of the session. Then we removed it to permit physical interactions between the young finch and the song tutor for the following 30 minutes. From day 3 to the end of the experiment, the tutor and the pupil could physically interact for the whole hour. At 100 dph when song is usually crystallized, the experiment ended. All birds were then transferred to aviaries with conspecifics.

2.1.2. Live group: exposure to a live tutor

In the live group, each pupil (called L1 to L12) was paired with a different adult tutor (called T1 to T12) that remained the same during the whole experiment. Tutors came from a previous experiment in which all founder males of a colony were trained to sing from the same song model (Le Maguer et al., submitted). Twice a week and for one hour, we placed a female in the tutor's cage to limit social isolation and to record song production in the presence of a female.

2.2. Exposure to a robotic finch, the MANDABOT

For a detailed description of the MANDABOT, see Araguas et al., 2021 (submitted). Briefly, the robot was fixed on a grey plastic box (16x8.5x8 cm) and had the size, the shape and the plumage colours of an adult male zebra finch. It was controlled via a custom program running on a Beaglebone Black Wireless and was connected to a Dell Latitude 5300 computer running on Microsoft Windows 10 (Figure 1). It was equipped with a speaker (AS01808MR from PUI Audio) to broadcast pre-recorded zebra finch songs and calls collected in the live group (see below). On each side of the robot, there was a perch (5 cm) connected to load cells (CZL639HD from Phidgets). We also developed a closed-loop program to mimic the vocal interactions between two finches. Such bird-machine vocal interactions have already been successfully applied in oscine songbirds including zebra finches (Lerch et al., 2011; Benichov et al., 2016). Sounds from the experimental cage were continuously recorded using a Behringer C-2 microphone connected to a PreSonus AudioBox 1818VSL (16 bits, 44.1 kHz). When a vocal sound produced by the pupil (either a call or a song syllable) was detected by the system, one pre-recorded call among the 50 calls from the database of the specific tutor was broadcast through the speaker. The robot was placed in a cage of a similar size as the one used for housing groups, and contains two perches, with food and water.

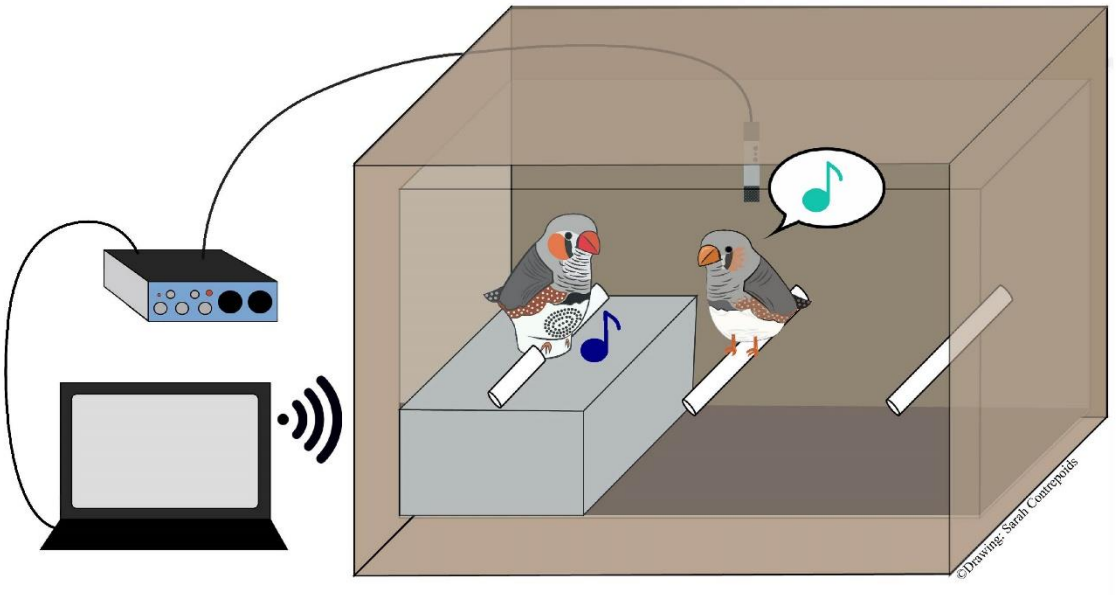


Figure 1. Experimental setup with the zebra finch robot and a pupil.

2.2.1. Contingent group

In the contingent group, the robotic zebra finch produced sounds (songs and calls), moved and reacted to the behaviour of the pupil. Each of the 12 pupils of the contingent group (called C1 to C12) was daily exposed to the robot producing songs and calls of one of the 12 tutors from the live group. When the pupil landed on the perch, the robot turned in his direction and randomly moved its head or broadcast a song among the 20 of his specific tutor's repertoire. When the pupil left the perch, the body of the robot turned back to the centre in a neutral position.

Following the same experimental schedule as in the live group, we launched a closed-loop program for 60 minutes for each pupil. At the end of the loop, the computer generated a script containing all the sequences of movements and tutor's specific vocalisations produced by the robot. For each pupil, 25 scripts were produced corresponding to the 25 daily exposure sessions to the contingent robot.

For the first day of exposure, a 60 min closed-loop was launched. The grid prevented physical interactions but the robot and the pupil interacted vocally via the vocal loop. To ensure that the pupil was exposed to a minimum quantity of robot's songs even if he did not interact with the robot later during the experiment, the robot was programmed to broadcast 10 songs. Due to technical problems, some birds were exposed to less songs: 4 birds were exposed to 9 songs, one bird to 6 songs and another one to 5 songs during the first day of exposure.

As explained previously, during the second day of exposure, the grid was removed after 30 min which allowed the pupil to interact with the robot for the last 30 min. The robot was programmed to broadcast one song during the first 30 min to ensure that the pupil heard the robot singing even if he did not land on perches later.

The number of songs heard by the pupil depended, except for day 1 (complete session) and day 2 (beginning of the session), on pupils' behaviours and on the number of times they landed on perches on either side of the robot.

2.2.2. Non-contingent group

Due to the Covid19, we could not run this group so far. Following the same procedure, 12 young male finches will be exposed to the same robot programmed with a non-contingent mode. Each pupil will be exposed to the robot running the 25 daily scripts recorded in the closed-loop interaction context between the robot and the pupil in the contingent group.

Therefore, each pupil will be exposed to the same sequence of movements and sounds produced by the robot in the contingent group, but there will be no contingency between its own behaviour and those of the robot.

2.3. Behavioural analysis

2.3.1.Sound recording and analysis

During the whole experiment, sounds were recorded continuously using SAP 2011 (Sound Analysis Pro software; Tchernichovski et al., 2000). Each cage was equipped with a Behringer C-2 microphone connected to a PreSonus AudioBox 1818VSL (24 bits, 96 kHz) controlled by a DELL Optilex GX620 PC on Windows 7. During recording, SAP detects and saves individual songs into separate files, whilst mostly discarding isolated calls and cage noises.

2.3.1.1. Song segmentation

The Zebra Finch song is hierarchically organized: songs are produced in bouts, each bout usually starting with short sounds called introductory notes followed by one or several motifs: a motif being a stereotyped sequence of song syllables separated by silent intervals. Each syllable is composed of one or several notes or elements; each element representing a different vocal gesture. Successive motifs in a bout can sometimes be separated by call-like syllables called connectors (Hyland Bruno & Tchernichovski, 2017).

We selected sound files containing songs produced at 100 dph for the pupils and during the last days of the live tutoring experiment for the tutors. Using the Goldwave software (v6.36), we applied a high pass filter at 420 Hz (Lachlan et al., 2016) and a volume maximization of 90% to sound files. Then, songs were segmented into syllables using SAP 2011 (see below) until we selected at least 300 syllables for each bird. For each syllable, SAP 2011 extracted 14 different parameters: duration, amplitude, pitch, frequency modulation (FM), squared amplitude modulation (AM^2), Wiener entropy, goodness of pitch, mean frequency and the variance in pitch, FM, entropy, pitch goodness, mean frequency and AM (see SAP 2011 manual: http://soundanalysispro.com/manual-1).

For each bird, the table containing the results of this analysis was loaded into the SongSeq software (Daou et al., 2012) to classify the different syllables into different clusters (syllable types). Once syllables were clustered into syllable types, data were then further processed using a custom-made program written in MATLAB (R2017a) in order to create one *.wav file per syllable and to classify each sound file created into different folders; each folder containing all sounds belonging to the same cluster. We checked that the clustering was appropriately done by visually inspecting spectrograms and eventually listening to sounds, using Sound Explorer (René Jansen, University of Amsterdam). If needed, syllables were re-assigned to the appropriate cluster.

In order to analyse sequence of syllables within a bout, we first had to set a value for the maximal gap between successive bouts. Based on previous analyses, we set this value at 200ms. Using a custom-made program written in MATLAB, we obtained for each song bout: 1) the total number of syllables; 2) the bout duration and 3) song linearity (Scharff & Nottebohm, 1991). The linearity score represents how often syllables are sung in a specific order. In our study we calculated a slightly modified linearity score, as follows: linearity = (number of different syllables – 1) / number of syllable-to-syllable transitions (Iyengar & Bottjer, 2002; Kao & Brainard, 2006; Zevin et al., 2004).

2.3.1.2. Similarity to the song model: analysis based on the motif

Song similarity was quantified using an automated procedure implemented in SAP that parametrically quantifies the similarity between songs. The similarity procedure implemented in SAP provides 3 components: % similarity, mean accuracy and sequential match. Percentage of similarity is the percentage of similar sounds included in final sections. Mean accuracy is the average local similarity scores across final sections. Sequential match is calculated by sorting

the final sections according to their temporal order in reference to the song model, and then examining their corresponding order in the song motif.

First, portions of songs were selected after a visual inspection of the spectrograms, by cutting the sound at the onset and offset of the selected sequence, using Sound Explorer (René Jansen, Amsterdam). Second, regions of high similarity between the segments of the pupil and the model songs were identified, and the results were aggregated into a global measure of acoustic similarity and sequence similarity. In asymmetric comparisons, the most similar sound elements of two sequences (tutor song and pupil song) are compared, independent of their position within the sequence. The smallest unit of comparison is 9.26-ms–long sound interval (FFT windows). Each interval is characterized by measures for five acoustic features: pitch, FM, amplitude modulation (AM), Wiener entropy, and pitch goodness. SAP calculates the Euclidean distance between all interval pairs from two songs, over the course of the motif, and determines a p-value for each interval pair. This p-value is based on p-value estimates derived from the cumulative distribution of Euclidean distances across 250.000 sound interval pairs, obtained from 25 random pairs of zebra finch songs. Neighbouring intervals that pass the p-threshold value ($p = 0.05$ in this study, default value of SAP) form larger similarity segments (70 ms). In asymmetric similarity measurements, we want to judge how good the copy is in reference to the song model. The song model is loaded as “sound 1” and the copies are loaded as “sound 2” in the batch module of the SAP software. Therefore, in our study, song models (tutor songs) were loaded as “sound 1” and pupil songs were loaded as “sound 2”. In summary, the amount of sound from the tutor song that is included into the similarity segments represents the similarity score; it thus reflects how much of the tutor’s song material is found in the pupil song. This procedure was repeated 100 times, comparing 10 different exemplars of the tutor song with 10 different exemplars of each pupil’s song. The mean value of these 100 comparisons was used for statistical analysis.

2.3.1.3. Preparation of acoustic stimuli for the robot

The live group (control group) allowed us to collect acoustic stimuli that were used in the two robot groups. For each of the 12 adult tutors, we selected 20 songs produced in the presence of the pupil and 50 contact calls emitted when he was alone in his cage. Using Avisoft SASLab Pro, we applied a high pass filter at 420 Hz and a volume maximization of 90% to all the vocalisations. It constitutes the vocal identity of each robotic tutor.

2.3.2.Video recording and analysis

All sessions were video recorded using a Logitech C920 webcam connected to ContaCam 4.9.9 software on HP ProBook 650 G1 computer running on Windows 7.

2.3.2.1. Video analysis for the live group

For each pupil, 23 out of 25 videos were analysed using the software BORIS v.7.9.16. Day 1 was excluded as the grid prevent physical contacts. As finches constantly moved in the cage, we decided to use the scan sampling method. Each minute, we assigned a proximity score between the pupil and his tutor regardless of the exhibited behaviour. We used custom-coded keystrokes in BORIS to score 3 different distances (Figure 2). We obtained 60 proximity scores for each video and we calculated a mean proximity score for each pupil for each day and for the whole experiment.

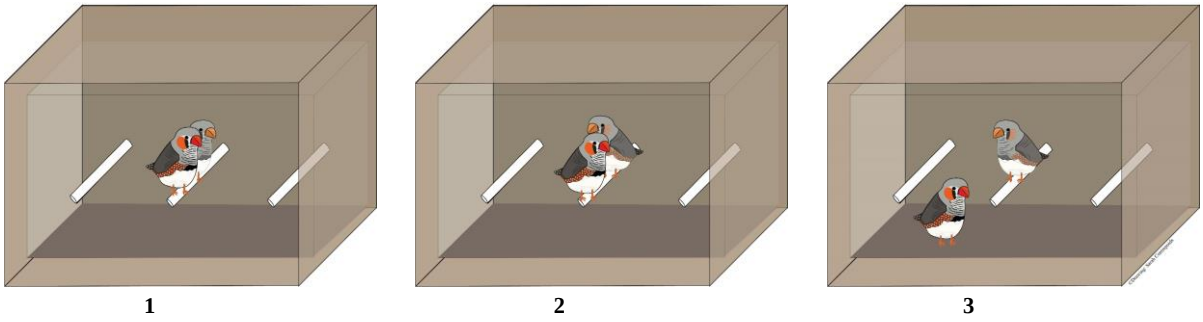


Figure 2. Proximity scores assessed between the adult tutor and the pupil. (1) Score of 1: from bodily contact between finches to the distance of a size of a zebra finch, in the ground or on the same perch. (2) Score of 2: from the size of a zebra finch to 10 cm, on the ground or on the same perch. (3) Score of 3: all distances upper 10 cm. When the two finches were on different perches, the score of 3 was automatically given.

2.3.2.2. Video analysis for the contingent group

For the contingent group, we conducted a more precise behavioural analysis. For time reasons, on the 25 videos recorded for each pupil, 6 videos were analysed using the software BORIS v.7.9.16. Days 1 and 2 of exposure were not analysed as a grid separated the pupil and the robot, preventing physical interactions between the bird and the robot respectively for the whole session and for half of the session. As described on Table 1, we coded videos of the day 3, the first time the bird interacted for 60 min with the robot and then, first days of each following week (days 6, 11, 16, 21) and the last day of the experiment (day 25).

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	day 1	day 2	day 3	day 4	day 5
Week 2	day 6	day 7	day 8	day 9	day 10
Week 3	day 11	day 12	day 13	day 14	day 15
Week 4	day 16	day 17	day 18	day 19	day 20
Week 5	day 21	day 22	day 23	day 24	day 25

Table 1. Experimental schedule. Each pupil was exposed to the robot one hour per day, 5 days a week during 5 weeks. Video analysis was performed for experimental days indicated in blue.

We created an ethogram to describe interactions between the pupil and the robot. We focused our analysis on behaviours oriented towards the robot. We used custom-coded keystrokes in BORIS to quantitatively score these behaviours from the videos (Table 2). On the fifteen behaviours coded, five (platform, close vicinity, clumping, preening, sleeping) were analysed for their duration because they occurred continuously for long periods of time (on the order of seconds). Pecking on the different body parts of the robot and on the substrate was coded in occurrences as their duration was less than one second.

Following Chen and al. (2016), we analysed pupils' reactions to the song tutor. We scored five possibilities of pupils' reactions when they heard the robot' song: stays still, stands up, moves on the platform, flies away and orientates his body towards the robot, flies away and does not orientate his body towards the robot.

Code	Behaviour	Description
q	Platform	Moving, lying down or standing on the platform for less than 10 s
a	Close vicinity	Lying down or standing on the platform for more than 10 s
u	Clumping	Bodily contact between the pupil and the robot, flank to flank (Sparks, 1964) for more than 10 s
t	Preening	The pupil turns his head to groom his feathers, his beak is hidden then put back his head in a neutral position
d	Sleeping	Pupil's eyes are closed and his head is under his wing
Pecking on the:		
b	Beak	Robot's beak
v	Head	Robot's head
c	Body	Robot's neck, flanks and legs
x	Tail	Robot's tail
m	Substrate	Platform, perches and cage bars
Pupil's reactions to the robot' song:		
s	Stays still	Pupil stays still
f	Stands up	Pupil stands up
g	Moves	Pupil moves on the platform
h	Flies away and orientated	Pupil leaves the platform and orients his body and his head in the direction of the robot
j	Flies away	Pupil leaves the platform and does not orient himself in in the direction of the robot

Table 2. Ethogram used for quantitative behavioural analysis. For each behaviour, we detailed the keyboard keys used on BORIS software, the category of behaviour and its description.

2.4. Statistical analysis

Statistical analyses were conducted in R Studio v. 1.4.1103 and Matlab R2017A. We checked distributions of data using Shapiro Wilk tests. Significance level was $p < 0.05$. We used non parametric tests or Generalized Linear Models (GLM) with non-normal distributions.

2.4.1.Acoustic parameters

Song similarities, song linearities, bout durations in both groups and number of syllables in the contingent group were normally distributed (Shapiro-Wilk tests: $p > 0.05$) whereas number of syllables in the live group followed a Gamma distribution (Shapiro-Wilk test: $W = 0.82$, $p = 0.02$). We used Welch Two Sample t-tests to compare song similarities song linearities and bout durations between the two groups and a Mann-Whitney test for the number of syllables.

2.4.2.Influences of song activity on song imitation

To predict song similarity, we used a Generalized Linear Mixed Effects Models (GLME) with the tutor's source of the songs as random effects, two components of the number of songs heard – a linear and a quadratic one – and an interaction between the linear component and the number of songs heard, as fixed effects (reciprocal link). As a reminder, songs from each live tutor were copied to each robot tutor. Therefore, using the tutor's source as random effects, it is possible to account for the effect of song complexity when predicting the pupil's learning. In addition, we worked with song dissimilarity (100 – song similarity, in %) as the dependent variable instead of song similarity, because song dissimilarity followed a more typical gamma distribution.

2.4.3.Evolution of behaviours over the experiment

2.4.3.1.Live group

We calculated a mean proximity score for each dyad, for each day and for the whole experiment. Proximity scores were not normally distributed (Shapiro-Wilk test: $W = 0.95$, $p = 0.02$). Inverted the scores (4 – scores) allowed to obtain a clear gamma distribution. Therefore, we modelled the relationship of these new scores with the predictors assuming a gamma distribution and a reciprocal link. To assess the evolution of the proximity scores across weeks for the live group, we conducted a GLME with the week of the experiment as fixed effects (both a quadratic and a linear component) and the pupils as random effects.

2.4.3.2. Contingent group

As durations on the platform encompassed durations in close vicinity that also included those in clumping, we preferred to use durations in an exclusive way. Exclusive platform durations were obtained by subtracting close vicinity durations from platform durations; exclusive close durations were obtained by subtracting clumping durations from close durations. All durations followed gamma distributions.

Evolution of durations over the experiment:

To assess the evolution of the durations across days, we conducted a GLM with the day of experiment as a predictor. We assumed that durations followed a gamma distribution (adding 1 to all durations to avoid zero) and we model the relationship with the predictors using a reciprocal link. Including the tutor source in a GLME did not increase the AIC (Akaike's Information Criteria).

Comparisons of durations for day 5 versus day 6 and day 10 versus day 11:

For days 5, 6, 10 and 11, durations spent on the platform and on close vicinity were normally distributed (Shapiro-Wilk tests: $p > 0.05$) whereas clumping durations were not (Shapiro-Wilk tests: $p < 0.05$). To compare day 5 to day 6 and day 10 to day 11, we used paired t-tests for platform and close vicinity durations and Wilcoxon tests for paired data for clumping durations.

Evolution of the number of pecking on the robot:

We compiled pecking on the robot's beak and head and pecking on the robot's body and tail. We estimated the mean number of pecking on the robot's beak and head and on the robot's body and tail each from separate GLME with the pupils as random effects and the day (through a linear and a quadratic component) as fixed effects. We assumed a Poisson distribution (given pecking are events) and a log link. We used a GLME to test whether the decreasing of the number of pecking is significant, with the tutor source as random effects.

Reactions to the robot songs:

On the 5 reactions described, we gathered "stays still" and "stands up" in a category called "attentive reactions" as proposed by Chen and al. (2016). Given we had proportions, we used normal distributions. We used a generalized linear mixed-effects model to test whether the number of attentive reactions decreased over the experiment. To characterize the evolution of the proportion of attentive reactions across days, we conducted a GLME with pupil as random effects and day (both with a linear and a quadratic component) as fixed effects. The proportion of attentive reactions did not follow a normal (Shapiro-Wilk test: $W = 0.84$, $p < 0.001$) or a clear gamma distribution. Therefore, we transformed it to the proportion of

inattentive reactions (1-proportion of attentive reactions) plus one (to avoid zero values), and used a gamma distribution with a reciprocal link to model its evolution across days.

Evolution of the number of calls over the experiment:

To characterize the evolution of the number of calls across days, we conducted a GLME with pupil as random effects and day (both with a linear and a quadratic component) as fixed effects. The number of calls did not follow a normal (Shapiro-Wilk test: $W = 0.65$, $p < 0.001$) but a clear gamma distribution. Therefore, we added one (to avoid zero values), and used a gamma distribution with a reciprocal link to model its evolution across days.

2.4.4.Effect of behaviours on song similarity

2.4.4.1. Live group

We used a GLM to characterize the relationship between proximity scores and song similarity with mean proximity scores as a linear and quadratic predictor (normal distribution and identity link).

2.4.4.2. Contingent group

We applied a GLM to study the link between song similarity and the time spent on the platform (not clumping, not close) as a linear and quadratic predictor (normal distribution and identity link). We conducted a similar analysis with the number of pecking on the beak or head of the robot, as a predictor, and independently, with the number of pecking on the body or tail, the proportion of attentive reactions and the number of calls produced by the robot during the experiment.

3. Results

3.1. Acoustic parameters

3.1.1.Song similarity between tutors and pupils

In both groups, song imitation was high as pupils copied on average 75% of the tutors' songs (Figure 4A) live group: mean ± SE: 75.30% ± 1.51, contingent group: mean ± SE: 76.63% ± 2.25). Nevertheless, we observed an inter-individual variability in song imitation in both groups: in the contingent group, 2 birds (C5 and C10) copied less than 50% of their tutors' songs (T5 and T10). In the live group, no pupil was under 50% of similarity but we can notice that the two poorest imitations were produced by the two pupils (L5 and L10) who learned their songs with the same tutors: tutors T5 and T10 (Figure 3). The tutor T3 was very well imitated in both groups (92.76% in the live group, 90.11% in the contingent group; Figure 3).

