Université de Lille

SCALab
SCIENCES COGNITIVES ET SCIENCES AFFECTIVES

Ecole doctorale
Sciences de l'Homme et de la Société
Université Lille Nord de France

Thèse de doctorat de

l'Université de Lille

École Doctorale N° 473
Sciences de l'Homme et de la Société
Spécialité : *Psychologie*

Par
Robin GIGANDET

Percevoir les Êtres Sociaux dans les Agents Artificiels

Perceiving Social Beings in Artificial Agents

Thèse présentée et soutenue à Tourcoing, le 10 décembre 2025
Unité de recherche : UMR CNRS 9193 - SCALab - Sciences Cognitives et Sciences Affectives

Composition du Jury :

Rapporteurs :	Clément BELLETIER	Maître de conférence, Université Clermont-Auvergne
	Thierry CHAMINADE	Chargé de recherche, Aix-Marseille Université
Examinatrices :	Malika AUVRAY	Directrice de recherche, Sorbonne Université
	Serena IVALDI	Directrice de recherche, Université de Lorraine
Président du jury :	Yann COELLO	Professeur des universités, Université de Lille
Directrice de thèse :	Tatjana NAZIR	Directrice de recherche, Université de Lille

Université de Lille

UMR CNRS 9193 - Laboratoire Sciences Cognitives et Sciences Affectives (SCALab)

École Doctorale n° 473 : Sciences de l'Homme et de la Société

Percevoir les Êtres Sociaux dans les Agents Artificiels

Thèse de doctorat de Psychologie

Présentée par **Robin GIGANDET**

Sous la direction de **Tatjana NAZIR**

Soutenue à Tourcoing, le 10 décembre 2025

Composition du jury :

Rapporteurs :	Clément BELLETIER	Maître de conférence, Université Clermont-Auvergne
	Thierry CHAMINADE	Chargé de recherche, Aix-Marseille Université
Examinatrices :	Malika AUVRAY	Directrice de recherche, Sorbonne Université
	Serena IVALDI	Directrice de recherche, Université de Lorraine
Président :	Yann COELLO	Professeur des universités, Université de Lille
Dir. de thèse :	Tatjana NAZIR	Directrice de recherche, Université de Lille

Remerciements

Je tiens à exprimer ma reconnaissance à ma directrice de thèse, Tatjana Nazir, pour sa confiance et la liberté intellectuelle qu'elle m'a toujours laissée. Ses conseils avisés et sa bienveillance constante ont guidé ce travail du début à la fin. Ensuite, je remercie également les membres de mon jury de thèse, Malika Auvray, Clément Belletier, Thierry Chaminade, Yann Coello et Serena Ivaldi, pour l'honneur qu'ils m'ont fait en acceptant d'évaluer cette thèse.

Ma reconnaissance va également à Bing Li et Melisa Yavuz, mes *camarades* de bureau. C'était quand même vachement bien les pauses thé à 15h30 tous les jours ! Ils resteront parmi mes meilleurs souvenirs. J'exprime également ma reconnaissance à l'ensemble du SCALab et tout particulièrement au personnel administratif, un grand merci à Emmanuelle Fournier (Manu) et Sabine Pierzchala ! ainsi qu'à toutes les personnes que j'ai eu l'occasion d'y croiser et avec qui j'ai pu échanger au cours de ces années. Un immense merci à toutes les personnes de la plateforme FR 2052 SCV à Tourcoing, pour leur aide et nos discussions enrichissantes durant la pause thé. J'ai une pensée toute particulière pour Omar Blanco qui était toujours disponible si besoin et à courir pour aider tout le monde, et pour discuter de faits intéressants. Également, à toutes celles et ceux que j'aurais pu oublier au moment d'écrire ces lignes : merci ! Je rends hommage à tous ceux qui m'ont inspiré, de près ou de loin, à toutes celles et ceux qui cherchent sincèrement la connaissance pour le bien.

Je souhaite exprimer une gratitude particulière à mon épouse, Inas, pour sa présence indéfectible, son intelligence, sa patience, son soutien dans certains moments, et sa capacité à être résiliente qui m'inspire chaque jour. Cette thèse lui doit beaucoup. À quand un article Inas Redjem & Robin Gigandet ?

Je souhaite remercier du plus profond du cœur ma mère, Bénédicte, elle est la cause première de tout ce que j'ai pu accomplir. Par son amour, sa patience et ses sacrifices silencieux, elle m'a tout donné. Également, je pense à Pride, à toute ma famille et celle de mon épouse, à mon petit frère, à ma grande sœur, tout comme à mes grands-parents et à mon père.

ٱلْحَمْدُ لِلّٰهِ ٱلَّذِي بِنِعْمَتِهِ تَتِمُّ ٱلصَّالِحَات

Après qu'*Il* m'a accordé la grâce d'achever cette thèse, je *Lui* rends grâce avant tout et en dernier lieu, car *Il* est le plus digne de reconnaissance.

Sommaire

Introduction Générale **18**

I Cadre théorique et conceptuel **23**

1 Agents sociaux artificiels et robots sociaux : définitions, typologies et enjeux **24**
1.1 Définitions des Agents Sociaux Artificiels et Robots Sociaux 24
1.2 Taxonomies et classifications 25
1.3 Enjeux, règles sociales et normes culturelles 29

2 Mécanismes de perceptions et d'interprétation appliqués aux agents artificiels **33**
2.1 L'Anthropomorphisme 33
2.1.1 Du biais perceptif à la projection de l'humain 33
2.1.2 Motivations et mécanismes de l'anthropomorphisme : le modèle SEEK 34
2.1.3 Anthropomorphisme dans l'interaction humain-robot 36
2.2 Attribution d'états mentaux et interprétation des agents artificiels 39
2.2.1 La Posture Intentionnelle 40
2.2.2 Réseau cérébral de mentalisation et attribution d'intentionnalité 43
2.2.3 Mécanismes d'attention sociale et attribution d'intentionnalité 46
2.2.4 Perception de l'esprit et agents artificiels 48
2.2.5 Modèle Like-me et Accordage Social 51
2.2.6 Théorie de la Media Equation et le paradigme CASA 52
2.3 Similitudes et différences des mécanismes face à un humain 55
2.3.1 Convergences des mécanismes 55
2.3.2 Divergences et limites des mécanismes 58

2.3.3 Entre similitudes et différences 59

3 Présence sociale : le sentiment d'être avec un autre 61
3.1 Le concept de présence sociale 61
3.1.1 Définitions de la présence sociale 61
3.1.2 Présence sociale dans l'interaction humain-robot 63
3.2 De la reconnaissance au sentiment d'être avec un autre 65
3.2.1 Une formalisation expérimentale du Perceptual Crossing . . 67
3.2.2 Modélisation du Perceptual Crossing par robotique évolutionnaire 70
3.2.3 Organisation multi-échelle de l'interaction dans le Perceptual Crossing 71
3.2.4 Clarté de la présence de l'autre et réussite conjointe de l'interaction 73
3.2.5 Reconnaissance de l'humain au travers des contingences réciproques 75
3.3 Émergence du sentiment de présence dans la relation 78

4 Discours et temporalité dans l'interaction humain-robot 80
4.1 Temporalité, tour de parole entre humains 80
4.1.1 Fluidité universelle du tour de parole 80
4.1.2 Variabilité et signification sociale des délais 81
4.2 Systèmes de dialogues et limites dans l'interaction humain-robot . 82
4.2.1 Contraintes techniques et modèles prédictifs 82
4.2.2 Incarnation, signaux visuels et attentes temporelles 83
4.3 Réaction cérébrale au discours des robots 85
4.3.1 Vigilance et confiance 85
4.3.2 Marqueurs neurocognitifs 88

II Émergence de l'interaction et Présence sociale 93

5 Adaptation du paradigme du perceptual crossing 95
5.1 Introduction 95
5.2 Méthode 99
5.2.1 Conception 99

5.2.2 Participants . . . 99
5.2.3 Matériel et Apparatus . . . 100
5.2.4 Procédure . . . 106
5.2.5 Analyse des données . . . 110

6 Impact de l'instruction 115
6.1 Hypothèses . . . 115
6.2 Résultats . . . 116
6.2.1 Présence sociale . . . 117
6.2.2 Nombre de croisements . . . 120
6.2.3 Structure temporelle des vélocités . . . 122
6.3 Discussion . . . 139
6.3.1 Présence sociale . . . 139
6.3.2 Nombre de croisement . . . 142
6.3.3 Analyses fractales et multifractales . . . 143
6.3.4 Limites . . . 147

III Attentes temporelles dans les discussions humain-robot 151

7 Style de communication et perception des délais 153
7.1 Introduction . . . 153
7.2 Méthode . . . 155
7.2.1 Conception . . . 155
7.2.2 Participants . . . 156
7.2.3 Matériel et Apparatus . . . 156
7.2.4 Procédure . . . 161
7.2.5 Analyse des données . . . 162
7.3 Résultats . . . 163
7.3.1 Point d'Égalité Subjective . . . 163
7.3.2 Sensibilité aux écarts temporels . . . 166
7.3.3 Perception du robot . . . 168
7.4 Discussion partielle . . . 171