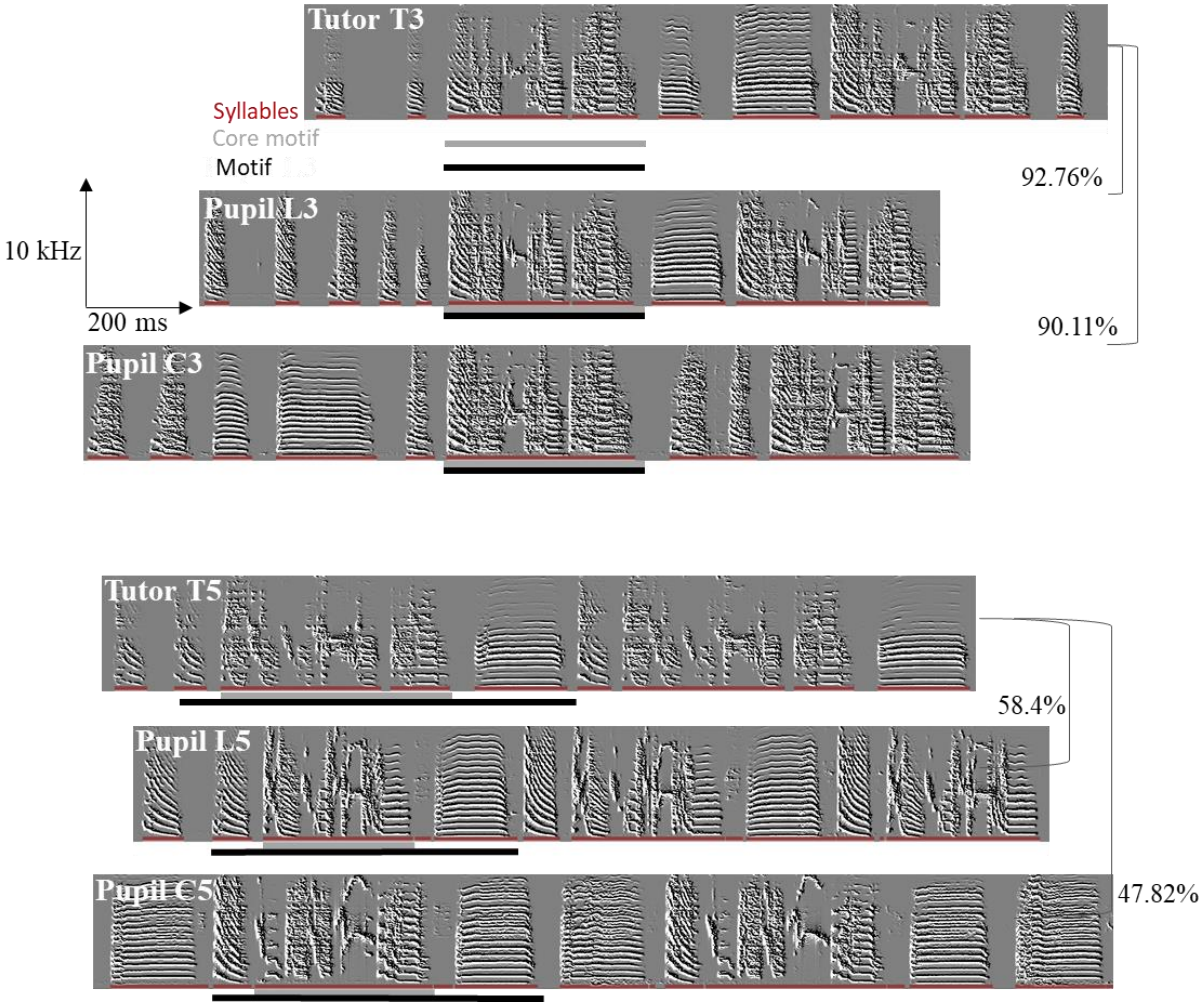


Figure 3. Examples of spectrograms of songs produced by tutors and pupils in the two groups. Frequency is in kHz and duration is in ms.

3.1.2.Comparison of acoustic parameters in the two groups

Four acoustic parameters were compared in both groups: song similarity, song linearity, bout duration and the number of syllables in songs. As shown in Figure 4, we did not find any significant difference between the two groups regarding song similarity (Welch Two Sample t-test: t = -0.24, df = 19.25, p = 0.81), song linearity (Welch Two Sample t-test: t = -0.65, df = 21.99, p = 0.53), bout duration (Welch Two Sample t-test: t = -0.26, df = 21.86, p = 0.80) and the number of syllables in songs (Mann-Whitney test: W = 77, p = 0.80).

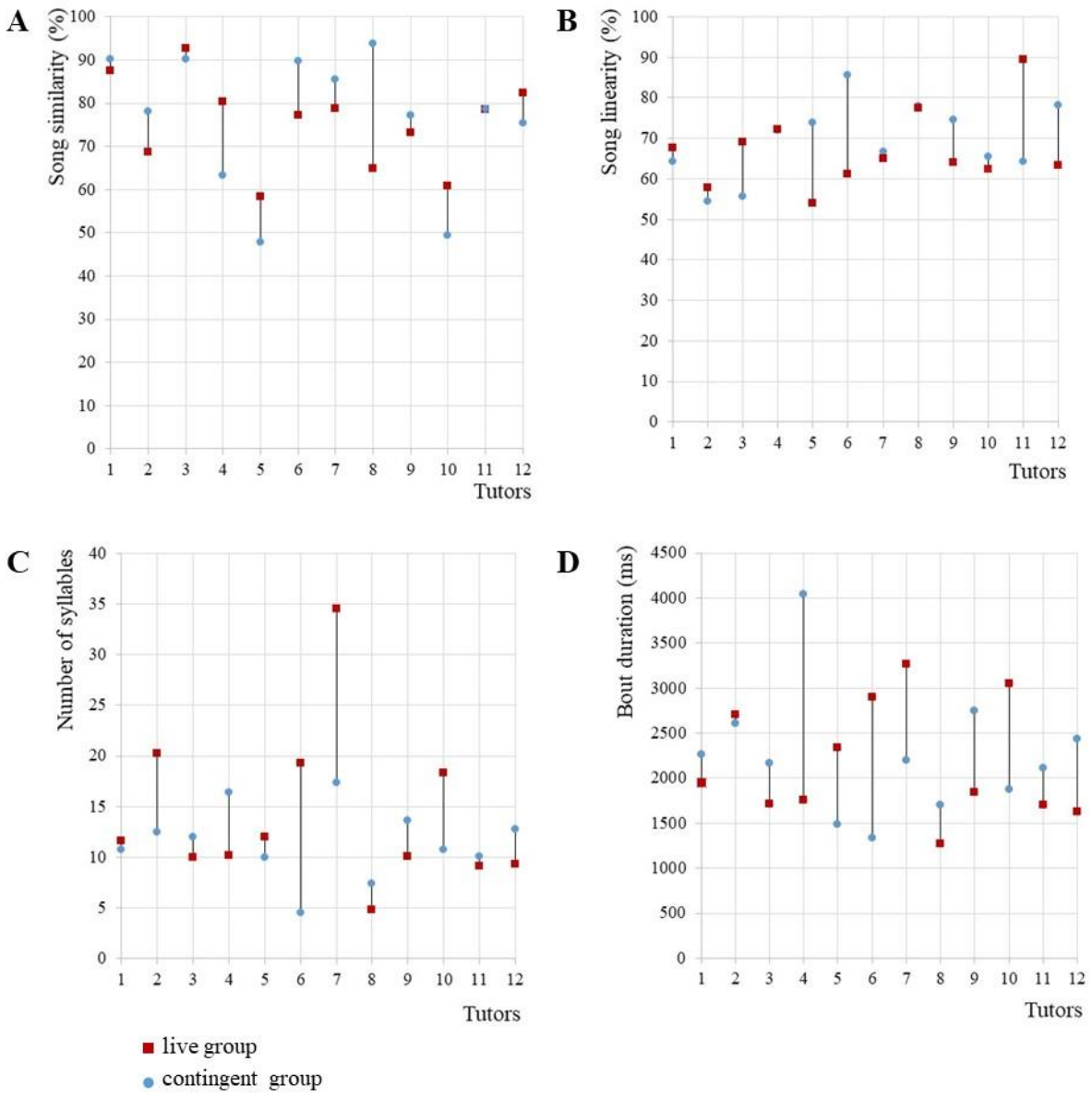


Figure 4. Comparison of songs parameters in live and in contingent groups. (A) Song similarity (mean in %), (B) song linearity (mean in %), (C) number of syllables (mean) and (D) bout duration (mean in ms). In the live group, each pupil (red square) was trained with a unique adult tutor. Songs of the tutor were broadcast by the robot to the paired pupil in the contingent group (blue circle).

3.2. Influences of song activity on song learning

In both groups, we observed an inter-individual variability in exposure to tutor songs (live group: mean per day ± SE = 38.07 songs ± 10.25, contingent group: mean per day ± SE = 10.06 songs ± 1.96). During the whole experiment, pupils L10 and L2 heard respectively 2 672 and 2 543 songs whereas pupils L8 and L9 heard respectively 334 and 218 songs. Pupils C1 and C2 heard respectively 565 and 547 songs whereas pupil C5 heard 26 songs and C10 heard 7 songs. As previously explained, we used song dissimilarity which depended on the number of songs heard, with a linear component ($t(20) = 3.57$, $p = 0.002$) and a quadratic component ($t(20) = -3.54$, $p = 0.002$), and the linear component interacts with the group ($t(20) = 2.31$, $p = 0.03$, Figure 5). More specifically, the song similarity varied with the number of songs heard following an inverted U, and the top of the inverted U was reached faster in the contingent group than in the live group. As a reminder, no difference was found in song similarities between the two groups.

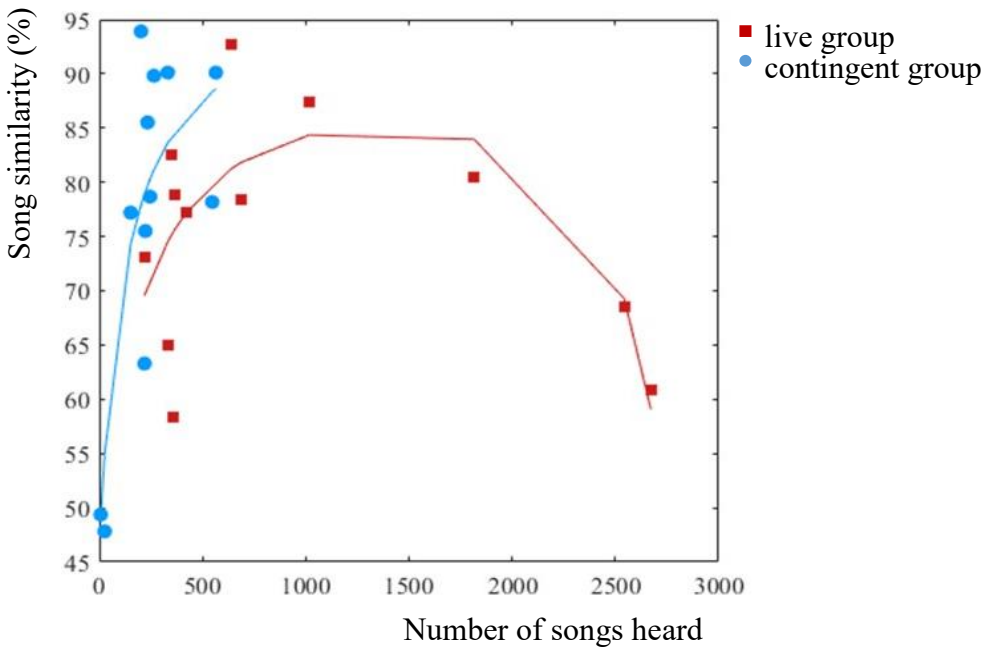


Figure 5. Song similarity scores (mean in %) as a function of the total number of songs produced by the tutors (live in red, robotic in blue) over the experiment.

3.3.Evolution of behaviours over the experiment

3.3.1. Live group

Over the experiment, the proximity scores linearly decreased with the experimental week ($t(57) = -2.2$, $p = 0.03$) with no quadratic component ($t(57)=1.85$, $p = 0.07$, Figure 6).

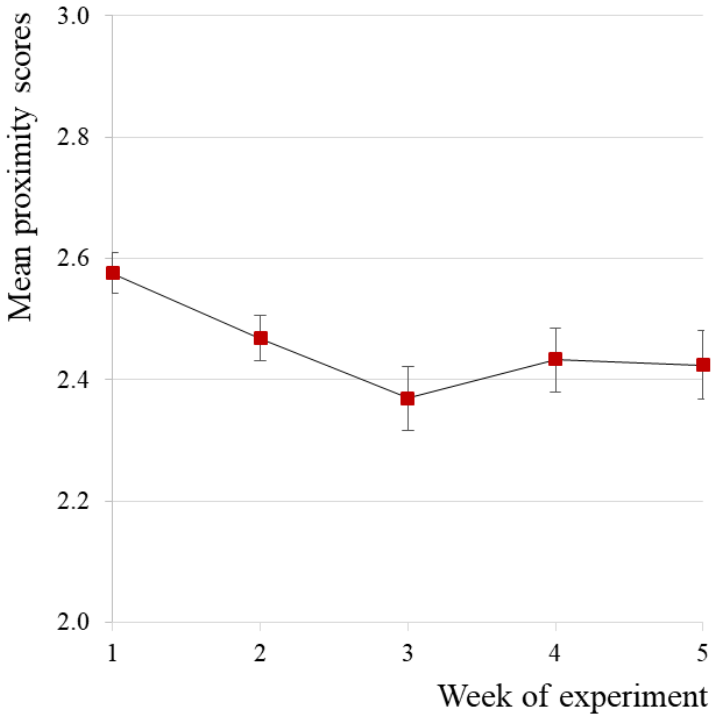


Figure 6. Evolution of the physical distance (mean proximity scores ± SE) between tutors and pupils over the experiment.

3.3.2. Contingent group

3.3.2.1. Times spent on the platform, in close vicinity and in clumping

During the first day of physical interactions with the robot (day 2):

During the first day of physical interactions, pupils spent 70.65% ± 4.92 (mean ± SE) of the 30 min-session on the platform and 64.57% ± 5.39 (mean ± SE) on close vicinity. Regarding physical interactions, 4 birds out of 12 exhibited clumping behaviours with the robot and one of them, C4, spent 73.96% of the session clumped to the robot.

Moreover, 9 birds out of 12 pecked the robot and 57.19% ± 2.38 (mean ± SE) of the total number of pecking was on the robot's body. These results suggest that most finches exhibited a strong interest for the robot from the first day when they could physically interact with it.

Comparison of days 3, 6, 11, 16, 21 and 25:

Over the experiment, the time spent in the platform increased with the experimental day (t(70) = -2.00), p = 0.049). But the time spent exclusively in close vicinity and the time spent clumping each decreased (t(70) = 3.08, p = 0.003 and t(70) = -2.14, p = 0.03 respectively) (Figure 7).

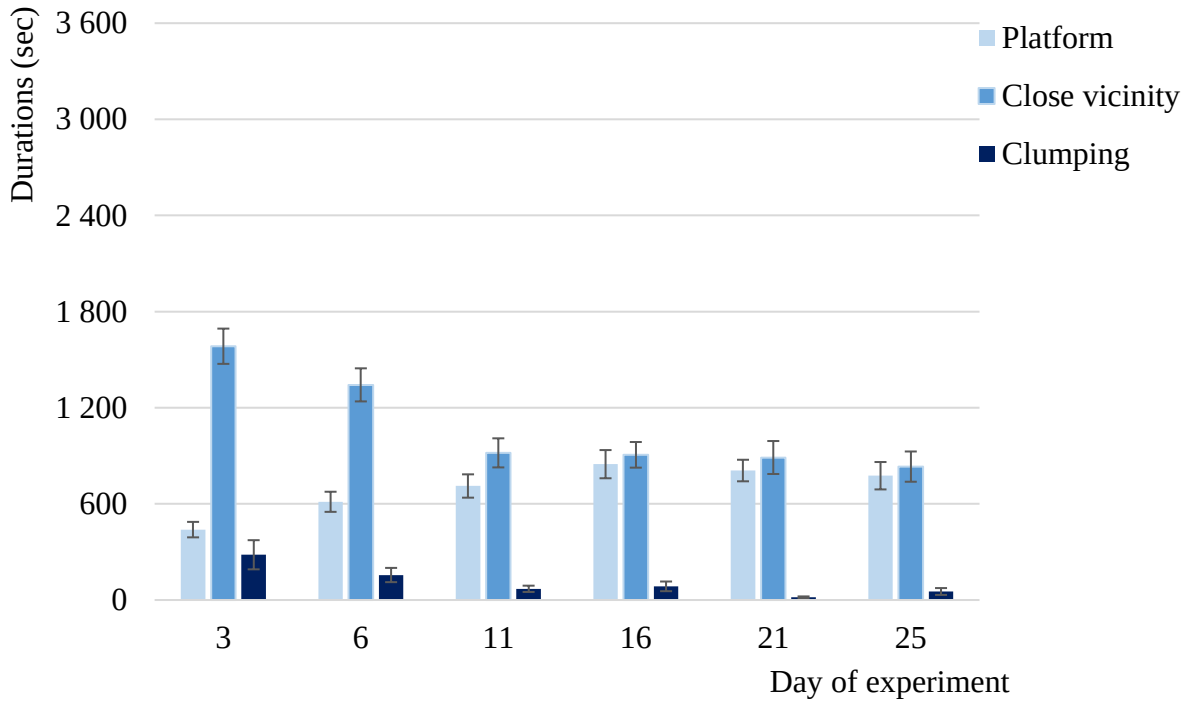


Figure 7. Evolution of durations (mean ± SE) spent in proximity to the robot over the experiment. For description of categories platform, close vicinity and clumping, see Table 2.

Each pupil had a different behavioural profile. For example, one bird (C10) spent only 53 s on the platform and no time in close vicinity on the 9 h 30 min of video analysed. He spent most of the time on other perches in the cage and on the ground.

Four birds out of 12 (C5, C6, C9 and C10) exhibited less than 10 s in clumping while 2 birds both spent 1 h 29 min in clumping on the total 9 h 30 min (C3: 1 h 29 min 18 sec, C4: 1 h 29 min 45 sec).

Comparisons of day 5 versus day 6 and day 10 versus day 11:

We also compared these 3 categories of behaviours between days 5 and 6 and between days 10 and 11 to analyse if we observed an upsurge of interest from pupils after two days without exposure to the robot. Regardless of the category, we found no difference between days 5 and 6 (t-test for platform durations: t = -0.08, df = 11, p = 0.43, t-test for close vicinity durations: t = 0.0007, df = 11, p = 0.99, Wilcoxon test for clumping durations: V = 10, p = 1) and between days 10 and 11 (t-test for platform durations: t = -0.69, df = 11, p = 0.50, t-test for close vicinity durations: t = 0.49, df = 11, p = 0.63, Wilcoxon test for clumping durations: V = 6, p = 0.40).

3.3.2.2. Number of pecking on the robot

The number of pecking on the robot's head significantly increased until day 11 and then decreased over the experiment, in an inverted U way (linear component: t(69) = 4.38, $p < 0.001$, quadratic component: t(69) = -3.55, $p < 0.001$, Figure 8). Similarly, the number of pecking on the robot's body significantly increased until day 16 then decreased over the experiment, in an inverted U way (linear component: t(69) = 18, $p < 0.001$, quadratic component: t(69) = -19, $p < 0.001$, Figure 8).

We observed a high inter-individual variability: the pupil C10 never pecked the robot whereas pupils C1 and C8 pecked respectively 1 014 and 1 305 times during the 10 sessions analysed.

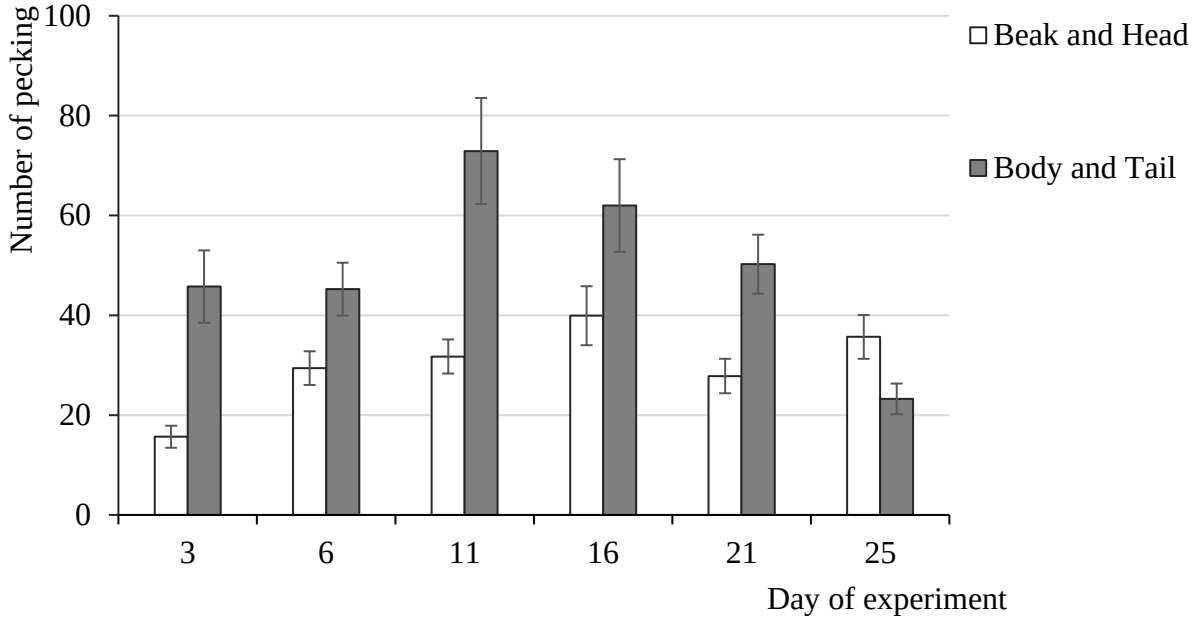


Figure 8. Evolution of pecking behaviour (mean ± SE) according to the different parts of the robot over the experiment.

3.3.2.3. Reactions to the robot songs

Following a song produced by the robot, “stays still” reactions were the most frequent ones (mean ± SE: mean number of “stays still”: 3.86 ± 0.22, mean number of “stands up”: 1.95 ± 0.15, mean number of “moves”: 0.53 ± 0.11, mean number of “flies away orientated”: 1.83 ± 0.18, mean number of “flies away”: 0.27 ± 0.03, Figure 13 in Supplementary Information).

The proportion of attentive reactions (“stays still” and “stands up” reactions) did not change significantly with time (linear component: $t(59) = -1.59$, $p = 0.12$; quadratic component: $t(59) = 1.51$, $p = 0.13$, Figure 9).

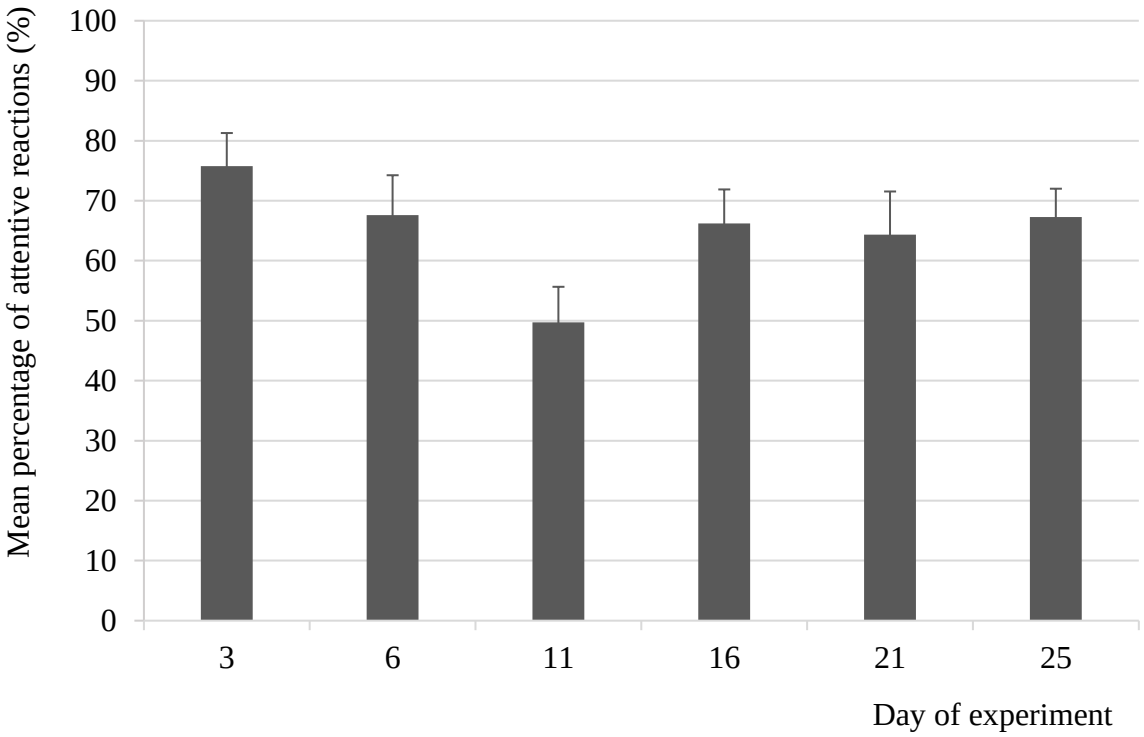


Figure 9. Evolution of proportion of "attentive reactions" (mean ± SE) to robot's songs over the experiment.

3.3.2.4. Evolution of number of calls over the experiment

The number of calls did not change significantly with time (linear component: t(297) = -0.98, p = 0.32; quadratic component: t(297) = 0.87, p = 0.38, Figure 10).

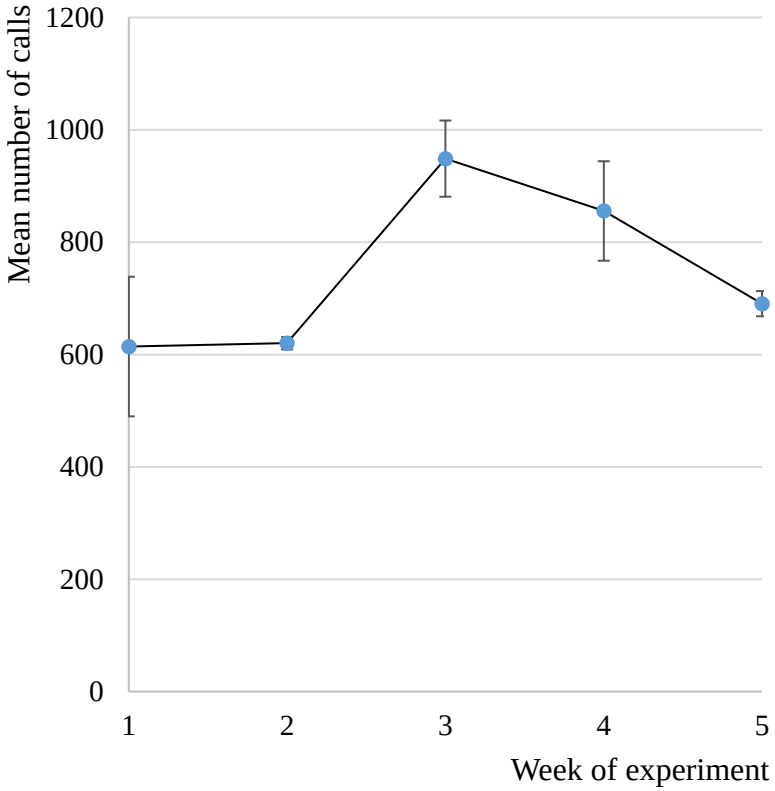


Figure 10. Evolution of the number of calls produced by the robot (mean proximity scores ± SE) over the experiment.

3.4.Effects of behaviours on song similarity

3.4.1.Live group

We attempted to characterize the relationship between proximity scores and song similarity through a GLM with mean proximity scores as a linear and quadratic predictor but none of the components reached significance (linear component: t-test t(9) = 1.11, p = 0.30; quadratic component: t(9) = -1.09, p = 0.30, Figure 11).

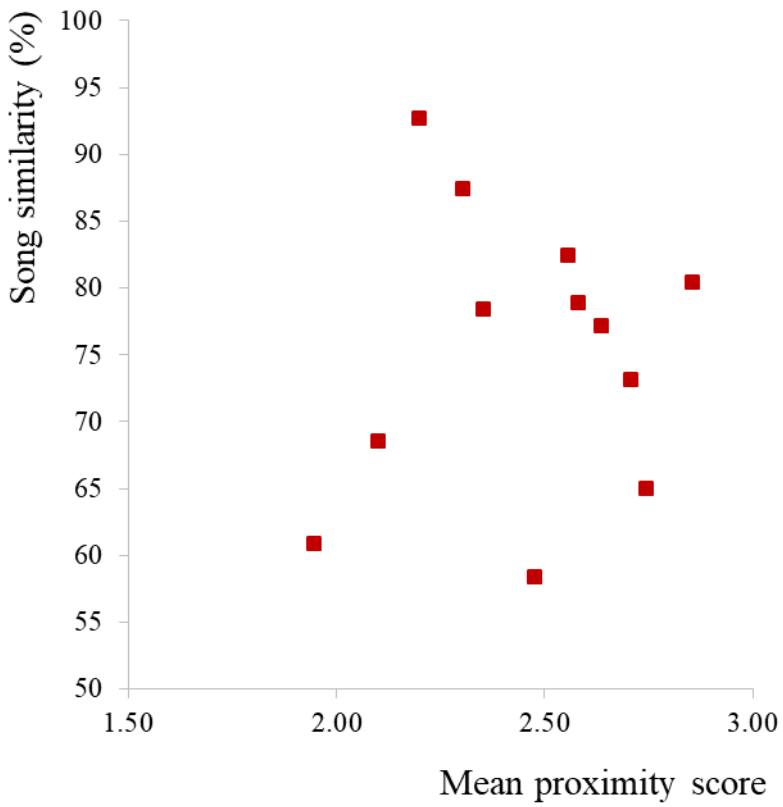


Figure 11. Song similarity as a function of mean proximity score over the experiment.

3.4.2.Contingent group

We attempted to characterize the relationship between time on the platform and song similarity through a GLM with time on the platform as a linear and quadratic predictor. Similarity increased linearly with the time on the platform (t-test t(9) = 2.52, p = 0.03) but not quadratically (t(9) = -1.91, p = 0.09, Figure 12A).

We reproduced this analysis with the time spent in close vicinity on the platform as a predictor, and independently, with the time spent in clumping. Neither the time spent close (linear component: t(9) = -1.73, p = 0.11; quadratic component: t(9) = 1.61, p = 0.14, Figure 12B), nor the time spent in clumping (linear component: t(9) = 1.22, p = 0.25; quadratic component: t(9) = -1.16, p = 0.28, Figure 12C) was related to song similarity.

We conducted a similar analysis with the number of pecking on the beak or head of the robot, as a predictor, and independently, with the number of pecking on the body or tail and the proportion of attentive reactions. Similarity increased with the number of pecking on beak or head and then decreased in an inverted U (linear component: $t(9) = 3.10$, $p = 0.01$; quadratic component: $t(9) = -2.35$, $p = 0.04$, Figure 12E) but did not vary with the number of pecking on body or tail (linear component: $t(9) = 1.08$, $p = 0.31$; quadratic component: $t(9) = -0.46$, $p = 0.65$, Figure 12F) or the proportion of attentive reactions (linear component: $t(7) = -0.26$, $p = 0.80$; quadratic component: $t(7) = 0.068$, $p = 0.95$, Figure 12D).

We conducted a similar analysis with the number of calls produced as a predictor but the similarity did not vary with the number of calls (linear component: $t(9) = -0.07$, $p = 0.94$; quadratic component: $t(9) = 0.27$, $p = 0.79$, Figure 12G).

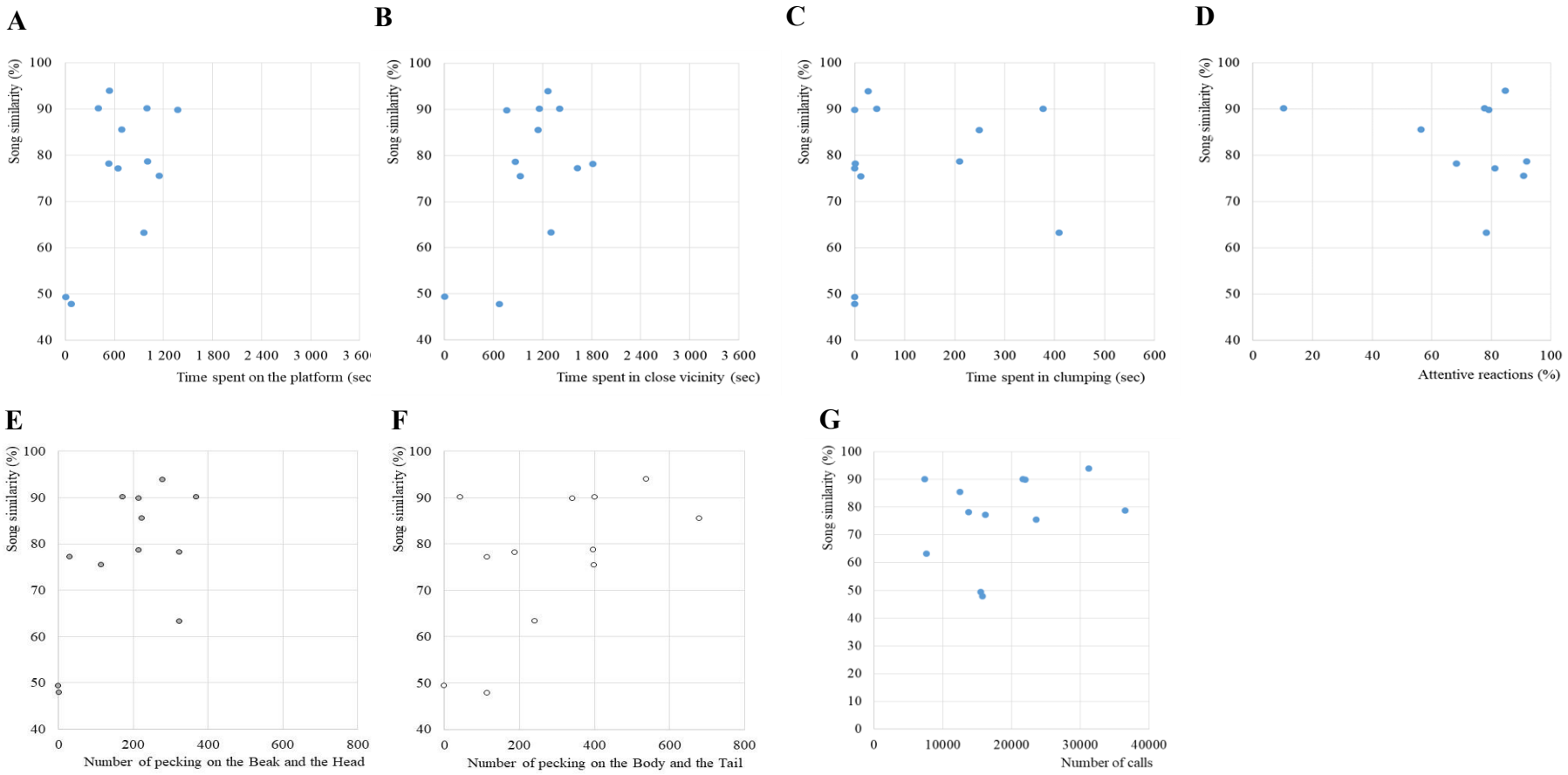


Figure 12. For the contingent group, song similarity (%) as a function of (A) the time spent on the platform (sec), (B) the time spent in close vicinity (sec), (C) the time spent in clumping (sec), (D) the proportion of attentive reactions (%), (E) the number of pecking on the robot's beak and head, (F) the number of pecking on the body and tail and (G) the number of calls produced by the robot.

4.Discussion

Our results are the first experimental demonstration that a robotic finch can be successfully used a song tutor for young finches. An experimental procedure consisting to expose a young finch to a tutor during the sensitive period for song learning, during one hour, 5 days a week during 5 weeks, is sufficient to trigger a good song imitation. Mimicking this experimental procedure by exposing young finches to a robotic finch exhibiting contingent behaviour, we did not observe a significant difference with a group of finches exposed to a live tutor.

Moreover, we did not find any significant differences between the two groups regarding other acoustic parameters as song linearity, bout duration and the number of syllables in songs. As in a previous study of Tchernichovski and al. (1999), we found that song imitation was linked to the number of songs heard by pupils in an inverted U way, meaning that an overexposure would not lead to better imitation. We noticed that the top of the inverted U was reached faster in the contingent group than in the live group suggesting that the zebra finch robot is particularly efficient for song learning.

From the first session analysed (day 2), pupils exhibited behaviours towards the robot that testify of its acceptance as a social agent. Pupils spent on average 70% of the 30 min-session on the platform and 65% in close vicinity. Regarding physical interactions, 4 birds out of 12 exhibited clumping behaviour to the robot. One of them, C4, spent 73.96% of the session clumped to the robot. Clumping is a specific behaviour, often used to evaluate bonding with mates and others social partners (Tomaszycki & Zatirka, 2014; Zann, 1996). Moreover, 9 birds out of 12 pecked the robot and 57.19% ± 2.38 of the total number of pecking was on the robot's body. Attentive reactions (the pupil stayed still or stood up after hearing the robot's song) were also predominant during the first session.

Over the experiment, time spent on the platform tended to increase and on day 25, pupils spent on average more than 45 % of the session on the platform. We observed a peak of pecking behaviours on days 11 and 16 then they slightly decreased. Attentive reactions to the robot's songs remained the most frequent and did not decrease. Taken together, these results showed that pupils remained interested in the robot despite the repetition of sessions.

We evaluated whether song imitation could be explained by pupils' behaviours towards the robot. We found that song similarity was linked with the number of pecking on the robot's beak and head in an inverted U way. As young male finches seem to solicit their song tutors to sing by pecking their beak (Adret, 1992), we propose that pupils who pecked on the beak and the head may be more motivated to learn songs. Other experiments have to be done to confirm this result.

We also evaluated the influence of pupils' attention to the robot's songs on song similarity. In contrary to the study of Chen and al. (2016), we found no correlation between the proportion of attentive reactions and song similarity. A possible explanation is that Chen and collaborators analysed only the first 5 days of pupil-tutor interactions whereas our analysis covered 6 days out of the 25 of the experiment. Attention to songs may influenced song similarity during the first 5 days but then this effect is less important. Furthermore, in Chen and collaborators' study, pupils could not physically interact with their tutors. We can also assume that attention to songs is different when pupils can have physical contacts with their tutor as in our experiment.

We assessed the influence of physical proximity on song learning. We observed two different trends. Even if pupils of the live group and their tutor were unfamiliar at the beginning of the experiment, they tended to get closer and closer over the experiment. On the contingent group, pupils spent increasing time on the platform but tended to spend less time in close contact with the robot as close vicinity and clumping decreased. The only link between pupil-tutor

physical proximity and song learning is in the contingent group where song similarity linearly increased with the time spent on the platform.

First of all, our sampling effort could be increased. For the live group, both birds freely interacted and permanently moved in the cage which complicated behavioural analysis. Assessing a proximity score each minute was an approximation and a more detailed and automated behaviour analysis including proximity durations, behaviours oriented towards the two partners to evaluate the quality of the relationship between the pupil and his tutor could conduct to different results. On the contingent group, behavioural analysis was more feasible as the robot remained at the same place in the cage and we could increase the number of coded sessions.

Moreover, beyond physical proximity, we make the hypothesis that the tutor's contingency to the pupil's behaviours impacts song learning more notably. To assess the influence of contingency, we will further test a non-contingent group. As previously described, 12 young male finches will be exposed to the robot playing the same sequence of movements and sounds produced in the contingent group. We predict that pupils of the non-contingent group will produce less quality songs than in the contingent and live group. Since we can control the degree of contingency, for example by varying the latency of the robot's responses, we could further examine more precisely the influence of contingency on song learning.

In animal-robot interactions studies, appearance and movements of the robot are often tested together and several studies showed that movements produced by the robot favoured attraction of conspecifics (Romano et al., 2019 for a review). Such characteristics are also important in human-robot interaction (Hameed & Tan, 2016; Hosseini et al., 2017). Using the finch robot, we could evaluate independently the impact of different factors on preference and song imitation such as the contingency between the robot's beak openings and sound

production. We could also manipulate the appearance of the robot (body colours, robot covered with a taxidermic skin covered with feathers…).

As often reviewed in the literature, song learning is a complex phenomenon with an important inter individual variability between young zebra finches (Derégnaucourt, 2011). Several authors showed that genetic factors have a strong influence on song learning but a recent study by Mets & Brainard (2019) highlighted that genetic effects can be balanced by exposing pupils Bengalese finches (*Lonchura striata domestica*) to songs specifically created to corresponding to their learning styles. This observation could be very useful for future experiments as we could use the zebra finch robot to develop experimental procedures to balance genetic predispositions.

Furthermore, animal robot could offer a solution to face challenges linked to endangered species. For example, the population of regent honeyeater (*Anthochaera phrygia),* an Australian songbird, severely decreased in the past years to the point that the number of conspecific song tutors is not sufficient anymore to transmit vocal culture to future generations of birds. Therefore, young males learn songs from other species which conduct to a loss of vocal culture (Crates et al. 2021). If song playback by loudspeakers could not trigger significant learning, robotic birds could be used to broadcast regent honeyeaters' songs in the wild and could be adopted as song tutors to preserve the vocal culture of the species.

In conclusion, our promising results show that a robot-finch could be considered as a valid song tutor for young finches, triggering better results of song imitation than those obtained with other experimental procedures in the laboratory so far (Derégnaucourt et al., 2013, Derégnaucourt, Araguas & Guellaï, in preparation). The use of robots in the context of vocal learning could open new perspectives of research and applications in different fields. Such applications could benefit both humans and non-human animals.

Acknowledgments

We thank Félix Bigand, Perrine Marsac, Philippine Prevost, Marie Soret and Lisa Jacquey for their help regarding the development of the vocal loop and Adeline Depierreux, Camille Nozières, Antoine Thomas, Khedidja Athamnia and Margaux Capezzera for their help in analysing data. We thank Philippe Groué, Emmanuelle Martin, Ophélie Bouillet, Marie Huet, Priscilla Roussel, Anaëlle Brunet and Margaux Laurent for taking care of the birds.

References

Adret, P. (1992). Vocal Learning Induced With Operant Techniques : An Overview. *Netherlands Journal of Zoology*, *43*, 125-142. https://doi.org/10.1163/156854293X00250

Araguas, A., Guellaï, B., Gauthier, P., Richer, F., Montone, G., & Derégnaucourt, S. (2021). Design of a robotic zebra finch for experimental studies on developmental song learning. *Journal of Experimental Biology* (submitted)

Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A., & Tchernichovski, O. (2016). The Forebrain Song System Mediates Predictive Call Timing in Female and Male Zebra Finches. *Current biology : CB*, *26*(3), 309-318. https://doi.org/10.1016/j.cub.2015.12.037

Butler, S., & Fernández-Juricic, E. (2014). European starlings recognize the location of robotic conspecific attention. *Biology letters*, *10*. https://doi.org/10.1098/rsbl.2014.0665

Carouso-Peck, S., & Goldstein, M. H. (2019). Female Social Feedback Reveals Non-imitative Mechanisms of Vocal Learning in Zebra Finches. *Current Biology: CB*, *29*(4), 631-636.e3. https://doi.org/10.1016/j.cub.2018.12.026

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), 6641-6646. https://doi.org/10.1073/pnas.1522306113

Crates, R., Langmore, N., Ranjard, L., Stojanovic, D., Rayner, L., Ingwersen, D., & Heinsohn, R. (2021). Loss of vocal culture and fitness costs in a critically endangered songbird. *Proceedings. Biological Sciences*, *288*(1947), 20210225. https://doi.org/10.1098/rspb.2021.0225

Daou, A., Johnson, F., Wu, W., & Bertram, R. (2012). A computational tool for automated large-scale analysis and measurement of bird-song syntax. *Journal of neuroscience methods*, *210*, 147-160. https://doi.org/10.1016/j.jneumeth.2012.07.020

Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the Zebra Finch (*Taeniopygia guttata*). *Interaction Studies*, *12*(2), 324–350. https://doi.org/10.1075/is.12.2.07der

Derégnaucourt, S., Araguas, A. & Guellaï, B. (2021). Technological advances for getting insight into the learning capacities of birds in the vocal domain. *Interaction Studies*. (in preparation)

Derégnaucourt, S., & Gahr, M. (2013). Horizontal transmission of the father's song in the zebra finch (Taeniopygia guttata). Biology Letters, 9(4), 20130247. https://doi.org/10.1098/rsbl.2013.0247

Derégnaucourt, S., Poirier, C., Kant, A. V. der, Linden, A. V. der, & Gahr, M. (2013). Comparisons of different methods to train a young zebra finch (Taeniopygia guttata) to learn a song. *Journal of Physiology, Paris*, *107*(3), 210-218. https://doi.org/10.1016/j.jphysparis.2012.08.003

Hameed, I. A., Tan, Z.-H., Thomsen, N., Duan, X. (2016). *User acceptance of social robots*. ACHI 2016: The Ninth International Conference on Advances in Computer-Human Interactions. pp. 274-279

Hauber, M. E., Louder, M. I., & Griffith, S. C. (2021). Neurogenomic insights into the behavioral and vocal development of the zebra finch. *eLife*, *10*, e61849. https://doi.org/10.7554/eLife.61849

Hosseini, S. M. F., Hilliger, S., Barnes, J., Jeon, M., Park, C. H., & Howard, A. M. (2017). Love at first sight : Mere exposure to robot appearance leaves impressions similar to interactions with physical robots. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 615-620. https://doi.org/10.1109/ROMAN.2017.8172366

Hyland Bruno, J., & Tchernichovski, O. (2017). Regularities in zebra finch song beyond the repeated motif. *Behavioural Processes*, *163*, 53-59. https://doi.org/10.1016/j.beproc.2017.11.001

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. *Bird Vocalizations*, *61*.