8 Perception sociale face aux délais non-optimaux 174
8.1 Introduction . . . 174

8.2 Méthode 175
8.2.1 Conception 175
8.2.2 Participants 175
8.2.3 Matériel et Apparatus 176
8.2.4 Procédure 176
8.2.5 Analyse des données 178
8.3 Résultats 178
8.3.1 Dimensions de Ho et MacDorman (2017) 178
8.3.2 Dimensions du modèle Almere (Heerink et al. 2010) 181
8.4 Discussion partielle 186
8.5 Discussion générale 187
8.5.1 Limites 190

IV Approche EEG des frontières et limites face au discours robotique 193

9 Un robot parlant d'actions physiques impossibles : validation du paradigme 195
9.1 Introduction 195
9.2 Méthode 197
9.2.1 Conception 197
9.2.2 Participants 197
9.2.3 Matériel et Apparatus 198
9.2.4 Procédure 204
9.2.5 Analyse des données 207
9.3 Résultats 208
9.3.1 Analyse des ERP 208
9.3.2 Analyse du questionnaire 212
9.3.3 Impact des croyances 215
9.4 Discussion partielle 219

10 Un robot parlant de ses émotions 222
10.1 Introduction 222
10.2 Méthode 223

10.3 Résultats 226
10.3.1 Analyse des ERP 226
10.3.2 Analyse du questionnaire 229
10.3.3 Impact de la croyance 231
10.4 Discussion partielle 235
10.5 Discussion générale 236
10.5.1 Limites 239

V Discussion Générale 241

11 Synthèse des contributions 244
11.1 Axe 1 - Présence sociale et dynamiques temporelles avec des agents artificiels dans un environnement minimaliste 244
11.1.1 Rôle de l'instruction sociale 244
11.1.2 Impact de la contingence de l'agent aux actions du participant sur la présence sociale et les comportements d'exploration 246
11.2 Axe 2 - Délai de réponse d'un robot à une question 248
11.2.1 Discussions humain-robot : un délai de réponse d'environ 700 ms 248
11.2.2 Modulation de la tolérance aux écarts temporels par les styles de communication 249
11.2.3 Impact des délais non-optimaux sur l'évaluation globale du robot 251
11.3 Axe 3 - Frontières cognitives face au discours d'un robot 253
11.3.1 Un robot sans bras et jambes parlant d'actions impossibles : validation du paradigme 253
11.3.2 Un robot parlant de ses émotions : une limite pour l'humain 254
11.3.3 Impact sur la N400 des croyances à propos des capacité du robot 255

12 Limites et perspectives 257
12.1 Échantillons 257
12.2 Validité écologique des protocoles expérimentaux et des mesures 258

12.3 Diversité limitée des plateformes robotiques étudiées 259

13 Implications **261**
13.1 Implications conceptuelles et méthodologiques 261
13.2 Implications pour la conception des interactions avec les robots sociaux 263
13.3 Implications éthiques 265

Conclusion **268**

Bibliographie **271**

Annexes **295**
A.1 Chapitre 7 : Script du robot selon le style de communication 297
A.1.1 Condition Autoritaire 297
A.1.2 Condition Soumis 298
A.1.3 Condition Enfantin 299
A.1.4 Condition Neutre (et Rideau) 299
A.1.5 En commun 300
A.2 Chapitre 7 : Liste des questions posées par l'humain au robot . . . 301
A.3 Chapitre 7 : Tableaux additionnels 307
A.4 Résultats de la manipulation de contrôle 309
A.5 Chapitre 8 : Discours du robot selon la condition 313
A.5.1 Condition Autoritaire 313
A.5.2 Condition Soumis 313
A.5.3 Condition Enfantin 314
A.5.4 Condition Neutre 314
A.6 Chapitre 8 : Liste des questions posées par l'humain au robot . . . 314
A.7 Chapitre 8 : Figures et tableaux additionnels 316
B.1 Annexe Chapitre 9 : Phrases prononcées par le robot (actions) . . . 318
B.2 Annexe Chapitre 10 : Phrases prononcées par le robot (émotions) . 323