Iyengar, S., & Bottjer, S. (2002). The Role of Auditory Experience in the Formation of Neural Circuits Underlying Vocal Learning in Zebra Finches. *The Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.22-03-00946.2002

Jolly, L., Pittet, F., Caudal, J.-P., Mouret, J.-B., Houdelier, C., Lumineau, S., & Margerie, E. (2016). Animal-to-robot social attachment : Initial requisites in a gallinaceous bird. *Bioinspiration & biomimetics*, *11*, 016007. https://doi.org/10.1088/1748-3190/11/1/016007

Kao, M. H., & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *Journal of Neurophysiology*, *96*(3), 1441-1455. https://doi.org/10.1152/jn.01138.2005

King, A. P., & West, M. J. (1989). Presence of female cowbirds (Molothrus ater ater) affects vocal imitation and improvisation in males. *Journal of Comparative Psychology*, *103*(1), 39-44. https://doi.org/10.1037/0735-7036.103.1.39

Lachlan, R. F., van Heijningen, C. A. A., Ter Haar, S. M., & Ten Cate, C. (2016). Zebra Finch Song Phonology and Syntactical Structure across Populations and Continents-A Computational Comparison. *Frontiers in Psychology*, *7*, 980. https://doi.org/10.3389/fpsyg.2016.00980

Lerch, A., Roy, P., Pachet, F., & Nagle, L. (2011). Closed-loop bird-computer interactions : A new method to study the role of bird calls. *Animal cognition*, *14*, 203-211. https://doi.org/10.1007/s10071-010-0353-6

Ljubičić, I., Hyland Bruno, J., & Tchernichovski, O. (2016). Social influences on song learning. *Current Opinion in Behavioral Sciences*, *7*, 101-107. https://doi.org/10.1016/j.cobeha.2015.12.006

Mets, D. G., & Brainard, M. S. (2019). Learning is enhanced by tailoring instruction to individual genetic differences. *ELife*, *8*, e47216. https://doi.org/10.7554/eLife.47216

Patricelli, G. L., Uy, J. A. C., Walsh, G., & Borgia, G. (2002). Male displays adjusted to female's response. *Nature*, *415*(6869), 279-280. https://doi.org/10.1038/415279a

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *Journal of Comparative and Physiological Psychology*, *93*(2), 260-277. https://doi.org/10.1037/h0077553

Romano, D., Benelli, G., Donati, E., Remorini, D., Canale, A., & Stefanini, C. (2017). Multiple cues produced by a robotic fish modulate aggressive behaviour in Siamese fighting fishes. *Scientific Reports*, *7*(1), 4667. https://doi.org/10.1038/s41598-017-04840-0

Romano, D., Donati, E., Benelli, G., & Stefanini, C. (2019). A review on animal-robot interaction : From bio-hybrid organisms to mixed societies. *Biological Cybernetics*, *113*(3), 201-225. https://doi.org/10.1007/s00422-018-0787-5

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype Me ! Socially Contingent Interactions Help Toddlers Learn Language. *Child Development*, *85*(3), 956-970. https://doi.org/10.1111/cdev.12166

Scharff, C., & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system : Implications for vocal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *11*, 2896-2913. https://doi.org/10.1523/JNEUROSCI.11-09-02896.1991

Sparks, J. H. (1964). Flock structure of the red avadavat with particular reference to clumping and allopreening. *Animal Behaviour*, *12*(1), 125-136. https://doi.org/10.1016/0003-3472(64)90113-7

Takahashi, D. Y., Liao, D. A., & Ghazanfar, A. A. (2017). Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Current Biology*, *27*(12), 1844-1852.e6. https://doi.org/10.1016/j.cub.2017.05.004

Tchernichovski, O., Lints, T., Mitra, P. P., & Nottebohm, F. (1999). Vocal imitation in zebra finches is inversely related to model abundance. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(22), 12901-12904.

Tchernichovski, O., & Nottebohm, F. (1998). Social inhibition of song imitation among sibling male zebra finches. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 8951-8956. https://doi.org/10.1073/pnas.95.15.8951

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, *59*(6), 1167-1176. https://doi.org/10.1006/anbe.1999.1416

Tomaszycki, M., & Zatirka, B. (2014). Same-Sex Partner Preference in Zebra Finches : Pairing Flexibility and Choice. *Archives of sexual behavior*, *43*. https://doi.org/10.1007/s10508-014-0377-0

West, M. J., & King, A. P. (1988). Female visual displays affect the development of male song in the cowbird. *Nature*, *334*(6179), 244-246. https://doi.org/10.1038/334244a0

Zann, R. A. (1996). *The zebra finch: A synthesis of field and laboratory studies* (Vol. 5). Oxford University Press.

Zevin, J. D., Seidenberg, M. S., & Bottjer, S. W. (2004). Limits on Reacquisition of Song in Adult Zebra Finches Exposed to White Noise. *Journal of Neuroscience*, *24*(26), 5849-5862. https://doi.org/10.1523/JNEUROSCI.1891-04.2004

Supplementary Information

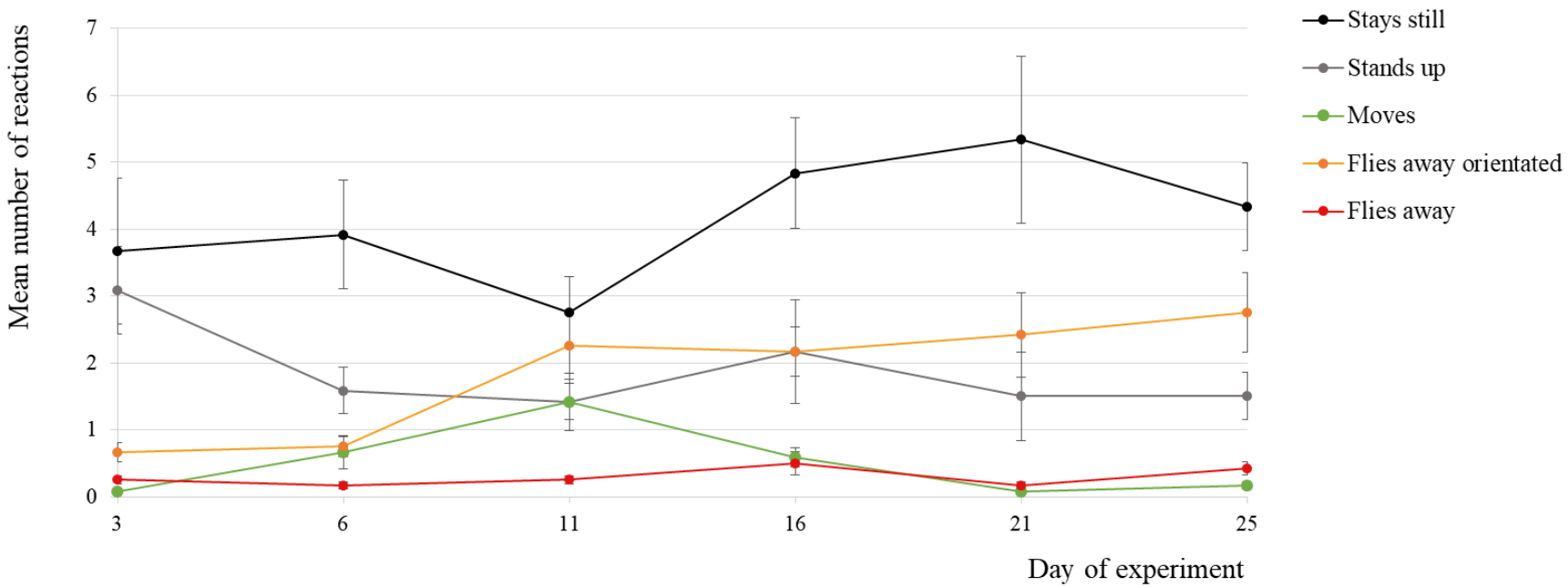


Figure 13. Evolution of pupils' reactions to robot's songs (mean ± SE) over the experiment.

DISCUSSION GENERALE

1. Résumé des résultats principaux, interprétations et limites

Nos trois expériences ont en commun la comparaison des interactions entre agent biologique-robot à celles entre deux agents biologiques de la même espèce. Trois populations différentes ont été testées : des enfants au développement typique (DT), des enfants présentant un Trouble du Spectre Autistique (TSA) léger ainsi que de jeunes diamants mandarins (*Taeniopygia guttata*).

Chez les enfants, le protocole expérimental des deux études réalisées avec les enfants a été pensé pour être accessible à la fois à des enfants DT et à des enfants TSA et englobe différents domaines de compétences (vocabulaire, conscience corporelle et comportement pro-social). Nous voulions, en effet, réaliser une étude comparative entre ces deux groupes. Nous avons trouvé que chez les enfants DT comme chez les enfants TSA, les scores aux différentes tâches étaient comparables que la démonstratrice soit un adulte ou le robot NAO. Chez les enfants DT, une amélioration des scores avec l'âge a été observée dans la majorité des tâches. Les temps de regard vers les démonstratrices humaine et robotique sont comparables dans toutes les tâches, à l'exception de la tâche des postures émotionnelles dans laquelle les enfants DT regardent davantage la démonstratrice robotique. Chez les enfants TSA, les temps de regard vers la démonstratrice robotique est significativement supérieur au temps de regard vers la démonstratrice humaine durant la familiarisation avec la comptine, la tâche d'imitation et la tâche d'identification des postures émotionnelles.

Chez le Diamant Mandarin, les résultats de notre étude tendent à montrer qu'un robot interactif (le MANDABOT) peut être accepté comme un tuteur de chant par de jeunes mâles pendant la période sensible d'apprentissage vocal. Les résultats d'imitation des modèles de chant diffusés par le robot sont bons et comparables à ceux obtenus chez de jeunes oiseaux élevés seuls avec un mâle adulte (scores de similarité acoustique, moyenne ± erreur standard,

dans le groupe tuteur robot interactif : 76.63% ± 2.25 et dans le groupe tuteur adulte : 75.30% ± 1.51).

La pandémie de Covid-19 a beaucoup affecté le déroulement de cette thèse. La durée des expériences et le nombre d'enfants TSA inclus dans notre étude ont été restreints. Chez le Diamant Mandarin, le suivi d'un groupe d'oiseaux entrainés avec un robot non interactif n'a pas pu être réalisé. Ainsi, sans ce groupe contrôle, nous ne pouvons pas conclure quant à l'importance de la contingence comportementale du robot sur le succès d'imitation.

1.1. Interactions enfant-robot dans un contexte d'apprentissage, de contrôle des connaissances sur le corps et de mesure de la pro-socialité: comparaison avec un partenaire humain

Au cours de cette étude inter-sujets, nous avons comparé les comportements d'enfants au développement typique (DT), âgés de 3 à 6 ans, en présence d'un instructeur délivrant des consignes: le robot humanoïde NAO (féminisé pour l'étude et appelé « NAOMIE », groupe robot) ou une adulte non familière (groupe humain). Les domaines testés étaient : 1) l'apprentissage de vocabulaire nouveau, 2) la connaissance des parties du corps, composée de plusieurs sous-tâches : la compréhension des noms des différentes parties du corps sur soi et sur la démonstratrice (robot ou humain), l'imitation des parties du corps pointées par la démonstratrice (l'enfant devait, dans un premier temps, nommer les parties du corps montrées par la démonstratrice puis imiter les séquences de parties du corps montrées), 3) la reconnaissance de quatre émotions primaires (joie, peur, tristesse, colère) sur différents supports (des images puis les postures de la démonstratrice) et l'expression des émotions par le corps de l'enfant et enfin, 4) une tâche d'aide spontanée de la démonstratrice. Les tâches expérimentales débutaient et se terminaient pas le même questionnaire visant à évaluer les caractéristiques et capacités attribuées par les enfants à la démonstratrice. Nous avons également analysé les temps

de regard vers la démonstratrice, l'endroit où l'enfant regardait lorsqu'il répondait aux questions posées par la démonstratrice durant les tâches expérimentales ainsi que le nombre de fois que l'enfant touchait le corps de la démonstratrice.

Nous avons constitué trois groupes d'âge (groupe 1 : 3 ans et 1 mois à 4 ans et 1 mois, groupe 2 : 4 ans et 2 mois à 5 ans et 1 mois, groupe 3 : 5 ans et 2 mois à 6 ans et 1 mois). Par la suite, le protocole a été adapté aux enfants du groupe 1 en enlevant la tâche de vocabulaire et en diminuant le nombre de questions dans les questionnaires pré et post tests (7 questions sur 17 ont été conservées).

Notre hypothèse principale était que les enfants des deux groupes réagiraient de manière comparable, ce que nous avons validé. En effet, nous n'avons pas trouvé d'effet du type de démonstratrice dans les tâches réalisées, les enfants ont répondu de manière similaire aux consignes qu'elles leur aient été données par le robot ou l'adulte. Notre seconde hypothèse était que les enfants les plus âgés auraient de meilleures performances que les enfants les plus jeunes.

Ceci a été vérifié dans la majorité des tâches expérimentales et les questionnaires pré et post test. Notons que le temps de réponse dans la tâche de vocabulaire, la reconnaissance des parties du corps sur soi, l'identification des émotions sur les images, l'expression des émotions ainsi que le temps mis pour ramasser le bâton dans la tâche d'aide spontanée ne sont pas influencés par l'âge des enfants.

Les tâches « reconnaissance des parties du corps sur soi » et « identification des émotions sur les images » ont obtenu des scores très élevés et on observe un effet plafond. Ceci est cohérent avec les données de la littérature, qui indiquent que les parties du corps que nous avons testées étaient en général connues entre 2,2 ans et 3,4 ans (Camões-Costa et al., 2011) et que les six émotions primaires définies par Eckman & Friesen (1976) à savoir la joie, la peur, la colère, la tristesse, la surprise et le dégoût, étaient reconnues à 3 ans (Widen & Russel, 2003). Les scores obtenus dans la tâche d'expression des émotions pourraient être expliqués par les

différences de personnalité et notamment la tendance à l'introversion/extraversion des enfants, paramètre que nous n'avons pas mesuré et qui impacterait fortement la capacité à exprimer ses émotions (Riggio & Riggio, 2002). Enfin, le temps de réaction dans la tâche d'aide spontanée pourrait être expliqué par d'autres variables comme le fait que les enfants n'aient pas clairement perçu la demande implicite de la démonstratrice (ramasser le bâton et lui rendre) ou encore la personnalité des enfants. En effet, les chercheurs en psychologie de la personnalité ont montré qu'il existait des liens entre certains traits de personnalité et le comportement pro-social (compris comme des actions interpersonnelles réalisées dans un contexte socio-culturel défini et qui bénéficient aux autres individus, (Dovidio et al., 2017). Par exemple, dans le modèle en cinq facteurs de McCrae & Costa (1997) (comprenant l'Ouverture à l'expérience, la Conscience, l'Extraversion/introversion, l'Agréabilité et le Névrosisme), l'agréabilité, soit « la manière dont les individus diffèrent en termes d'attitude dans les relations interpersonnelles » (Graziano & Tobin, 2009) est le facteur le plus souvent mis en lien avec les comportements d'aide (Graziano et al., 2007).

Le temps de regard en direction de la démonstratrice (robotique ou humaine) durant la totalité de l'expérience n'a pas été influencé ni par le type de démonstratrice ni par l'âge. Cependant, dans la tâche d'identification des postures émotionnelles de la démonstratrice, les enfants ont davantage regardé la démonstratrice dans le groupe robot que dans le groupe humain. Ceci laisse penser qu'ils avaient besoin de plus de temps pour décrypter la posture du robot.

Nous avons également retrouvé que lorsqu'il s'agissait de répondre aux questions posées par la démonstratrice pendant l'expérience, les enfants regardaient plus l'expérimentatrice en répondant quand ils étaient face au robot que quand ils étaient face à l'adulte. Notre hypothèse est que les enfants étaient moins sûrs de leurs réponses et avaient besoin de chercher davantage une validation extérieure quand ils étaient face au robot. Ceci

concorde avec une précédente étude indiquant que les enfants regardent plus l'adulte qui les accompagne (dans cette étude, leur parent) lorsqu'ils interagissent avec le robot NAOMIE que lorsqu'ils interagissent avec un adulte (Tolksdorf et al., 2021).

Le nombre de contacts physiques de la part des enfants était plus important à la fois dans la condition robot et chez les enfants les plus âgés. Nous faisons l'hypothèse que toucher un adulte inconnu est inhabituel pour les enfants alors que le robot est plus perçu comme un objet que les enfants peuvent toucher.

Pour finir, nous avons déterminé, par les résultats aux questionnaires pré et post tests, que les enfants avaient une représentation plus homogène des catégories d'appartenance et des capacités de la démonstratrice humaine que du robot. De plus, il a été démontré que lorsqu'un robot présente des caractéristiques humaines, les enfants s'attendent à ce qu'il se comporte comme un humain (Woods, 2006).

Une des principales applications de ce type d'étude est de déterminer dans quelle mesure les robots sociaux peuvent être utilisés auprès des enfants par les enseignants. En ce sens, notre étude présente plusieurs limites. Nous ne sommes pas parvenus à équilibrer complètement les groupes expérimentaux selon l'âge, le genre et la condition expérimentale : le groupe 1 (enfants les plus jeunes) est composé de 14 enfants contre 26 enfants dans le groupe 2 et 22 enfants dans le groupe 3. Nous avons eu plus de difficultés à faire passer l'expérience aux enfants du groupe 1, soit parce que les enfants ne voulaient pas venir dans la salle où se déroulait l'expérience, soit parce qu'ils demandaient à arrêter la passation en cours.

Au-delà de cette première impression, maintenir l'engagement et l'intérêt des enfants tout au long de l'interaction représente un des défis de l'utilisation des robots sociaux à l'école. Nous avons observé que plusieurs enfants demandaient si l'expérience était bientôt terminée et ont exprimé leur déception face aux capacités limitées du robot. De plus, à la fin des expériences, nous avons présenté le protocole et le robot en groupe aux enfants qui n'avaient

pas pu participer, soit parce que leurs parents ne les y avaient pas autorisés, soit parce qu'ils étaient trop âgés. Ceci nous a permis d'échanger plus spontanément avec les enfants concernant leur perception du robot. Plusieurs enfants ont dit que le robot faisait peur, à cause de son regard fixe. Pour Tolksdorf et ses collègues (2021), il faut garder à l'esprit que les enfants constituent une population vulnérable : leur perception, leur besoin de communication et leur expression et régulation des émotions sont différents de ceux des adultes. En cela, il est indispensable d'être attentifs à ce qu'ils expriment pendant les expériences et de prendre en compte les éventuelles conséquences de ce type d'étude.

Prises ensemble, ces observations nous laissent penser que tester des enfants de ces tranches d'âge en utilisant un médiateur inconnu et potentiellement impressionnant comme le robot, nécessite sûrement un temps de familiarisation avec les démonstratrices et l'expérimentatrice plus important. Intégrer les démonstratrices dans des activités routinières en classe permettrait que les enfants sachent ce que peut faire le robot avant l'expérience, qu'ils puissent en parler avec leurs enseignants et leurs parents. Tolksdorf et collègues (2021) suggèrent également d'intégrer un accompagnateur familier (un adulte qui s'occupe de lui à l'école comme l'enseignant ou l'Agent Territorial Spécialisé des Ecoles Maternelles) lorsque l'on réalise des études sur les interactions entre un enfant de moins de 6 ans et un robot plutôt que de demander à l'enfant de réaliser l'expérience seul avec l'expérimentateur. En effet, il a été mis en évidence que les enfants de 5 ans ont besoin d'un soutien émotionnel important de la part de leur accompagnateur pour réguler leurs émotions dans cette situation nouvelle qu'est l'interaction avec le robot NAO et que l'accompagnateur joue un rôle majeur dans le maintien de l'interaction avec le robot (Rohlfing et al., 2020).

1.2. Conscience corporelle et comportement pro-social chez des enfants TSA, interagissant avec un humain ou un robot

Notre seconde étude a été réalisée auprès de six enfants présentant un Trouble du Spectre Autistique léger (correspondant à un besoin de soutien minimal par rapport aux autres formes de TSA, comme défini dans le DSM-5 (American Psychiatric Association, 2013), âgés de 4 à 8 ans. Les tâches proposées étaient identiques à certaines de l'étude réalisée avec les enfants DT : 1) la connaissance des parties du corps avec la compréhension des noms des différentes parties du corps sur soi et sur la démonstratrice (humaine ou robotique), l'imitation des parties du corps pointées par la démonstratrice (nommer les parties du corps montrées par la démonstratrice puis imiter les séquences de parties du corps montrées), 2) la reconnaissance de quatre émotions primaires (joie, peur, tristesse, colère) sur différents supports (des images puis les postures de la démonstratrice) et l'expression des émotions par le corps de l'enfant et enfin, 3) une tâche d'une tâche d'aide spontanée de la démonstratrice. Nous avons également analysé les temps de regard vers la démonstratrice, l'endroit où l'enfant regardait lorsqu'il répondait aux questions posées par la démonstratrice durant les tâches expérimentales ainsi que le nombre de fois que l'enfant touchait le corps de la démonstratrice.

Nous avons mené l'étude en intra-sujet et sur plusieurs séances afin de déterminer si la répétition des interactions avec les démonstratrices impactait favorablement leurs réponses. Chaque enfant a participé à quatre sessions réparties sur deux semaines consécutives, à raison de 2 sessions par semaine : deux sessions avec le robot NAOMIE comme démonstratrice (condition robot) et deux sessions avec l'adulte non familier comme démonstratrice (condition humain), attribuées au hasard.

Notre hypothèse était que les enfants réagiraient de manière comparable dans les deux conditions et que les temps de regard vers la démonstratrice seraient plus importants dans la condition robot que dans la condition humain. A l'échelle de l'échantillon, nous n'avons pas

trouvé de différences concernant les résultats des enfants au cours des quatre sessions aussi avons-nous analysé les résultats des deux sessions « humain » rassemblées d'une part et des deux sessions « robot » rassemblées d'autre part pour la suite des analyses. Nous avons trouvé que les scores des enfants dans la condition robot étaient similaires à ceux de la condition humaine excepté pour la production des noms des parties du corps dans la tâche d'imitation et pour le respect de l'ordre dans la séquence imitée. Reconnaître les parties du corps montrées par la démonstratrice a été plus difficile sur le corps du robot NAO que sur le corps de l'adulte. Nous supposons que cela est dû aux capacités motrices limitées du robot et à la taille de la main du robot par rapport aux autres parties de son corps qui rend son geste de pointage imprécis. Cependant, ils ont aussi bien imité les parties du corps montrées par le robot que par l'adulte mais le respect de l'ordre est plus élevé dans la condition humaine. Ainsi, il apparait plus facile d'imiter correctement une séquence instruite par un démonstrateur humain que par un robot.

Concernant les tâches dans lesquelles il n'y avait pas de différence entre les conditions robot et humain, les pourcentages de bonnes réponses étaient égaux ou supérieurs à 50% pour toutes les tâches, sauf pour l'identification des postures émotionnelles. Une explication possible est que les enfants de ces âges sont encore en train d'apprendre à identifier les émotions et qu'ils ont besoin du contexte pour décoder la posture émotionnelle.

Enfin, le temps de réaction pour ramasser le bâton dans la tâche d'aide est comparable dans les deux conditions, ce qui montre que les enfants TSA peuvent répondre à un besoin instrumental qu'il soit manifesté par un robot ou par un adulte.

Dans les deux conditions, les enfants ont regardé la démonstratrice plus de la moitié du temps de la session. Nous avons trouvé que, sur toute la session et notamment au cours de la familiarisation avec la comptine, l'imitation et l'identification des postures émotionnelles, les enfants regardaient plus la démonstratrice robotique que la démonstratrice humaine. Ceci est cohérent avec deux études de cas réalisées chez des personnes TSA. Dans la première étude

(Tapus et al., 2012), certains enfants TSA regardaient plus vers le démonstrateur robotique (NAO) que vers le démonstrateur humain. Dans la seconde étude, Damm et collègues (2013) ont montré, en utilisant le suivi du regard (eye-tracker), que des adultes TSA regardaient significativement plus le visage du robot et maintenaient plus longtemps le regard avec le robot qu'avec le démonstrateur humain.

En comparaison avec le protocole expérimental de l'étude auprès des enfants typiques, la tâche d'apprentissage de vocabulaire a été supprimée sur conseil de l'équipe éducative du SESSAD (Service d'Education Spéciale et de Soins à Domicile). Les enfants TSA ont, en effet, eu des difficultés dans le domaine de la production langagière. Acquérir du vocabulaire nouveau et usuel leur permettant de communiquer dans leur vie quotidienne, nécessite souvent un entraînement répété sur plusieurs semaines voire plusieurs mois. Ainsi, leur proposer d'apprendre des mots inexistants comme « dax » et « grid » n'aurait pas été utile pour leur vie quotidienne. Nous avons pensé à personnaliser cette tâche en remplaçant les mots et les images d'animaux imaginaires par des noms d'objets ou d'animaux réels encore non acquis ou inconnus pour les enfants, en fonction des observations des parents et de l'équipe du SESSAD. Cependant, ceci aurait nécessité une nouvelle phase de programmation du robot, ce qui n'était pas envisageable dans le cadre de notre étude. Par ailleurs, nous avons également retiré les questionnaires pré et post-test sur conseil de l'équipe éducative pour réduire la durée des sessions et maximiser les chances que les enfants restent engagés durant toute l'expérience. Par ailleurs, demander aux enfants de répondre verbalement aux questions aurait pu les mettre en échec.

Enfin, pour les expériences menées au SESSAD, l'expérimentatrice n'était pas un adulte inconnu, comme pour les expériences en école, mais l'adulte de référence de l'enfant au sein de la structure (la psychomotricienne ou l'éducatrice) qui avait été préalablement formée à la passation du protocole expérimental.

Les résultats de cette étude nous ont permis de valider notre protocole expérimental auprès des enfants de 4 à 8 ans avec un TSA léger. Nous souhaiterions poursuivre cette étude en incluant plus d'enfants ayant des profils variés pour analyser plus en détails les intérêts et limites du robot NAO par rapport à l'adulte selon les enfants.

1.3. Utilisation d'un robot comme tuteur de chant pour de jeunes mâles de Diamant Mandarin

Dans le cadre de cette thèse visant à étudier l'utilisation de robots avec de jeunes enfants et oiseaux, nous avons développé un robot-diamant mandarin (MANDABOT) en partenariat avec les équipes de l'Institut des Systèmes Intelligents et de Robotique de Sorbonne Université et du Laboratoire de Psychologie de la Perception de l'Université Paris Descartes. MANDABOT a l'apparence d'un diamant mandarin mâle adulte (taille et couleurs du plumage et du bec), il est fixé sur un axe rotatif lui permettant de bouger son corps et il peut également effectuer des mouvements de tête. Il peut diffuser des sons, en l'occurrence des cris et du chant de diamant mandarin grâce à un haut-parleur situé dans son abdomen en résine. MANDABOT est conçu pour exprimer des comportements contingents à ceux de l'oiseau avec lequel il interagit. De part et d'autre du corps du robot, deux perches équipées de capteurs lui permettent de détecter la présence d'un oiseau et de se tourner dans sa direction. Il répond par un cri au cri poussé par l'oiseau grâce à l'intégration d'une boucle d'interaction vocale. Ainsi, ce robot permet d'étudier la contribution relative de différents facteurs (contingence, apparence, multimodalité sensorielle) sur son acceptation et son utilisation pour l'apprentissage du chant en laboratoire (Araguas et al., en révision).

Comme nous l'avons exposé précédemment, le but originel de notre étude était de comparer les résultats d'apprentissage du chant chez trois groupes d'oiseaux exposés individuellement à un tuteur : 1) un mâle adulte ; 2) un robot interactif présentant des comportements contingents aux comportements exprimés par le jeune oiseau ; 3) un robot non

interactif programmé à produire les séquences enregistrées dans la situation interactive (et de ce fait, non contingentes aux comportements exprimés par le jeune oiseau élevé avec le robot non interactif).

Pour rappel, chez le Diamant Mandarin, on peut obtenir une imitation très fidèle du chant lorsqu'un jeune oiseau est élevé seul avec un adulte (Derégnaucourt, 2011). Jusqu'à présent, des résultats hétérogènes ont été obtenus avec les autres méthodes utilisées (playback passif ou actif du chant, exposition à des vidéos) : on observe soit, une imitation pauvre du modèle diffusé (exemple : playback passif) soit, une forte variabilité inter-individuelle : certains apprennent bien et d'autres moins (Derégnaucourt, 2010, Derégnaucourt et al., 2013).

Nous avons fait l'hypothèse que les jeunes exposés individuellement au MANDABOT imiteraient aussi bien les chants que ceux du groupe d'oiseaux élevés individuellement avec un mâle adulte. De plus, nous nous attendions à observer de moins bons résultats d'apprentissage dans le groupe d'oiseaux exposés individuellement à un robot non interactif. Comme indiqué précédemment, le suivi de ce groupe n'a pas été possible en raison de problèmes technologiques (fonctionnement de la boucle ouverte du robot permettant la diffusion de séquences enregistrées) rencontrés pendant la crise sanitaire. En effet, l'équipe de l'ISIR n'a pas pu intervenir pour régler ce problème à temps.

Chaque oiseau a été exposé pendant la phase sensible d'apprentissage du chant (à partir du 35ème jour post éclosion) au tuteur de chant, 5 jours par semaine, pendant 5 semaines. Chaque session d'exposition a été filmée à l'aide de caméras afin d'étudier les comportements des jeunes et des tuteurs et nous avons enregistré en continu l'activité vocale des jeunes mandarins et des tuteurs. A la fin de l'expérience (100ème jour post-éclosion), nous avons calculé des scores de similarité acoustique entre les chants des jeunes et de leurs tuteurs. Notre résultat principal est que les mandarins ont aussi bien imité le chant produit par le tuteur oiseau que par le tuteur robotique

Nous avons également étudié l'impact des comportements du jeune sur son apprentissage du chant. Concernant le comportement vocal du tuteur, plus le tuteur produit de chants, plus le jeune produit une bonne imitation. Ceci est valable jusqu'à un certain seuil à partir duquel la copie devient moins fidèle. Cet effet est plus important dans le groupe robot, et peut en partie être expliqué par les résultats d'oiseaux peu actifs en présence du robot et qui ont de ce fait, été exposés à un moins grand nombre de chants diffusés.

Nous avons aussi analysé la distance physique entre le jeune et son tuteur dans les deux groupes. Pour le groupe robot uniquement, nous sommes allés plus en détails en étudiant le temps passé près du robot et en contact physique (clumping), le nombre de coups de bec du jeune sur le robot, l'attention du jeune au chant du robot. Nous avons trouvé que plus les jeunes passaient du temps proche du robot, plus leur imitation du chant était élevée. Le nombre de coups de bec sur la tête et le bec du robot présente une corrélation en U inversé à l'imitation du chant. Enfin, nous n'avons pas trouvé d'effet du nombre de réactions attentives au chant sur l'apprentissage. Ainsi, notre étude tend à montrer qu'un robot interactif peut être considéré comme un tuteur de chant acceptable pour de jeunes oiseaux.

L'importance de la contingence comportementale sur l'apprentissage du chant a été montrée récemment chez le Diamant Mandarin : de jeunes mâles exposés à des vidéos de femelles faisant des postures particulières pendant qu'ils chantent, produisent une imitation paternelle du chant meilleure que leurs frères exposés à ces mêmes vidéos mais de manière non contingente à leur production vocale (Carouso-Peck & Goldstein, 2019). Néanmoins, des études complémentaires sont nécessaires pour valider ce résultat. Tout d'abord, il faudrait tester des oiseaux en présence d'un robot non interactif comme indiqué ci-dessus. Comme les problèmes techniques rencontrés pendant la pandémie sont à présent résolus, cette expérience sera réalisée au printemps 2022. Cela permettra également de mettre en perspective les résultats obtenus au cours d'une précédente étude-pilote dans laquelle de jeunes diamants mandarins ont

été exposés à des modèles de chant diffusé soit 1) par un robot mobile produisant une séquence de mouvements prédéfinis par un programme (boucle ouverte), soit 2) par un robot immobile. Les auteurs de cette étude n'ont pas observé de différence significative entre les groupes et une imitation plutôt faible du modèle de chant diffusé (Derégnaucourt, communication personnelle).

Il faudrait aussi pouvoir tester le MANDABOT avec un plus grand nombre d'oiseaux. Malgré des scores de similarité élevés, nous avons observé une variabilité inter-individuelle comme cela a déjà été décrit au cours de précédentes études (Derégnaucourt et al., 2005, Derégnaucourt et al., 2013, Phan et al. 2006). Cette variabilité inter-individuelle peut présenter des limites. Le Diamant Mandarin est couramment utilisé pour étudier l'apprentissage vocal, et certaines procédures expérimentales de neuro-éthologie sont complexes, coûteuses et présentent surtout des aspects éthiques (exemple : pinéalectomie, transgenèse : Derégnaucourt et al., 2012 ; Haesler et al., 2005) impliquant de réduire le nombre d'oiseaux testés. Ainsi, la procédure la plus couramment utilisée est celle consistant à élever un jeune oiseau avec un tuteur adulte car elle garantit presque toujours une imitation quasi-parfaite du chant. L'utilisation de ce paradigme permet ainsi de vérifier l'impact d'une manipulation expérimentale (ex : lésion, transfection) sur le succès d'imitation en utilisant une taille d'échantillon réduite. Mais la variabilité inter-individuelle présente également des aspects intéressants, en particulier dans le contexte de l'évolution culturelle du chant. Au cours d'une étude menée dans notre laboratoire, il a ainsi été montré que cette variabilité inter-individuelle permettait de maintenir une certaine hétérogénéité vocale au sein d'une colonie de diamants mandarins fondée à partir d'individus entrainés expérimentalement à produire le même chant (Le Maguer, 2019 ; Le Maguer et al., en révision). Cette variabilité inter-individuelle pourrait être liée à des profils comportementaux individuels ; la personnalité chez les animaux en

général et les oiseaux en particulier, fait l'objet de nombreuses études, en particulier dans le domaine de la biologie évolutive.

2. Perspectives pour de futures études

2.1. Chez les enfants DT et présentant un TSA

Les remarques suivantes concernent à la fois les futurs projets de recherche menés auprès des enfants DT et des enfants TSA. Nous souhaitons en effet développer un protocole strictement identique afin de pouvoir comparer les résultats obtenus auprès de ces deux populations.

Le premier point concerne le comportement des démonstratrices robotique et humaine. L'interactivité a été volontairement minimisée pour l'expérience, ce qui a pu dérouter certains enfants. Nous voulions limiter le nombre de facteurs qui pouvaient expliquer des différences de comportements et d'apprentissage chez les enfants. Nous étions également contraints d'un point de vue technologique par le fait que le robot NAO était télé-opéré par l'expérimentatrice, via la tablette tactile, et entièrement programmé pour énoncer une fois chaque consigne. Son comportement était toujours strictement identique quelles que soient les réactions des enfants, et la démonstratrice humaine a, le plus possible, calqué son comportement sur celui du robot. Dans nos expériences, c'est l'expérimentatrice qui jouait le rôle de médiatrice entre l'enfant et la démonstratrice : elle pouvait répéter la consigne si besoin ou intervenir si l'enfant s'impatientait ou avait un comportement inadapté. Il s'agissait donc plus d'une simulation d'interaction avec l'enfant que d'une interaction réelle. Bien que la tendance soit à l'augmentation de l'autonomie et de l'intelligence artificielle des robots sociaux, il est, à l'heure actuelle, très compliqué de laisser les enfants interagir seuls avec les robots sans surveillance, ce qui correspondrait à la situation expérimentale idéale pour comparer l'influence de la

démonstratrice robotique à celui de la démonstratrice humaine. C'est d'ailleurs ce paradigme qui a été utilisé sans réserve dans notre étude avec les oiseaux.

Une amélioration possible pour limiter les interventions de l'expérimentatrice et rendre le comportement de la démonstratrice plus flexible dans notre protocole, serait que la démonstratrice puisse proposer un étayage standardisé et comportant plusieurs niveaux d'aide pour chaque tâche. Pour la tâche de reconnaissance des postures émotionnelles, le premier niveau de consigne donné par la démonstratrice pourrait être « Quand je suis dans cette position, quelle émotion est-ce que j'exprime ? » alors que le second niveau serait une énumération « Est-ce que je suis plutôt joyeuse, triste, en colère ou est-ce que j'ai peur ? ».

La majorité des tâches proposées correspondait à une évaluation des connaissances, des compétences et des représentations des enfants à un moment donné. Seule la tâche de vocabulaire dans le protocole des enfants typiques correspondait à une situation d'apprentissage immédiat. Or, le domaine de l'apprentissage à partir des robots sociaux semble prometteur et pourrait offrir de nouveaux outils pédagogiques aux enseignants, notamment pour les enfants les plus en difficultés. Utiliser les robots à l'école nécessite de déterminer les situations dans lesquelles ils peuvent servir de supports d'apprentissage complémentaires aux autres outils présents en classe. Afin de participer au développement de ce champ disciplinaire, nous pourrions transformer l'ensemble de notre protocole en situation d'apprentissage personnalisé. L'expérience, telle que présentée actuellement correspondrait à l'évaluation initiale des connaissances de l'enfant. Ensuite, selon ses résultats, nous pourrions lui proposer plusieurs sessions d'apprentissage ciblées sur les parties du corps ou les postures émotionnelles les moins reconnues par exemple. Pour déterminer si l'enfant s'améliore et apprend de nouvelles connaissances au fur et à mesure des sessions, il serait nécessaire que la démonstratrice le corrige lorsqu'il se trompe et lui donne la bonne réponse et qu'elle le félicite lorsque sa réponse est juste.

A ce sujet, une étude récente de Haas et ses collègues (2020) a mis en évidence que l'engagement des enfants de 5 et 6 ans dans une tâche d'apprentissage de mots nouveaux variait en fonction du feedback reçu par le robot NAO. Plus précisément, ils ont dans un premier temps réalisé une enquête auprès d'enseignants pour déterminer de quelle manière ces derniers estimaient que le robot devait exprimer son feedback à l'enfant dans une tâche d'apprentissage. Le feedback le plus adapté selon les enseignants combinait une phrase explicite variable comme « c'est bien, c'est un oiseau ! », « bravo, c'est un oiseau ! » ou « malheureusement, c'est faux, essaye encore », « c'est incorrect, essaye encore », avec une répétition du mot à apprendre et la possibilité pour l'enfant d'avoir un essai supplémentaire lorsque sa réponse était incorrecte. Ils ont comparé l'engagement des enfants dans la tâche d'apprentissage selon que le feedback était adapté, imprécis (le robot disait alors simplement « c'est bien » ou « c'est faux ») ou que le robot ne donnait pas de feedback. Ils ont trouvé que les enfants se montraient plus engagés dans la tâche lorsque le robot leur avaient fait un retour adapté.

Par ailleurs, Chen et collègues (2020) suggèrent de rendre l'enfant actif dans son apprentissage en combinant des scénarii de type « apprendre en enseignant » dans lequel l'enfant apprend des connaissances au robot avec une approche plus classique dans laquelle le robot joue le rôle de tuteur. Ils ont montré que cette alternance de rôles chez le robot Tega et les enfants de 5 à 7 ans conduisait à de meilleurs résultats concernant l'apprentissage de vocabulaire et à un engagement affectif (mesuré par les expressions faciales des enfants) plus important que dans les conditions où le robot avait, soit le rôle de tuteur, soit le rôle d'apprenant.

Enfin, malgré les temps de présentation organisés en amont de la venue en école, il est arrivé à plusieurs reprises que des parents expriment leur désaccord au sujet de la mise en place du projet en disant avoir peur. Ces inquiétudes ont également été exprimées par toutes les équipes éducatives rencontrées. Nous pensons qu'il est indispensable de réaliser ce type de projet en étroite collaboration avec l'Education Nationale et les institutions médico-sociales

afin que les résultats apportés par la recherche puissent être utilisés par les professionnels de l'éducation. Tolksdorf et collègues (2021) soulignent que la plupart des études actuelles adoptent une microperspective focalisée sur les compétences cognitives, les comportements et émotions suscitées par l'interaction avec le robot. Ils proposent de développer plutôt une macroperspective visant à étudier l'impact de l'utilisation des robots à un niveau plus large, comme son utilisation en institution. Ces auteurs insistent également sur le fait que les projets utilisant les robots sociaux avec des enfants de moins de 6 ans ne devraient pas être menés sans intervention d'experts, et être portés uniquement par les enseignants ou les éducateurs. Ces derniers ont en effet, en général des compétences limitées concernant le fonctionnement du robot et sa réparation s'il vient à arrêter de fonctionner.

Pour conclure, notre étude est, à notre connaissance, la première étude développementale qui compare les réactions des enfants à la fois face au robot NAO ou face à un adulte non familier, et dans des tâches variées allant de l'évaluation des représentations des enfants à l'expression d'un comportement prosocial. Nos résultats sont cohérents avec la littérature existante, indiquant que le robot NAO peut être utilisé en situation d'interaction directe avec les enfants de moins de 8 ans, ce qui nous permet de valider le protocole expérimental. Nous avons également pu déterminer les améliorations méthodologiques à réaliser pour de prochaines études.

2.2. Chez le Diamant Mandarin

Chez le Diamant Mandarin, les interactions physiques et vocales avec un tuteur semblent indispensables pour obtenir une bonne imitation du chant (Derégnaucourt, 2010). Nos résultats semblent montrer qu'un robot social contingent présente les caractéristiques suffisantes pour être adopté comme tuteur de chant par de jeunes diamants mandarins. De plus, Chen et collègues (2016) ont mis en évidence l'existence de chants adressés aux jeunes

diamants mandarins. Ils ont montré que les tuteurs mâles adultes modifiaient la structure de leurs chants pour en faciliter l'apprentissage par les jeunes mandarins (en ajoutant, par exemple, des syllabes introductives et en augmentant l'intervalle entre les motifs de chant).

Nous utiliserons nos données expérimentales pour comparer les chants produits par les tuteurs lorsqu'ils étaient seuls dans leur cage, à ceux produits en présence d'une femelle, et à ceux produits en présence des jeunes diamants mandarins.

Afin d'aller plus en profondeur dans la compréhension des mécanismes influençant l'apprentissage du chant, nous pourrions imaginer comparer l'apprentissage du chant des mandarins selon que l'exposition au chant du tuteur est réalisée en utilisant une vidéo du robot ou une vidéo d'un mâle adulte, ce qui, à notre connaissance, n'a encore jamais été testé chez aucune espèce d'animal non-humain. Nous pourrions aussi étudier l'importance de l'apparence physique du robot (recouvert de plumes, couleur) et d'autres caractéristiques (production thermique, odeurs…) sur son acceptation même si nos résultats avec un robot composé d'une coque en résine peinte aux couleurs d'un mâle de diamant mandarin sont déjà très prometteurs. Ce résultat n'est pas vraiment étonnant lorsque l'on parcourt la littérature très riche en éthologie concernant l'utilisation de leurres pour étudier le comportement des animaux (Tinbergen, 1951).

Le Diamant Mandarin vivant en grands groupes, nous pourrions également utiliser le robot en volière, face à des groupes d'oiseaux pour évaluer la préférence des oiseaux et l'apprentissage du chant en groupe. En effet, au cours de notre étude, les jeunes oiseaux ne pouvaient pas choisir leur tuteur alors que plusieurs études tendent à montrer qu'ils choisissent activement un tuteur en fonction de ses caractéristiques individuelles (exemple : l'agressivité, Clayton, 1987). Plusieurs robots-oiseaux, chantant chacun un chant différent pourraient, par exemple, être placés dans une même volière, dans un contexte plus proche des conditions naturelles d'apprentissage du chant. Nous pourrions également étudier l'évolution culturelle du

chant au sein d'une volière fondée avec des tuteurs robotiques. Des études ont montré que des robots étaient bien acceptés par des oiseaux en milieu naturel et déclenchaient des comportements de parade (robot Jardinier satiné (*Ptilonorhynchus violaceus)*, Patricelli et al., 2006 ; robot Tétras des Armoises (*Centrocercus urophasianus*), Patricelli & Krakauer, 2009).