Table des figures

5.1 Aperçu de l'interface pour différentes phases 101
5.2 Exemples de trajectoires d'agents sur un essai 104

6.1 Scores aux deux dimensions de présence sociale investiguées . . . 118
6.2 Nombre de croisements selon l'agent et l'Étude 121
6.3 Vélocité relative : Exemples de tracés DFA (log-log) pour la vélocité relative selon le type d'agent rencontré 124
6.4 Vélocité relative : Exposant d'échelle α selon l'instruction 125
6.5 Vélocité relative : Pente spectrale β selon l'instruction 127
6.6 Vélocité relative : Largeur du spectre multifractal Δh selon l'instruction 129
6.7 Vélocité des agents : Exposant d'échelle α selon l'instruction 131
6.8 Vélocité des agents : Pente spectrale β selon l'instruction 132
6.9 Vélocité des agents : Largeur du spectre multifractal Δh selon l'instruction 132
6.10 Vélocité des participants : Exposant d'échelle α selon l'instruction et le partenaire 134
6.11 Vélocité des participants : Pente spectrale β selon l'instruction et le partenaire 134
6.12 Vélocité des participants : Largeur du spectre multifractal Δh selon l'instruction et le partenaire 135

7.1 Aperçu des interfaces, vidéos et visages du robot 157
7.2 Fonction psychométrique globale pour les jugements des délais de réponse 165
7.3 Courbes psychométriques issues du GLMM des cinq conditions . . 166
7.4 Scores selon la condition aux dimensions de Ho et MacDorman (2017) 169

8.1 Scores selon le style de communication aux dimensions de Ho et MacDorman (2017) . . . 180
8.2 Scores selon le style de communication au questionnaire du modèle Almere (Heerink et al. 2010) . . . 184

9.1 Captures d'écran des vidéos des conditions HEAD et BODY . . . 199
9.2 Électrodes d'intérêt et configuration de salle d'expérimentation . . 204
9.3 Schéma d'une séquence pour un essai de l'expérience . . . 206
9.4 Tracés des potentiels évoqués des deux conditions aux électrodes Cz et Pz . . . 209
9.5 Amplitudes moyennes pour les treize électrodes d'intérêt . . . 211
9.6 Scores aux cinq affirmations explorant les perceptions à l'égard du robot . . . 213
9.7 Scores au questionnaire de Ho et MacDorman (2010) selon la condition . . . 214
9.8 Amplitudes moyennes des ERP pour les treize électrodes d'intérêt selon la croyance . . . 216
9.9 Scores pour les cinq traits explorant les perceptions à l'égard du robot selon la croyance . . . 218
9.10 Scores selon le sous-groupe de croyance aux dimensions du questionnaire de Ho et MacDorman (2010) . . . 219

10.1 Tracés des potentiels évoqués des deux conditions aux électrodes Cz et Pz . . . 227
10.2 Amplitudes moyennes pour les treize électrodes d'intérêt . . . 228
10.3 Scores pour les cinq affirmations explorant les perceptions à l'égard des agents . . . 229
10.4 Scores au questionnaire de Ho et MacDorman (2010) selon la condition . . . 230
10.5 Amplitudes moyennes des ERP pour les treize électrodes d'intérêt selon la croyance . . . 232
10.6 Scores pour les cinq affirmations explorant les perceptions à l'égard du robot selon la croyance . . . 233
10.7 Scores selon le sous-groupe de croyance aux dimensions du questionnaire de Ho et MacDorman (2010) . . . 234

Liste des tableaux

6.1 Moyennes marginales estimées de la coprésence selon l'agent . . . 119
6.2 Moyennes marginales estimées de l'interdépendance comportementale perçue selon l'étude . . . 119
6.3 Nombre moyen estimé de croisements par essai selon l'agent et l'étude . . . 121
6.4 Vélocité relative : Exposant d'échelle α selon l'agent et l'étude . . . 126
6.5 Vélocité relative : Exposant spectral β selon l'agent et l'étude . . . 127
6.6 Vélocité relative : Largeur du spectre multifractal (Δh) selon l'agent 128
6.7 Vélocité de l'agent : Exposant α . . . 130
6.8 Vélocité de l'agent : Exposant spectral β . . . 130
6.9 Vélocité de l'agent : Largeur du spectre multifractal Δh . . . 131
6.10 Vélocité relative au bloc 1 : Exposant α . . . 136
6.11 Vélocité relative au bloc 1 : Exposant β . . . 137

7.1 Point d'Égalité Subjective moyen selon la condition . . . 164
7.2 Points d'Égalité Subjective estimés à partir du GLMM selon la condition . . . 167
7.3 Moyennes des scores aux dimensions de Ho et MacDorman (2017) par condition . . . 170