En conditions naturelles, des robots-oiseaux pourraient être utilisés comme tuteurs de chant lorsque le nombre de tuteurs conspécifiques disponibles n'est plus assez important. Une étude récente a montré que la diminution des populations de Méliphage Régent (*Anthochaera phrygia*) était associée à une perte de la complexité acoustique des chants produits, liée à une diminution du nombre de modèles produits par des individus conspécifiques (Crates et al., 2021). A terme, les auteurs indiquent que cela pourrait conduire à une extinction des populations d'oiseaux déjà déclinantes : les femelles pouvant potentiellement ne plus être attirées par des mâles ne chantant plus les séquences spécifiques. Ainsi, des robots ayant l'apparence d'un méliphage régent diffusant du chant pourraient servir de tuteurs de chant. Par le contrôle des chants diffusés, il serait possible d'assurer la conservation d'une diversité vocale au sein de populations vulnérables.

De manière plus globale, le MANDABOT s'inscrit dans la lignée de robots sociaux animaloïdes et en ce sens, il pourrait également être utilisé dans l'interaction avec les humains. Le champ des interventions assistées par l'animal englobe toutes les situations qui utilisent les animaux dans le but d'améliorer la condition physique, mentale, et sociale des humains (Kamioka et al., 2014). Pour les humains, les bénéfices apportés par la thérapie assistée par l'animal semblent dépasser de loin les risques comprenant les allergies, les maladies ou les accidents (Bert et al., 2016 pour une revue). Cependant, les études mesurant l'impact des sessions sur les réponses physiologiques et comportementales des animaux non-humains sont encore rares et les résultats sont mitigés. Dans sa revue, Glenk (2017) souligne que les études montrent à la fois des effets apaisants et stimulants pour les chiens. Mendoça et ses collègues

(2019) ont montré que chez les chevaux, les sessions n'étaient vécues ni comme des évènements stressants ni comme des événements positifs. Enfin, chez le cochon d'Inde (Gut et al., 2018), les sessions provoquent un stress limité, diminué par la possibilité pour l'animal de se retirer de l'interaction à tout moment et peuvent même constituer un enrichissement pour le cochon d'Inde. Récemment, plusieurs équipes de recherche ont cherché à déterminer si les robots animaloïdes pouvaient remplacer les animaux dans ce contexte. Si l'impact sur le bien-être animal est incontestable, les bénéfices pour la santé humaine et la régulation émotionnelle restent encore à évaluer (Glenk & Foltin, 2021). Barber et ses collègues (2020) ont comparé les comportements d'enfants de 11 à 12 ans selon qu'ils interagissaient avec un chien de thérapie ou avec le robot MiRo-E, présenté dans cette étude, comme un robot-chien. Les enfants passaient autant de temps à caresser le chien et le robot mais de manière générale, ils interagissaient plus longuement avec le robot, ce dernier répondant plus aux comportements d'initiation de l'enfant. Interrogés sur leurs ressentis, les enfants ont dit avoir significativement préféré interagir avec le chien de thérapie qu'avec le robot même si l'interaction avec le robot était elle aussi associée à un ressenti positif. Enfin, citons le robot-dauphin Delle[1], particulièrement réaliste, capable de nager, sauter et pousser des cris. Il a été conçu dans le but de mettre fin à la captivité des mammifères marins dans les parcs d'attraction aux Etats-Unis tout en offrant aux visiteurs la possibilité de continuer à assister à des spectacles dont le premier est prévu pour l'été 2022.

[1] https://www.edgedolphinspirit.com/





A gauche, le robot MiRo-E[2], dont l'espèce est délibérément ambiguë : il a des oreilles de lapin, une tête de cerf avec de grands yeux et un corps ovale (Coghlan et al., 2018) et à droite, le robot-dauphin Delle[3], tous deux en interaction avec des humains.

Conclusion

Nous avons développé, dans ce projet de recherche, une approche pluridisciplinaire alliant les champs de l'éthologie, de la psychologie du développement et de la psychomotricité. Nos études ont été réalisées auprès de deux espèces hautement sociales, l'humain et le diamant mandarin. Nous avons comparé les interactions sociales entre deux individus non familiers de la même espèce à celles avec un robot non familier présentant certaines des caractéristiques physiques et comportementales de l'espèce étudiée. Nous espérons que nos résultats contribueront au développement de robots sociaux adaptés à leurs partenaires, avec différentes applications dans des champs aussi divers que la biologie de la conservation ou la pédagogie.

[2] https://www.miro-e.com/
[3] https://www.edgefx.com/real-time-animatronics

REFERENCES

Références

Adret, P. (1992). Vocal Learning Induced With Operant Techniques : An Overview. *Netherlands Journal of Zoology*, *43*, 125-142. https://doi.org/10.1163/156854293X00250

Adriaense, J. E. C., Martin, J. S., Schiestl, M., Lamm, C., & Bugnyar, T. (2019). Negative emotional contagion and cognitive bias in common ravens (Corvus corax). *Proceedings of the National Academy of Sciences of the United States of America*, *116*(23), 11547-11552. https://doi.org/10.1073/pnas.1817066116

Alaria., L. & Gentaz, P. (2018). Le développement du langage oral chez les jeunes enfants mono ou multilingues, repères théoriques et pratiques destinés aux professionnel-les de l'enfance à Genève.

Allen, J. (2019). Community through Culture : From Insects to Whales. *BioEssays*, *41*. https://doi.org/10.1002/bies.201900060

American Psychiatric Association, DSM-5 Task Force. (2013). Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.). American Psychiatric Publishing, Inc..

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, ed. 5, 2013) https://doi.org/10.1176/appi.books.9780890425596

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell*, *157*(1), 187-200. https://doi.org/10.1016/j.cell.2014.03.003

Andreu-Perez, J., Deligianni, F., Ravì, D., & Yang, G.-Z. (2017). *Artificial Intelligence and Robotics*.

Ahtam, B., Braeutigam, S., & Bailey, A. (2020). Semantic Processing in Autism Spectrum Disorders Is Associated With the Timing of Language Acquisition : A Magnetoencephalographic Study. *Frontiers in Human Neuroscience*, *14*, 267. https://doi.org/10.3389/fnhum.2020.00267

Asimov, I. (1942). “Runaround” from I, Robot. *Street and Smith Publications.*

Araguas, A., Guellaï, B., Gauthier, P., Richer, F., Montone, G., & Derégnaucourt, S. (2021). Design of a robotic zebra finch for experimental studies on developmental song learning. *Journal of Experimental Biology* (submitted)

Baraka, K., Alves-Oliveira, P., & Ribeiro, T. (2020). *An Extended Framework for Characterizing Social Robots* (p. 21-64). https://doi.org/10.1007/978-3-030-42307-0_2

Barber, O., Somogyi, E., McBride, A. E., & Proops, L. (2021). Children's Evaluations of a Therapy Dog and Biomimetic Robot : Influences of Animistic Beliefs and Social Interaction. *International Journal of Social Robotics*, *13*(6), 1411-1425. https://doi.org/10.1007/s12369-020-00722-0

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2008). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, *1*, 71-81. https://doi.org/10.1007/s12369-008-0001-3

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of Emotional Body Language Displayed by a Humanoid Robot : A Case Study with Children. *International Journal of Social Robotics*, *5*(3), 325-334. https://doi.org/10.1007/s12369-013-0193-z

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, *3*(2), 74-99. https://doi.org/10.5898/JHRI.3.2.Beer

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education : A review. *Science Robotics*, *3*(21). https://doi.org/10.1126/scirobotics.aat5954

Benge, C., Onwuegbuzie, A., & Robbins, M. (2012). A Model for Presenting Threats to Legitimation at the Planning and Interpretation Phases in the Quantitative, Qualitative, and Mixed Research Components of a Dissertation. *International Journal of Education*, *4*. https://doi.org/10.5296/ije.v4i4.2360

Benskin, C. M. H., Mann, N. I., Lachlan, R. F., & Slater, P. J. B. (2002). Social learning directs feeding preferences in the zebra finch, *Taeniopygia guttata*. *Animal Behaviour*, *64*(5), 823–828. https://doi.org/10.1006/anbe.2002.2005

Bergès, J., & Lézine, I. (1963). *Test of imitation of gestures: Methods for plotting body image and movements of children from 3 to 6 years of age*. Masson & Cie.

Bert, F., Gualano, M. R., Camussi, E., Pieve, G., Voglino, G., & Siliquini, R. (2016). Animal assisted intervention : A systematic review of benefits and risks. *European Journal of Integrative Medicine*, *8*(5), 695-706. https://doi.org/10.1016/j.eujim.2016.05.005

Bilodeau, R. A., & Kramer, R. K. (2017). Self-Healing and Damage Resilience for Soft Robotics : A Review. *Frontiers in Robotics and AI*, *4*, 48. https://doi.org/10.3389/frobt.2017.00048

Birch, J., Schnell, A. K., & Clayton, N. S. (2020). Dimensions of Animal Consciousness. *Trends in Cognitive Sciences*, *24*(10), 789-801. https://doi.org/10.1016/j.tics.2020.07.007

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by Teaching : A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence*, *19*(3-4), 363-392. https://doi.org/10.1080/08839510590910200

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, *11*(11), 747–759. https://doi.org/10.1038/nrn2931

Bonnet, F., Cazenille, L., Gribovskiy, A., Halloy, J., & Mondada, F. (2017b). Multi-robot control and tracking framework for bio-hybrid systems with closed-loop interaction. *2017 IEEE International Conference on Robotics and Automation (ICRA)*. https://doi.org/10.1109/ICRA.2017.7989515

Bonnet, F., Cazenille, L., Séguret, A., Gribovskiy, A., Collignon, B., Halloy, J., & Mondada, F. (2017a). Design of a modular robotic system that mimics small fish locomotion and body movements for ethological studies. *International Journal of Advanced Robotic Systems*, *14*(3), 1729881417706628. https://doi.org/10.1177/1729881417706628

Bonnet, F., Crot, N., Burnier, D., & Mondada, F. (2016a). Design methods for miniature underwater soft robots. *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. https://doi.org/10.1109/BIOROB.2016.7523823

Bonnet, F., Kato, Y., Halloy, J., & Mondada, F. (2016b). Infiltrating the zebrafish swarm : Design, implementation and experimental tests of a miniature robotic fish lure for fish-robot interaction studies. *Artificial Life and Robotics*. https://doi.org/10.1007/s10015-016-0291-8

Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1392), 227-233.

Bouman, A., Ginting, M. F., Alatur, N., Palieri, M., Fan, D. D., Touma, T., Pailevanian, T., Kim, S.-K., Otsu, K., Burdick, J., & Agha-mohammadi, A. (2020). Autonomous Spot : Long-Range Autonomous Exploration of Extreme Environments with Legged Locomotion. *2020 IEEE/RSJ*

International Conference on Intelligent Robots and Systems (IROS). https://doi.org/10.1109/IROS45743.2020.9341361

Box, H. (1984). Primate Behavior and Social Ecology. Chapman & Hall, London.

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, *417*(6886), 351–358.

Briefer, E. F. (2018). Vocal contagion of emotions in non-human animals. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873), 20172783. https://doi.org/10.1098/rspb.2017.2783

Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., & Schwartz, D. (1999). *Teachable Agents : Combining Insights from Learning Theory and Computer Science*.

Broussard, M. (2018). *Artificial Unintelligence : How Computers Misunderstand the World*. MIT Press.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, *84*(2), 422-428. https://doi.org/10.1111/j.1467-8624.2012.01860.x

Camões-Costa, V., Erjavec, M., & Horne, P. J. (2011). The impact of body-part-naming training on the accuracy of imitative performances in 2- to 3-year-old children. *Journal of the Experimental Analysis of Behavior*, *96*(3), 291-315. https://doi.org/10.1901/jeab.2011.96-291

Carouso-Peck, S., & Goldstein, M. H. (2019). Female Social Feedback Reveals Non-imitative Mechanisms of Vocal Learning in Zebra Finches. *Current Biology: CB*, *29*(4), 631-636.e3. https://doi.org/10.1016/j.cub.2018.12.026

Carr, K., Kendal, R. L., & Flynn, E. G. (2015). Imitate or innovate? Children's innovation is influenced by the efficacy of observed behaviour. *Cognition*, *142*, 322-332. https://doi.org/10.1016/j.cognition.2015.05.005

Catchpole, C. K., & Slater, P. J. B. (2008). Bird Song: Biological Themes and Variations. *Cambridge University, Cambridge*.

Chatila, R., Renaudo, E., Andries, M., Chavez-Garcia, R.-O., Luce-Vayrac, P., Gottstein, R., Alami, R., Clodic, A., Devin, S., Girard, B., & Khamassi, M. (2018). Toward Self-Aware Robots. *Frontiers in Robotics and AI*, *5*, 88. https://doi.org/10.3389/frobt.2018.00088

Chen, Z., Jia, X., Riedel, A., & Zhang, M. (2014). A bio-inspired swimming robot. *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2564-2564. https://doi.org/10.1109/ICRA.2014.6907221

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), 6641-6646. https://doi.org/10.1073/pnas.1522306113

Chen, H., Park, H. W., & Breazeal, C. (2020). Teaching and learning with children : Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education*, *150*, 103836. https://doi.org/10.1016/j.compedu.2020.103836

Chin, D. B., Dohmen, I. M., & Schwartz, D. L. (2013). Young Children Can Learn Scientific Reasoning with Teachable Agents. *IEEE Transactions on Learning Technologies*, *6*(3), 248-257. https://doi.org/10.1109/TLT.2013.24

Church, R. M. (1959). Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology*, *52*(2), 132-134. https://doi.org/10.1037/h0043531

Clark, R. (1983). *Reconsidering Research on Learning from Media*. https://doi.org/10.3102/00346543053004445

Clayton, N. S. (1987). Song tutor choice in zebra finches. *Animal Behaviour*, *35*(3), 714-721. https://doi.org/10.1016/S0003-3472(87)80107-0

Coelho, C. G., Falótico, T., Izar, P., Mannu, M., Resende, B. D., Siqueira, J. O., & Ottoni, E. B. (2015). Social learning strategies for nut-cracking by tufted capuchin monkeys (Sapajus spp.). *Animal Cognition*, *18*(4), 911-919. https://doi.org/10.1007/s10071-015-0861-5

Cohen, I., Looije, R., & Neerincx, M. A. (2014). Child's Perception of Robot's Emotions : Effects of Platform, Context and Experience. *International Journal of Social Robotics*, *6*(4), 507-518. https://doi.org/10.1007/s12369-014-0230-6

Cominelli, L., Mazzei, D., & De Rossi, D. E. (2018). SEAI : Social Emotional Artificial Intelligence Based on Damasio's Theory of Mind. *Frontiers in Robotics and AI*, *5*, 6. https://doi.org/10.3389/frobt.2018.00006

Connolly, J., Mocz, V., Salomons, N., Valdez, J., Tsoi, N., Scassellati, B., & Vázquez, M. (2020). Prompting Prosocial Human Interventions in Response to Robot Mistreatment. *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 211-220.

Copeland, B. J. (2000). The Turing Test*. *Minds and Machines*, *10*(4), 519-539. https://doi.org/10.1023/A:1011285919106

Coslett, H. B., Saffran, E. M., & Schwoebel, J. (2002). Knowledge of the human body : A distinct semantic domain. *Neurology*, *59*(3), 357-363. https://doi.org/10.1212/wnl.59.3.357

Coulson, M. (2004). Attributing Emotion to Static Body Postures : Recognition Accuracy, Confusions, and Viewpoint Dependence. *Journal of Nonverbal Behavior*, *28*(2), 117-139. https://doi.org/10.1023/B:JONB.0000023655.25550.be

Crates, R., Langmore, N., Ranjard, L., Stojanovic, D., Rayner, L., Ingwersen, D., & Heinsohn, R. (2021). Loss of vocal culture and fitness costs in a critically endangered songbird. *Proceedings. Biological Sciences*, *288*(1947), 20210225. https://doi.org/10.1098/rspb.2021.0225

Cueto, S., Prieto, J. A., Nistal, P., Abelairas-Gómez, C., Barcala-Furelos, R., & López, S. (2017). Teachers' Perceptions of Preschool Children's Psychomotor Development in Spain. *Perceptual and Motor Skills*, *124*(4), 725-739. https://doi.org/10.1177/0031512517705534

Cully, A., Clune, J., Tarapore, D., & Mouret, J.-B. (2015). Robots that can adapt like animals. *Nature*, *521*(7553), 503-507. https://doi.org/10.1038/nature14422

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion (Washington, D.C.)*, *12*(5), 1085-1101. https://doi.org/10.1037/a0025737

Damm, O., Malchus, K., Jaecks, P., Krach, S., Paulus, F., Naber, M., Jansen, A., Kamp-Becker, I., Einhaeuser-Treyer, W., Stenneken, P., & Wrede, B. (2013). Different gaze behavior in human-robot interaction in Asperger's syndrome : An eye-tracking study. *2013 IEEE RO-MAN*, 368-369. https://doi.org/10.1109/ROMAN.2013.6628501

Danchin, E., Nöbel, S., Pocheville, A., Dagaeff, A.-C., Demay, L., Alphand, M., Ranty-Roby, S., van Renssen, L., Monier, M., Gazagne, E., Allain, M., & Isabel, G. (2018). Cultural flies : Conformist social learning in fruitflies predicts long-lasting mate-choice traditions. *Science (New York, N.Y.)*, *362*(6418), 1025-1030. https://doi.org/10.1126/science.aat1590

Dang, Q., Nguyen, T., Jeong, S., & Chong, N. (2017). *Encoding Cultures in Robot Emotion Representation*. https://doi.org/10.1109/ROMAN.2017.8172356

Darwin, C. (1872). The expression of the emotions in man and animals (3rd ed.). London, UK: HarperCollins.

Decety, J., Bartal, I., Uzefovsky, F., & Knafo-Noam, A. (2016). Empathy as a driver of prosocial behaviour : Highly conserved neurobehavioural mechanisms across species. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2015.0077

de Haas, M., Vogt, P., & Krahmer, E. (2020). The Effects of Feedback on Children's Engagement and Learning Outcomes in Robot-Assisted Second Language Learning. *Frontiers in Robotics and AI*, *7*, 101. https://doi.org/10.3389/frobt.2020.00101

de Margerie, E., Lumineau, S., Houdelier, C., & Yris, M.-A. R. (2011). Influence of a mobile robot on the spatial behaviour of quail chicks. *Bioinspiration & Biomimetics*, *6*(3), 034001. https://doi.org/10.1088/1748-3182/6/3/034001

de Margerie, E., Peris, A., Pittet, F., Houdelier, C., Lumineau, S., & Richard-Yris, M.-A. (2013). Effect of mothering on the spatial exploratory behavior of quail chicks. *Developmental Psychobiology*, *55*(3), 256-264. https://doi.org/10.1002/dev.21019

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach–Major, S., & Queenan, P. (2003). Preschool Emotional Competence : Pathway to Social Competence? *Child Development*, *74*(1), 238-256. https://doi.org/10.1111/1467-8624.00533

Derégnaucourt, S. (2010). Interspecific Hybridization as a Tool to Understand Vocal Divergence : The Example of Crowing in Quail (Genus Coturnix). *PloS one*, *5*, e9451. https://doi.org/10.1371/journal.pone.0009451

Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the Zebra Finch (*Taeniopygia guttata*). *Interaction Studies*, *12*(2), 324–350. https://doi.org/10.1075/is.12.2.07der

Derégnaucourt, S., Araguas, A. & Guellaï, B. (2021). Technological advances for getting insight into the learning capacities of birds in the vocal domain. *Interaction Studies*. (in preparation)

Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, *433*(7027), 710-716. https://doi.org/10.1038/nature03275

Derégnaucourt, S., Poirier, C., Van der Kant, A., Van der Linden, A., & Gahr, M. (2013). Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *Journal of Physiology-Paris*, *107*(3), 210–218.

Derégnaucourt, S., Saar, S., & Gahr, M. (2012). Melatonin affects the temporal pattern of vocal signatures in birds. *Journal of pineal research*, 53(3), 245-258.

Desmedt, L., George, I., Mohamed Benkada, A., Hervé, M., Aubin, T., Derégnaucourt, S., & Lumineau, S. (2020). Maternal presence influences vocal development in the Japanese quail (Coturnix c. Japonica). *Ethology*, *126*(5), 553-562. https://doi.org/10.1111/eth.13007

de Waal, F. B. M. (2008). Putting the altruism back into altruism : The evolution of empathy. *Annual Review of Psychology*, *59*, 279-300. https://doi.org/10.1146/annurev.psych.59.103006.093625

de Waal, F. B. M. (2011). What is an animal emotion? *Annals of the New York Academy of Sciences*, *1224*, 191-206. https://doi.org/10.1111/j.1749-6632.2010.05912.x

de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy : Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498-509. https://doi.org/10.1038/nrn.2017.72

Dou, X., Wu, C.-F., Lin, K.-C., & Tseng, T.-M. (2019). The Effects of Robot Voice and Gesture Types on the Perceived Robot Personalities. In M. Kurosu (Éd.), *Human-Computer Interaction. Perspectives on Design* (p. 299-309). Springer International Publishing. https://doi.org/10.1007/978-3-030-22646-6_21

Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. A. (2017). *The Social Psychology of Prosocial Behavior*. Psychology Press. https://doi.org/10.4324/9781315085241

Driskell, J., Goodwin, G., Salas, E., & O'Shea, P. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice*, *10*, 249-271. https://doi.org/10.1037/1089-2699.10.4.249

Duffy, B. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*, 177-190. https://doi.org/10.1016/S0921-8890(02)00374-3

Dunfield, K. A., Best, L. J., Kelley, E. A., & Kuhlmeier, V. A. (2019). Motivating Moral Behavior : Helping, Sharing, and Comforting in Young Children With Autism Spectrum Disorder. *Frontiers in Psychology*, *10*, 25. https://doi.org/10.3389/fpsyg.2019.00025

Dunfield, K. A., & Kuhlmeier, V. A. (2013). Classifying prosocial behavior : Children's responses to instrumental need, emotional distress, and material desire. *Child Development*, *84*(5), 1766-1776. https://doi.org/10.1111/cdev.12075

Eisenberg, N. (2006). Prosocial Behavior. In *Children's needs III: Development, prevention, and intervention* (p. 313-324). National Association of School Psychologists.

Ekman, P. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science*, *11*, 31-34. https://doi.org/10.1177/1745691615596992

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56-75. https://doi.org/10.1007/BF01115465

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., Le Compte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712-717. https://doi.org/10.1037/0022-3514.53.4.712

Erden, M. S. (2013). Emotional Postures for the Humanoid-Robot Nao. *International Journal of Social Robotics*, *5*(4), 441-456. https://doi.org/10.1007/s12369-013-0200-4

Faraj, Z., Selamet, M., Morales, C., Torres, P., Hossain, M., Chen, B., & Lipson, H. (2021). Facially expressive humanoid robotic face. *HardwareX*, *9*, e00117. https://doi.org/10.1016/j.ohx.2020.e00117

Fernández-Juricic, E., Gilak, N., Mcdonald, J., Pithia, P., & Valcarcel, A. (2006). A dynamic method to study the transmission of social foraging information in flocks using robots. *Animal Behaviour*, *71*, 901-911. https://doi.org/10.1016/j.anbehav.2005.09.008

Ferretti, V., & Papaleo, F. (2019). Understanding others : Emotion recognition in humans and other animals. *Genes, Brain and Behavior*, *18*(1), e12544. https://doi.org/10.1111/gbb.12544

Fiorito, G., & Scotto, P. (1992). Observational Learning in Octopus vulgaris. *Science (New York, N.Y.)*, *256*(5056), 545-547. https://doi.org/10.1126/science.256.5056.545

Fischer, A., Manstead, A., Lewis, I., Haviland-Jones, J., & Barrett, L. (2016). *Social Functions of Emotion and Emotion Regulation*.

Floreano, D., & Wood, R. (2015). Science, technology and the future of small autonomous drones. *Nature*. https://doi.org/10.1038/nature14542

Floridi, L. (2019). What the Near Future of Artificial Intelligence Could Be. *Philosophy & Technology*, *32*(1), 1-15. https://doi.org/10.1007/s13347-019-00345-y

Flynn, E., Turner, C., & Giraldeau, L.-A. (2016). Selectivity in social and asocial learning : Investigating the prevalence, effect and development of young children's learning preferences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*, 20150189. https://doi.org/10.1098/rstb.2015.0189

Flynn, E., & Whiten, A. (2012). Experimental « microcultures » in young children : Identifying biographic, cognitive, and social predictors of information transmission. *Child Development*, *83*(3), 911-925. https://doi.org/10.1111/j.1467-8624.2012.01747.x

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3), 143-166. https://doi.org/10.1016/S0921-8890(02)00372-X

Francis, P., Eastwood, K. W., Bodani, V., Looi, T., & Drake, J. M. (2018). Design, Modelling and Teleoperation of a 2 mm Diameter Compliant Instrument for the da Vinci Platform. *Annals of Biomedical Engineering*, *46*(10), 1437-1449. https://doi.org/10.1007/s10439-018-2036-4

Frijda, N. H. (1986). *The emotions* (p. xii, 544). Editions de la Maison des Sciences de l'Homme.

Galef, B. (1988). Imitations in animals : History, definitions, and interpretations of data from the psychological laboratory. In *Social Learning : Psychological and Biological Perspectives* (p. 3-28).

Geva, N., Uzefovsky, F., & Levy-Tzedek, S. (2020). Touching the social robot PARO reduces pain perception and salivary oxytocin levels. *Scientific Reports*, *10*(1), 9814. https://doi.org/10.1038/s41598-020-66982-y

Glenk, L. M. (2017). Current Perspectives on Therapy Dog Welfare in Animal-Assisted Interventions. *Animals*, *7*(2), 7. https://doi.org/10.3390/ani7020007

Glenk, L. M., & Foltin, S. (2021). Therapy Dog Welfare Revisited : A Review of the Literature. *Veterinary Sciences*, *8*(10), 226. https://doi.org/10.3390/vetsci8100226

Gosling, S. D. (2008). Personality in Non-human Animals. *Social and Personality Psychology Compass*, *2*(2), 985-1001. https://doi.org/10.1111/j.1751-9004.2008.00087.x

Grandgeorge, M. (2020). Evaluating Human-Robot Interaction with Ethology. *Human-Robot Interaction* (p. 257-268). https://doi.org/10.1007/978-3-030-42307-0_10

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping : A person × situation perspective. *Journal of Personality and Social Psychology*, *93*(4), 583-599. https://doi.org/10.1037/0022-3514.93.4.583

Graziano, W. G., & Tobin, R. M. (2009). Agreeableness. In *Handbook of individual differences in social behavior* (p. 46-61). The Guilford Press.

Gribovskiy, A., Halloy, J., Deneubourg, J., Bleuler, H., & Mondada, F. (2010). Towards mixed societies of chickens and robots. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. https://doi.org/10.1109/IROS.2010.5649542

Gribovskiy, A., Mondada, F., Deneubourg, J.-L., Cazenille, L., Bredeche, N., & Halloy, J. (2015). *Automated Analysis of Behavioural Variability and Filial Imprinting of Chicks (G. gallus), using Autonomous Robots*.

Gribovskiy, A., Halloy, J., Deneubourg, J. L., & Mondada, F. (2018). Designing a socially integrated mobile robot for ethological research. *Robotics and Autonomous Systems*, *103*, 42-55. https://doi.org/10.1016/j.robot.2018.02.003

Guillette, L. M., Scott, A. C. Y., & Healy, S. D. (2016). Social learning in nest-building birds: A role for familiarity. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1827), 20152685. https://doi.org/10.1098/rspb.2015.2685

Gunderson, J., & Gunderson, L. (2004). *Intelligence (is not equal to) Autonomy (is not equal to) Capability*.

Gunkel, D. J. (2018). *Robot Rights*. https://doi.org/10.7551/mitpress/11444.001.0001

Guo, K., Li, Z., Yan, Y., & Li, W. (2019). Viewing heterospecific facial expressions : An eye-tracking study of human and monkey viewers. *Experimental Brain Research*, *237*(8), 2045-2059. https://doi.org/10.1007/s00221-019-05574-3

Gut, W., Crump, L., Zinsstag, J., Hattendorf, J., & Hediger, K. (2018). The effect of human interaction on guinea pig behavior in animal-assisted therapy. *Journal of Veterinary Behavior*, *25*, 56-64. https://doi.org/10.1016/j.jveb.2018.02.004

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology*, *5*(12), e321. https://doi.org/10.1371/journal.pbio.0050321

Hameed, I., Tan, Z.-H., Thomsen, N., & Duan, X. (2016). *User Acceptance of Social Robots*.

Haute Autorité de Santé. (2018). Trouble du spectre de l'autisme - Signes d'alerte, repérage, diagnostic et évaluation chez l'enfant et l'adolescent. Recommandation de bonne pratique.

Hildt, E. (2019). Artificial Intelligence : Does Consciousness Matter? *Frontiers in Psychology*, *10*, 1535. https://doi.org/10.3389/fpsyg.2019.01535

Hockstein, N., Gourin, C., Faust, R., & Terris, D. (2007). A history of robots : From science fiction to surgical robotics. *Journal of Robotic Surgery*, *1*, 113-118. https://doi.org/10.1007/s11701-007-0021-2

Holzhaider, J. C., Hunt, G. R., & Gray, R. D. (2010). Social learning in New Caledonian crows. *Learning & Behavior*, *38*(3), 206-219. https://doi.org/10.3758/LB.38.3.206

Hong, A., Lunscher, N., Hu, T., Tsuboi, Y., Zhang, X., Alves, S. F. dos R., Nejat, G., & Benhabib, B. (2020). A Multimodal Emotional Human-Robot Interaction Architecture for Social Robots Engaged in Bidirectional Communication\vspace*7pt. *IEEE Transactions on Cybernetics*, 1-15. https://doi.org/10.1109/TCYB.2020.2974688

Hosseini, S. M. F., Hilliger, S., Barnes, J., Jeon, M., Park, C. H., & Howard, A. M. (2017). Love at first sight : Mere exposure to robot appearance leaves impressions similar to interactions with physical robots. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 615-620. https://doi.org/10.1109/ROMAN.2017.8172366

Houx, A. B., & Ten cate, C. (1999). Song learning from playback in zebra finches : Is there an effect of operant contingency? *Animal Behaviour*, *57*(4), 837-845. https://doi.org/10.1006/anbe.1998.1046

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays : Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *33*(1), 3-13. https://doi.org/10.1023/a:1022270118899

Hudlicka, E. (2011). Guidelines for Designing Computational Models of Emotions. *International Journal of Synthetic Emotions (IJSE)*, *2*(1), 26-79. https://doi.org/10.4018/jse.2011010103

Hung, L., Liu, C., Woldum, E., Au-Yeung, A., Berndt, A., Wallsworth, C., Horne, N., Gregorio, M., Mann, J., & Chaudhury, H. (2019). The benefits of and barriers to using a social robot PARO in care settings : A scoping review. *BMC Geriatrics*, *19*(1), 232. https://doi.org/10.1186/s12877-019-1244-6

Hyland Bruno, J., & Tchernichovski, O. (2017). Regularities in zebra finch song beyond the repeated motif. *Behavioural Processes*, *163*, 53-59. https://doi.org/10.1016/j.beproc.2017.11.001

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. *Bird Vocalizations*, *61*.

Iverson, J. M. (2010). Developing language in a developing body : The relationship between motor development and language development. *Journal of child language*, *37*(2), 229-261. https://doi.org/10.1017/S0305000909990432

Jamet, F., Masson, O., Jacquet, B., Stilgenbauer, J.-L., & Baratgin, J. (2018). Learning by Teaching with Humanoid Robot : A New Powerful Experimental Tool to Improve Children's Learning Ability. *Journal of Robotics*, *2018*, e4578762. https://doi.org/10.1155/2018/4578762

Janik, V. M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology*, *28*, 60-65. https://doi.org/10.1016/j.conb.2014.06.010

Johal, W. (2020). Research Trends in Social Robots for Learning. *Current Robotics Reports*, *1*, 1-9. https://doi.org/10.1007/s43154-020-00008-3

Johansson, R., Skantze, G., & Jönsson, A. (2017). *A Psychotherapy Training Environment with Virtual Patients Implemented Using the Furhat Robot Platform*. 184-187. https://doi.org/10.1007/978-3-319-67401-8_20

Jolly, L., Pittet, F., Caudal, J.-P., Mouret, J.-B., Houdelier, C., Lumineau, S., & de Margerie, E. (2016). Animal-to-robot social attachment : Initial requisites in a gallinaceous bird. *Bioinspiration & Biomimetics, 11*(1), 016007. https://doi.org/10.1088/1748-3190/11/1/016007

Kamioka, H., Okada, S., Tsutani, K., Park, H., Okuizumi, H., Handa, S., Oshio, T., Park, S.-J., Kitayuguchi, J., Abe, T., Honda, T., & Mutoh, Y. (2014). Effectiveness of animal-assisted therapy : A systematic review of randomized controlled trials. *Complementary Therapies in Medicine, 22*(2), 371-390. https://doi.org/10.1016/j.ctim.2013.12.016

Kanazawa, S. (1996). Recognition of facial expressions in a Japanese monkey (*Macaca fuscata*) and humans (*Homo sapiens*). *Primates, 37*(1), 25-38. https://doi.org/10.1007/BF02382917

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive Robots as Social Partners and Peer Tutors for Children : A Field Trial. *Human Computer Interaction (Special issues on human-robot interaction), 19*, 61-84. https://doi.org/10.1207/s15327051hci1901&2_4

Kappassov, Z., Corrales Ramon, J. A., & Perdereau, V. (2015). Tactile sensing in dexterous robot hands—Review. *Robotics and Autonomous Systems, 74*. https://doi.org/10.1016/j.robot.2015.07.015

Katz, M., & Lachlan, R. F. (2003). Social learning of food types in zebra finches (*Taenopygia guttata*) is directed by demonstrator sex and feeding activity. *Animal Cognition, 6*(1), 11–16. https://doi.org/10.1007/s10071-003-0158-y

Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech : A review. *International Journal of Speech Technology, 15*(2), 99-117. https://doi.org/10.1007/s10772-011-9125-1

Kopman, V., Laut, J., Polverino, G., & Porfiri, M. (2013). Closed-loop control of zebrafish response using a bioinspired robotic-fish in a preference test. *Journal of The Royal Society Interface, 10*(78), 20120540. https://doi.org/10.1098/rsif.2012.0540

Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., & Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction, 13*, 1-9. https://doi.org/10.1016/j.ijcci.2017.04.001

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Interactive robots for communication-care : 14th IEEE Workshop on Robot and Human Interactive Communication, RO-MAN 2005. *14th IEEE*

Workshop on Robot and Human Interactive Communication, RO-MAN 2005, 341-346. https://doi.org/10.1109/ROMAN.2005.1513802

Kremer, L., Klein Holkenborg, S. E. J., Reimert, I., Bolhuis, J. E., & Webb, L. E. (2020). The nuts and bolts of animal emotion. *Neuroscience & Biobehavioral Reviews*, *113*, 273-286. https://doi.org/10.1016/j.neubiorev.2020.01.028

Kühnlenz, B., Busse, F., Förtsch, P., Wolf, M., & Kühnlenz, K. (2018, août 28). *Effect of Explicit Emotional Adaptation on Prosocial Behavior of Humans towards Robots Depends on Prior Robot Experience*. https://doi.org/10.1109/ROMAN.2018.8525515

Lachlan, R. F., Crooks, L., & Laland, K. N. (1998). Who follows whom? Shoaling preferences and social learning of foraging information in guppies. *Animal Behaviour*, *56*(1), 181–190. https://doi.org/10.1006/anbe.1998.0760

Lambert, A., Norouzi, N., Bruder, G., & Welch, G. (2020). A Systematic Review of Ten Years of Research on Human Interaction with Social Robots. *International Journal of Human–Computer Interaction*, *36*(19), 1804-1817. https://doi.org/10.1080/10447318.2020.1801172

Landgraf, T., Bierbach, D., Nguyen, H., Muggelberg, N., Romanczuk, P., & Krause, J. (2016). RoboFish : Increased acceptance of interactive robotic fish with realistic eyes and natural motion patterns by live Trinidadian guppies. *Bioinspiration & Biomimetics*, *11*(1), 015001. https://doi.org/10.1088/1748-3190/11/1/015001

Lang, P. J. (2010). Emotion and Motivation : Toward Consensus Definitions and a Common Research Purpose. *Emotion Review*, *2*(3), 229-233. https://doi.org/10.1177/1754073910361984

Laschi, C., Mazzolai, B., Patane, F., Mattoli, V., Dario, P., Ishii, H., Ogura, M., Kurisu, S., Komura, A., & Takanishi, A. (2006). Design and Development of a Legged Rat Robot for Studying Animal-Robot Interaction. *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006. BioRob 2006.*, 631-636. https://doi.org/10.1109/BIOROB.2006.1639160

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Ledoux, J., Alberini, C., & Phelps, L. (2016). What We Talk About When We Talk About Emotions. *Cell*. https://doi.org/10.1016/j.cell.2016.11.029

Leenes, R., Palmerini, E., Koops, B.-J., Bertolini, A., Salvini, P., & Lucivero, F. (2017). Regulatory challenges of robotics : Some guidelines for addressing legal and ethical issues. *Law, Innovation and Technology*, *9*(1), 1-44. https://doi.org/10.1080/17579961.2017.1304921

Leite, I., Martinho, C., & Paiva, A. (2013). Social Robots for Long-Term Interaction : A Survey. *International Journal of Social Robotics*, *5*(2), 291-308. https://doi.org/10.1007/s12369-013-0178-y

Le Maguer, L. (2019). *Cultural evolution of birdsong in the laboratory* [Phd thesis, Université Paris Nanterre]. https://tel.archives-ouvertes.fr/tel-03091982

Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition and Emotion*, *13*(5), 481-504. https://doi.org/10.1080/026999399379159

Li, J. (2015). The benefit of being physically present : A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, *77*, 23-37. https://doi.org/10.1016/j.ijhcs.2015.01.001

Li, J., & Chignell, M. (2011). Communication of Emotion in Social Robots through Simple Head and Arm Movements. *International Journal of Social Robotics*, *3*(2), 125-142. https://doi.org/10.1007/s12369-010-0071-x

Li, Y., Dai, Q., Hou, R., Zhang, Z., Chen, P., Xue, R., Feng, F., Chen, C., Liu, J., Gu, X., Zhang, Z., & Qi, D. (2017). Giant pandas can discriminate the emotions of human facial pictures. *Scientific Reports*, *7*(1), 8437. https://doi.org/10.1038/s41598-017-08789-y

Liu, W.-C., Kohn, J., Szwed, S. K., Pariser, E., Sepe, S., Haripal, B., Oshimori, N., Marsala, M., Miyanohara, A., & Lee, R. (2015). Human mutant huntingtin disrupts vocal learning in transgenic songbirds. *Nature Neuroscience*, *18*(11), 1617-1622. https://doi.org/10.1038/nn.4133

Löffler, D., Dörrenbächer, J., & Hassenzahl, M. (2020). The Uncanny Valley Effect in Zoomorphic Robots : The U-Shaped Relation Between Animal Likeness and Likeability. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 261-270. https://doi.org/10.1145/3319502.3374788

Löffler, D., Schmidt, N., & Tscharn, R. (2018). Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 334-343. https://doi.org/10.1145/3171221.3171261

Lorenz, K. (1935). Der kumpan in der umwelt des vogels. *Journal für Ornithologie* 83(3):289–413. [Translated in K. Lorenz. 1970. Studies in Animal and Human Behaviour. Harvard University Press, Cambridge.]

Man, K., & Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence*, *1*(10), 446-452. https://doi.org/10.1038/s42256-019-0103-7

Manzotti, R., & Chella, A. (2018). Good Old-Fashioned Artificial Consciousness and the Intermediate Level Fallacy. *Frontiers in Robotics and AI*, *5*, 39. https://doi.org/10.3389/frobt.2018.00039

Marinetti, C., Moore, P., Lucas, P., & Parkinson, B. (2011). Emotions in Social Interactions : Unfolding Emotional Experience. In *Cognitive Technologies* (p. 31-46). https://doi.org/10.1007/978-3-642-15184-2_3

Marras, S., & Porfiri, M. (2012). Fish and robots swimming together : Attraction towards the robot demands biomimetic locomotion. *Journal of The Royal Society Interface*, *9*(73), 1856-1868. https://doi.org/10.1098/rsif.2012.0084

Masserman, J. H., Wechkin, S., & Terris, W. (1964). « Altruistic » behavior in rhesus monkeys. *The American Journal of Psychiatry*, *121*(6), 584-585. https://doi.org/10.1176/ajp.121.6.584

Mavridis, N. (2015). A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, *63*, 22-35. https://doi.org/10.1016/j.robot.2014.09.031

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Mag.* https://doi.org/10.1609/aimag.v27i4.1904

McCrae, R., & Costa, P. (1997). Personality Trait Structure as a Human Universal. *The American psychologist*, *52*, 509-516. https://doi.org/10.1037//0003-066X.52.5.509

McNeill, D. (1996). *Hand and Mind : What Gestures Reveal about Thought*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/H/bo3641188.html

Maréchal, L., Levy, X., Meints, K., & Majolo, B. (2017). Experience-based human perception of facial expressions in Barbary macaques (Macaca sylvanus). *PeerJ*, *5*, e3413. https://doi.org/10.7717/peerj.3413

Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour*, *58*(4), 825-830. https://doi.org/10.1006/anbe.1999.1219

Meghdari, A., Alemi, M., Zakipour, M., & Kashanian, S. A. (2019). Design and Realization of a Sign Language Educational Humanoid Robot. *Journal of Intelligent & Robotic Systems*, *95*(1), 3-17. https://doi.org/10.1007/s10846-018-0860-2

Meltzoff, A. N. (2005). Imitation and Other Minds : The « Like Me » Hypothesis. In *Perspectives on imitation : From neuroscience to social science : Vol. 2 : Imitation, human development, and culture* (p. 55-77). MIT Press.

Meltzoff, A. N. (2007). 'Like me' : A foundation for social cognition. *Developmental Science*, *10*(1), 126-134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Mendonça, T., Bienboire-Frosini, C., Menuge, F., Leclercq, J., Lafont-Lecuelle, C., Arroub, S., & Pageat, P. (2019). The Impact of Equine-Assisted Therapy on Equine Behavioral and Physiological Responses. *Animals*, *9*(7), 409. https://doi.org/10.3390/ani9070409

Miklósi, Á., Korondi, P., Matellán, V., & Gácsi, M. (2017). Ethorobotics : A New Approach to Human-Robot Relationship. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00958

Miklosi, A., Polgárdi, R., Topál, J., & Csányi, V. (2000). Intentional behaviour in dog-human communication : An experimental analysis of « showing » behaviour in the dog. *Animal Cognition*, *3*, 159-166. https://doi.org/10.1007/s100710000072

Miklósi, A., & Topál, J. (2013). What does it take to become « best friends »? Evolutionary changes in canine social competence. *Trends in Cognitive Sciences*, *17*(6), 287-294. https://doi.org/10.1016/j.tics.2013.04.005

Miriyev, A., & Kovač, M. (2020). Skills for physical artificial intelligence. *Nature Machine Intelligence*, *2*(11), 658-660. https://doi.org/10.1038/s42256-020-00258-y

Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Rocella, S., Carrozza, M. C., Dario, P., & Takanishi, A. (2004). Effective emotional expressions with expression humanoid robot WE-4RII : Integration of humanoid robot hand RCH-1. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, *3*, 2203-2208 vol.3. https://doi.org/10.1109/IROS.2004.1389736

Moore, C. (2006). *The Development of Commonsense Psychology*. Psychology Press. https://doi.org/10.4324/9780203843246

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine*, *19*(2), 98-100. https://doi.org/10.1109/MRA.2012.2192811

Müller, C. A., Schmitt, K., Barber, A. L. A., & Huber, L. (2015). Dogs Can Discriminate Emotional Expressions of Human Faces. *Current Biology*, *25*(5), 601-605. https://doi.org/10.1016/j.cub.2014.12.055

Nahm, F. K., Perret, A., Amaral, D. G., & Albright, T. D. (1997). How do monkeys look at faces? *Journal of Cognitive Neuroscience*, *9*(5), 611-623. https://doi.org/10.1162/jocn.1997.9.5.611

Neethirajan, S., Reimert, I., & Kemp, B. (2021). Measuring Farm Animal Emotions—Sensor-Based Approaches. *Sensors*, *21*(2), 553. https://doi.org/10.3390/s21020553

Nicol, C. J., & Pope, S. J. (1994). Social learning in small flocks of laying hens. *Animal Behaviour*, *47*(6), 1289–1296. https://doi.org/10.1006/anbe.1994.1177

Nicol, C. J., & Pope, S. J. (1999). The effects of demonstrator social status and prior foraging success on social learning in laying hens. *Animal Behaviour*, *57*(1), 163–171. https://doi.org/10.1006/anbe.1998.0920

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *The Journal of Comparative Neurology*, *165*(4), 457-486. https://doi.org/10.1002/cne.901650405

Nowak, K., & Biocca, F. (2003). The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence Teleoperators & Virtual Environments*, *12*, 481-494. https://doi.org/10.1162/105474603322761289

Oliveira, R., Arriaga, P., Santos, F. P., Mascarenhas, S., & Paiva, A. (2021). Towards prosocial design : A scoping review of the use of robots and virtual agents to trigger prosocial behaviour. *Computers in Human Behavior*, *114*, 106547. https://doi.org/10.1016/j.chb.2020.106547

Ostojić, L., & Clayton, N. S. (2014). Behavioural coordination of dogs in a cooperative problem-solving task with a conspecific and a human partner. *Animal Cognition*, *17*(2), 445-459. https://doi.org/10.1007/s10071-013-0676-1

Over, H., & Carpenter, M. (2011). Putting the social into social learning : Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*(2), 182-192. https://doi.org/10.1037/a0024555

Pagliarini, S., Hinaut, X., & Leblois, A. (2018). A bio-inspired model towards vocal gesture learning in songbird. *2018 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. https://hal.inria.fr/hal-01906459

Palagi, E., Celeghin, A., Tamietto, M., Winkielman, P., & Norscia, I. (2020). The neuroethology of spontaneous mimicry and emotional contagion in human and non-human animals. *Neuroscience & Biobehavioral Reviews*, *111*, 149-165. https://doi.org/10.1016/j.neubiorev.2020.01.020

Palagi, E., Nicotra, V., & Cordoni, G. (2015). Rapid mimicry and emotional contagion in domestic dogs. *Royal Society Open Science*, *2*(12), 150505. https://doi.org/10.1098/rsos.150505

Panksepp J. (2004). Affective neuroscience: The foundations of human and animal emotions. Oxford University Press.

Panksepp, J. (2005). Affective consciousness : Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30-80. https://doi.org/10.1016/j.concog.2004.10.004

Panksepp, J. (2010). The Evolutionary Sources of Jealousy. In *Handbook of Jealousy* (p. 101-120). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444323542.ch6

Park, H. W., & Howard, A. M. (2014). *Robot Learners : Interactive Instance-based Learning and Its Application to Therapeutic Tasks*. https://smartech.gatech.edu/handle/1853/53773

Parker, A. E., Mathis, E. T., & Kupersmidt, J. B. (2013). How is this child feeling? Preschool-aged children's ability to recognize emotion in faces and body poses. *Early education and development*, *24*(2), 188-211. https://doi.org/10.1080/10409289.2012.657536

Parr, L. A. (2003). The Discrimination of Faces and Their Emotional Content by Chimpanzees (Pan troglodytes). In *Emotions inside out : 130 years after Darwin's : The expression of the emotions in man and animals* (p. 56-78). New York Academy of Sciences.

Parr, L. A., Dove, T., & Hopkins, W. D. (1998). Why faces may be special : Evidence of the inversion effect in chimpanzees. *Journal of Cognitive Neuroscience*, *10*(5), 615-622. https://doi.org/10.1162/089892998563013

Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying Chimpanzee Facial Expressions Using Muscle Action. *Emotion (Washington, D.C.)*, *7*(1), 172-181. https://doi.org/10.1037/1528-3542.7.1.172

Partan, S. R., Larco, C. P., & Owens, M. J. (2009). Wild tree squirrels respond with multisensory enhancement to conspecific robot alarm behaviour. *Animal Behaviour*, *77*(5), 1127-1135. https://doi.org/10.1016/j.anbehav.2008.12.029

Patane, F., Mattoli, V., Laschi, C., Mazzolai, B., Dario, P., Ishii, H., & Takanishi, A. (2007). Biomechatronic design and development of a legged rat robot. *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. https://doi.org/10.1109/ROBIO.2007.4522273

Patricelli, G. L., Coleman, S. W., & Borgia, G. (2006). Male satin bowerbirds, *Ptilonorhynchus violaceus*, adjust their display intensity in response to female startling : An experiment with robotic females. *Animal Behaviour*, *71*(1), 49-59.

Patricelli, G., & Krakauer, A. (2009). Tactical allocation of effort among multiple signals in sage grouse : An experiment with a robotic female. *Behavioral Ecology*, *21*, 97-106. https://doi.org/10.1093/beheco/arp155

Paul, E. S., & Mendl, M. T. (2018). Animal emotion : Descriptive and prescriptive definitions and their implications for a comparative perspective. *Applied Animal Behaviour Science*, *205*, 202-209. https://doi.org/10.1016/j.applanim.2018.01.008

Peeters, M., Tuijl, H., Rutte, C., & Reymen, I. (2006). Personality and team performance : A meta-analysis. *European Journal of Personality*, *20*, 377-396. https://doi.org/10.1002/per.588

Perloff, R. M. (1982). Gender constancy and same-sex imitation : A developmental study. *The Journal of Psychology: Interdisciplinary and Applied*, *111*(1), 81-86. https://doi.org/10.1080/00223980.1982.9923516

Petersen, K. H., Napp, N., Stuart-Smith, R., Rus, D., & Kovac, M. (2019). A review of collective robotic construction. *Science Robotics*, *4*(28), eaau8479. https://doi.org/10.1126/scirobotics.aau8479

Peterson, C. C., Slaughter, V., & Brownell, C. (2015). Children with autism spectrum disorder are skilled at reading emotion body language. *Journal of Experimental Child Psychology*, *139*, 35-50. https://doi.org/10.1016/j.jecp.2015.04.012

Petkov, C., & Jarvis, E. (2012). Birds, primates, and spoken language origins : Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*, 12. https://doi.org/10.3389/fnevo.2012.00012

Phamduy, P., Polverino, G., Fuller, R. C., & Porfiri, M. (2014). Fish and robot dancing together : Bluefin killifish females respond differently to the courtship of a robot with varying color morphs. *Bioinspiration & Biomimetics*, *9*(3), 036021. https://doi.org/10.1088/1748-3182/9/3/036021

Phan, M. L., Pytte, C. L., & Vicario, D. S. (2006). Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 1088-1093. https://doi.org/10.1073/pnas.0510136103

Phillips, E., Ullman, D., de Graaf, M., & Malle, B. (2017). What Does A Robot Look Like? : A Multi-Site Examination of User Expectations About Robot Appearance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*, 1215-1219. https://doi.org/10.1177/1541931213601786

Pitti, A., Quoy, M., Boucenna, S., & Lavandier, C. (2021). Brain-inspired model for early vocal learning and correspondence matching using free-energy optimization. *PLOS Computational Biology*, *17*(2), e1008566. https://doi.org/10.1371/journal.pcbi.1008566

Polverino, G., & Porfiri, M. (2013). Mosquitofish (Gambusia affinis) responds differentially to a robotic fish of varying swimming depth and aspect ratio. *Behavioural Brain Research*, *250*, 133-138. https://doi.org/10.1016/j.bbr.2013.05.008

Prelock, P. J., & Nelson, N. W. (2012). Language and communication in autism : An integrated view. *Pediatric Clinics of North America*, *59*(1), 129-145, xi. https://doi.org/10.1016/j.pcl.2011.10.008

Preston, S. D., & de Waal, F. B. M. (2002). Empathy : Its ultimate and proximate bases. *The Behavioral and Brain Sciences*, *25*(1), 1-20; discussion 20-71. https://doi.org/10.1017/s0140525x02000018

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *Journal of Comparative and Physiological Psychology*, *93*(2), 260-277. https://doi.org/10.1037/h0077553

Proops, L., Grounds, K., Smith, A. V., & McComb, K. (2018). Animals Remember Previous Facial Expressions that Specific Humans Have Exhibited. *Current Biology*, *28*(9), 1428-1432.e4. https://doi.org/10.1016/j.cub.2018.03.035

Quinn, L., Schuster, L., Aguilar-Rivera, M., Arnold, J., Ball, D., Gygi, E., Holt, J., Lee, D., Taufatofua, J., Wiles, J., & Chiba, A. (2018). When rats rescue robots. *Animal Behavior and Cognition*, *5*, 368-379. https://doi.org/10.26451/abc.05.04.04.2018

Raimo, S., Iona, T., Di Vita, A., Boccia, M., Buratin, S., Ruggeri, F., Iosa, M., Guariglia, C., Grossi, D., & Palermo, L. (2021). The development of body representations in school-aged children. *Applied Neuropsychology: Child*, *10*(4), 327-339. https://doi.org/10.1080/21622965.2019.1703704

Raj, M., & Seamans, R. (2019). Primer on artificial intelligence and robotics. *Journal of Organization Design*, *8*(1), 11. https://doi.org/10.1186/s41469-019-0050-0

Rawal, N., & Stock-Homburg, R. M. (2021). Facial emotion expressions in human-robot interaction : A survey. *arXiv:2103.07169 [cs]*. http://arxiv.org/abs/2103.07169

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, *14*(4), 302-308. https://doi.org/10.1111/1467-9280.14431

Reed, C. L., Stone, V. E., Grubb, J. D., & McGoldrick, J. E. (2006). Turning configural processing upside down : Part and whole body postures. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 73-87. https://doi.org/10.1037/0096-1523.32.1.73

Reichmuth, C., & Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology*, *28*, 66-71. https://doi.org/10.1016/j.conb.2014.06.011

Reimert, I., Bolhuis, J. E., Kemp, B., & Rodenburg, T. B. (2013). Indicators of positive and negative emotions and emotional contagion in pigs. *Physiology & Behavior*, *109*, 42-50. https://doi.org/10.1016/j.physbeh.2012.11.002

Rice, G. E., & Gainer, P. (1962). « Altruism » in the albino rat. *Journal of Comparative and Physiological Psychology*, *55*(1), 123-125. https://doi.org/10.1037/h0042276

Richards, N., & Smart, W. (2013). How Should the Law Think About Robots? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2263363

Riek, L. (2012). Wizard of Oz Studies in HRI : A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction*, *1*. https://doi.org/10.5898/JHRI.1.1.Riek

Riggio, H. R., & Riggio, R. E. (2002). Emotional Expressiveness, Extraversion, and Neuroticism : A Meta-Analysis. *Journal of Nonverbal Behavior*, *26*(4), 195-218. https://doi.org/10.1023/A:1022117500440

Robert, L. P., Alahmad, R., Esterwood, C., Kim, S., You, S., & Zhang, Q. (2020). A Review of Personality in Human Robot Interactions. *arXiv:2001.11777 [cs]*. http://arxiv.org/abs/2001.11777

Roberts, B. W. (2009). Back to the Future : Personality and Assessment and Personality Development. *Journal of research in personality*, *43*(2), 137-145. https://doi.org/10.1016/j.jrp.2008.12.015

Roberts, T. F., Tschida, K. A., Klein, M. E., & Mooney, R. (2010). Rapid spine stabilization and synaptic enhancement at the onset of behavioural learning. *Nature*, *463*(7283), 948-952. https://doi.org/10.1038/nature08759

Rohlfing, K. J., Grimminger, A., & Wrede, B. (2020). The caregiver's role in keeping a child–robot interaction going. In *International Perspectives on Digital Media and Early Literacy*. Routledge.

Rolls, E. T. (2014). *Emotion and decision-making explained* (p. xvii, 710). Oxford University Press.

Romano, D., Donati, E., Benelli, G., & Stefanini, C. (2019). A review on animal-robot interaction : From bio-hybrid organisms to mixed societies. *Biological Cybernetics*, *113*(3), 201-225. https://doi.org/10.1007/s00422-018-0787-5

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding Tutor Learning : Knowledge-Building and Knowledge-Telling in Peer Tutors' Explanations and Questions. *Review of Educational Research*, *77*(4), 534-574. https://doi.org/10.3102/0034654307309920

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, *5*(1), 17-34. https://doi.org/10.1007/s12369-012-0173-8

Russo, L., Craig, F., Ruggiero, M., Mancuso, C., Galluzzi, R., Lorenzo, A., Fanizza, I., & Trabacca, A. (2018). Exploring Visual Perspective Taking and body awareness in children with Autism Spectrum Disorder. *Cognitive Neuropsychiatry*, *23*(4), 254-265. https://doi.org/10.1080/13546805.2018.1486182

Saunders, J., Syrdal, D. S., Koay, K. L., Burke, N., & Dautenhahn, K. (2016). "Teach Me–Show Me"—End-User Personalization of a Smart Home and Companion Robot. *IEEE Transactions on Human-Machine Systems*, *46*(1), 27-40. https://doi.org/10.1109/THMS.2015.2445105

Schleihauf, H., Hoehl, S., Tsvetkova, N., König, A., Mombaur, K., & Pauen, S. (2021). Preschoolers' Motivation to Over-Imitate Humans and Robots. *Child Development*, *92*(1), 222-238. https://doi.org/10.1111/cdev.13403

Schuppli, C., Meulman, E. J. M., Forss, S. I. F., Aprilinayati, F., van Noordwijk, M. A., & van Schaik, C. P. (2016). Observational social learning and socially induced practice of routine skills in immature wild orang-utans. *Animal Behaviour*, *119*, 87-98. https://doi.org/10.1016/j.anbehav.2016.06.014

Schwoebel, J., Boronat, C. B., & Branch Coslett, H. (2002). The man who executed « imagined » movements : Evidence for dissociable components of the body schema. *Brain and Cognition*, *50*(1), 1-16. https://doi.org/10.1016/s0278-2626(02)00005-2

Schwoebel, J., & Coslett, H. B. (2005). Evidence for multiple, distinct representations of the human body. *Journal of Cognitive Neuroscience*, *17*(4), 543-553. https://doi.org/10.1162/0898929053467587

Shank, S. S., & Margoliash, D. (2009). Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature*, *458*(7234), 73-77. https://doi.org/10.1038/nature07615

Shibata, T., & Wada, K. (2011). Robot therapy : A new approach for mental healthcare of the elderly - a mini-review. *Gerontology*, *57*(4), 378-386. https://doi.org/10.1159/000319015

Shimpi, P. M., Akhtar, N., & Moore, C. (2013). Toddlers' imitative learning in interactive and observational contexts : The role of age and familiarity of the model. *Journal of Experimental Child Psychology*, *116*(2), 309-323. https://doi.org/10.1016/j.jecp.2013.06.008

Sirigu, A., Grafman, J., Bressler, K., & Sunderland, T. (1991). Multiple representations contribute to body knowledge processing. Evidence from a case of autotopagnosia. *Brain: A Journal of Neurology*, *114 (Pt 1B)*, 629-642. https://doi.org/10.1093/brain/114.1.629

Sossinka, R., & Böhner, J. (1980). Song types in the zebra finch *Poephila guttata castanotis*. *Zeitschrift Für Tierpsychologie*, *53*(2), 123–132.

Spatola, N., & Wudarczyk, O. A. (2021). Ascribing emotions to robots : Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior*, *124*, 106934. https://doi.org/10.1016/j.chb.2021.106934

Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion Recognition for Human-Robot Interaction : Recent Advances and Future Perspectives. *Frontiers in Robotics and AI*, *7*, 145. https://doi.org/10.3389/frobt.2020.532279

Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences*, *106*(43), 18362-18366. https://doi.org/10.1073/pnas.0910063106

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies : An ERP study on the body-inversion effect. *Neuroreport*, *15*(5), 777-780. https://doi.org/10.1097/00001756-200404090-00007

Stiegler, L. N. (2015). Examining the Echolalia Literature : Where Do Speech-Language Pathologists Stand? *American Journal of Speech-Language Pathology*, *24*(4), 750-762. https://doi.org/10.1044/2015_AJSLP-14-0166

Stoeger, A. S., & Manger, P. (2014). Vocal learning in elephants : Neural bases and adaptive context. *Current Opinion in Neurobiology*, *28*, 101-107. https://doi.org/10.1016/j.conb.2014.07.001

Sung, J., Christensen, H. I., & Grinter, R. E. (2009). Robots in the wild : Understanding long-term use. *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 45-52. https://doi.org/10.1145/1514095.1514106

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and Communication in Autism. In *Handbook of autism and pervasive developmental disorders : Diagnosis, development, neurobiology, and behavior, Vol. 1, 3rd ed* (p. 335-364). John Wiley & Sons Inc.

Takahashi, D. Y., Liao, D. A., & Ghazanfar, A. A. (2017). Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Current Biology*, *27*(12), 1844-1852.e6. https://doi.org/10.1016/j.cub.2017.05.004

Takanishi, A., Aoki, T., Ito, M., Ohkawa, Y., & Yamaguchi, J. (1998). Interaction between creature and robot : Development of an experiment system for rat and rat robot interaction. *Proceedings. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications (Cat. No.98CH36190)*, *3*, 1975-1980 vol.3. https://doi.org/10.1109/IROS.1998.724896

Tanaka, F., & Matsuzoe, S. (2012). Children Teach a Care-Receiving Robot to Promote Their Learning : Field Experiments in a Classroom for Vocabulary Learning. *Journal of Human-Robot Interaction*, *1*. https://doi.org/10.5898/JHRI.1.1.Tanaka

Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., Rusu, A. S., & David, D. O. (2012). Children with autism social engagement in interaction with Nao, an imitative robot : A series of single case experiments. *Interaction Studies*, *13*(3), 315-347. https://doi.org/10.1075/is.13.3.01tap

Tasa, K., Sears, G. J., & Schat, A. C. H. (2011). Personality and teamwork behavior in context : The cross-level moderating role of collective efficacy. *Journal of Organizational Behavior*, *32*(1), 65-85. https://doi.org/10.1002/job.680

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots : The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75-84. https://doi.org/10.1016/j.chb.2014.05.014

Taylor, R., Klein, B., Stein, J., & Ryan, M. (2008). Faux frogs : Multimodal signalling and the value of robotics in the study of animal behaviour. *Animal Behaviour*, *76*, 1089-1097. https://doi.org/10.1016/j.anbehav.2008.01.031

ten Cate, C. (1991). Behaviour-contingent exposure to taped song and zebra finch song learning. *Animal Behaviour*, *42*(5), 857–859. https://doi.org/10.1016/S0003-3472(05)80131-9

Tinbergen, N. (1951). *The study of instinct* (p. xii, 237). Clarendon Press/Oxford University Press.

Tolksdorf, N. F., Crawshaw, C. E., & Rohlfing, K. J. (2021). Comparing the Effects of a Different Social Partner (Social Robot vs. Human) on Children's Social Referencing in Interaction. *Frontiers in Education*, *5*, 295. https://doi.org/10.3389/feduc.2020.569615

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In *Handbook of emotions, 3rd ed* (p. 114-137). The Guilford Press.

Topál, J., Gácsi, M., Miklósi, Á., Virányi, Z., Kubinyi, E., & Csányi, V. (2005). Attachment to humans : A comparative study on hand-reared wolves and differently socialized dog puppies. *Animal Behaviour*, *70*(6), 1367-1375. https://doi.org/10.1016/j.anbehav.2005.03.025

Topál, J., Gergely, G., Erdohegyi, A., Csibra, G., & Miklósi, A. (2009). Differential sensitivity to human communication in dogs, wolves, and human infants. *Science (New York, N.Y.)*, *325*(5945), 1269-1272. https://doi.org/10.1126/science.1176960

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

Turner, T. J., & Ortony, A. (1992). Basic emotions : Can conflicting criteria converge? *Psychological Review*, *99*(3), 566-571. https://doi.org/10.1037/0033-295X.99.3.566

Tuyen, N. T. V., Jeong, S., & Chong, N. (2018). Emotional Bodily Expressions for Culturally Competent Robots through Long Term Human-Robot Interaction. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. https://doi.org/10.1109/IROS.2018.8593974

Uzgiris, I. C. (1981). Two Functions of Imitation During Infancy. *International Journal of Behavioral Development*, *4*(1), 1-12. https://doi.org/10.1177/016502548100400101

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, *7*(3), 487-494. https://doi.org/10.1037/1528-3542.7.3.487

Van Kleef, G. A. (2017). The social effects of emotions are functionally equivalent across expressive modalities. *Psychological Inquiry*, *28*(2-3), 211-216. https://doi.org/10.1080/1047840X.2017.1338102

Van Kleef, G. A. & Côté, S. (2007). Expressing anger in conflict : When it helps and when it hurts. *Journal of Applied Psychology*, *92*(6), 1557-1569. https://doi.org/10.1037/0021-9010.92.6.1557

Wainer, J., Feil-Seifer, D., Shell, D., & Mataric, M. (2006). *The role of physical embodiment in human-robot interaction*. 117-122. https://doi.org/10.1109/ROMAN.2006.314404

Wan, M., Bolger, N., & Champagne, F. A. (2012). Human Perception of Fear in Dogs Varies According to Experience with Dogs. *PLOS ONE*, *7*(12), e51775. https://doi.org/10.1371/journal.pone.0051775

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science (New York, N.Y.)*, *311*(5765), 1301-1303. https://doi.org/10.1126/science.1121448

Westerlund, M. (2020). *An Ethical Framework for Smart Robots*. https://doi.org/10.22215/timreview/1312

White, S. J., Pascall, D. J., & Wilson, A. J. (2020). Towards a comparative approach to the structure of animal personality variation. *Behavioral Ecology*, *31*(2), 340-351. https://doi.org/10.1093/beheco/arz198

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, *39*(1), 114-128. https://doi.org/10.1037//0012-1649.39.1.114

Wilkinson, A., Kuenstner, K., Mueller, J., & Huber, L. (2010). Social learning in a non-social reptile (Geochelone carbonaria). *Biology Letters*, *6*(5), 614-616. https://doi.org/10.1098/rsbl.2010.0092

Williams, H. (1990). Models for song learning in the zebra finch: Fathers or others? *Animal Behaviour*, *39*(4), 745–757. https://doi.org/10.1016/S0003-3472(05)80386-0

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2012). Context-dependent model-based biases in cultural transmission : Children's imitation is affected by model age over model knowledge state. *Evolution and Human Behavior*, *33*(4), 387-394. https://doi.org/10.1016/j.evolhumbehav.2011.11.010

Woods, S. (2006). Exploring the design space of robots : Children's perspectives. *Interacting with Computers*, *18*(6), 1390-1418. https://doi.org/10.1016/j.intcom.2006.05.001

Zabala, U., Rodriguez Rodriguez, I., Martinez-Otzeta, J. M., & Lazkano, E. (2021). Expressing Robot Personality through Talking Body Language. *Applied Sciences*, *11*, 4639. https://doi.org/10.3390/app11104639

Zann, R. A. (1996). *The zebra finch: A synthesis of field and laboratory studies* (Vol. 5). Oxford University Press.

Zecca, M., Mizoguchi, Y., Endo, K., Iida, F., Kawabata, Y., Endo, N., Itoh, K., & Takanishi, A. (2009). *Whole body emotion expressions for KOBIAN humanoid robot—Preliminary experiments with different Emotional patterns—*. 381-386. https://doi.org/10.1109/ROMAN.2009.5326184

Zhao, J., Liang, B., & Chen, Q. (2018). The key technology toward the self-driving car. *International Journal of Intelligent Unmanned Systems*, *6*(1), 2-20. https://doi.org/10.1108/IJIUS-08-2017-0008