8.1 Moyennes des scores selon le style de communication et le délai de réponse aux dimensions de Ho et MacDorman (2017) . . . 179
8.2 Moyennes des scores selon le style de communication et le délai au questionnaire de Almere (Heerink et al. 2010) . . . 183

9.1 Caractéristiques des mots cibles dans l'Étude 1 : mesures linguistiques et phonologiques . . . 202
9.2 Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon la condition . . . 213
9.3 Scores relatifs à la perception de membres cachés . . . 214

9.4 Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon le sous-groupe de croyance 218

10.1 Caractéristiques des mots cibles dans l'Étude 2 : mesures linguistiques et phonologiques 225
10.2 Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon l'agent 230
10.3 Scores relatifs aux capacités émotionnelles perçues 231
10.4 Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon le sous-groupe de croyance 235

Liste des abréviations

— **ASA** : agent social artificiel
— **CASA** : *Computers Are Social Actors*
— **DFA** : analyse des fluctuations redressées (Detrended Fluctuation Analysis)
— **EEG** : électroencéphalographie
— **EMM** : moyennes marginales estimées
— **ERP** : potentiels évoqués (*Event-Related Potential*)
— **GLMM** : modèles linéaires généralisés à effets mixtes
— **HADD** : *Hyperactive/hypersensitive Agent/Agency Detection Device*
— **HRI** : interaction humain-robot
— **ICC** : coefficient de corrélation intraclasse
— **IRMf** : imagerie par résonance magnétique fonctionnelle
— **IST** : *Intentional Stance Test*
— **LM** : modèle linéaire
— **LMM** : modèle linéaire mixte
— **MDFA** : analyse multifractale des fluctuations redressées
— **MSE** : erreur quadratique moyenne des résidus
— **mPFC** : cortex préfrontal médian
— **aPCC** : aire paracingulaire antérieure
— **PSE** : point d'égalité subjective (Point of Subjective Equality)
— **RR** : rapports de taux
— **SEEK** : *Sociality, Effectance, and Elicited agent Knowledge*
— **STS** : sillon temporal supérieur
— **ToM** : théorie de l'esprit (*Theory of Mind*)
— **TPJ** : jonction temporo-pariétale
— **UP** : Point d'unicité

Introduction Générale

Introduction Générale

Lorsque j'étais étudiant en Licence de Psychologie, j'attendais avec excitation la commercialisation du robot compagnon *Buddy* (*Blue Frog Robotics*). À mes yeux, ce petit robot sur roues, sans bras ni jambes, représentait l'incarnation idéale d'une nouvelle génération de robots sociaux et émotionnels, semblant promettre une interaction naturelle, une présence amicale au quotidien et une capacité à jouer tout un éventail d'émotions. Je m'imaginais un agent autonome capable de converser et de co-exister naturellement avec les humains, réactif, s'adaptant à mes humeurs et donnant l'illusion d'être véritablement présent plutôt que d'être une simple statue technologique. La réalité de la première rencontre fut nettement moins idyllique. Mon premier contact avec *Buddy* eut lieu lors de la conférence *Embodied and Situated Language Processing* à Tourcoing en 2022. Durant cette conférence, *Buddy* trônait au centre de l'attention : contrôlé manuellement selon le principe du Magicien d'Oz, il avait été utilisé pour ponctuer le rythme des présentations. Bien que *Buddy* fût mignon, il avait une voix aiguë très agaçante et présentait une latence très importante dans ses réponses : chaque phrase prononcée semblait suspendue dans un temps dilaté. Seul face à lui pendant une pause, je découvris une coquille vide : un visage ponctuellement animé sur un écran, une absence de suivi du regard ou de mouvement de tête convaincant, et aucune autonomie réelle. « *Il est nul, ce robot !* », me suis-je même surpris à penser. Cette déception personnelle reflète un paradoxe plus large qui traverse notre société en pleine transformation technologique.

En dépit de la sophistication croissante des machines et de la prolifération apparemment triomphante des robots, de nombreux échecs persistent. Les robots sociaux envahissent, en effet, nos environnements et sont déjà utilisés dans une diversité de contextes. Ils servent notamment d'outils pour la recherche, d'aides en thérapie, de jouets ou encore d'outils éducatifs (Fong et al., 2003). Par exemple, le robot phoque *PARO* utilisé pour la prise en charge non pharmacologique de la démence (Shibata & Wada, 2011) comptait en 2021, plus de

6000 unités déployées dans plus de 30 pays (Paro, 2021). Le célèbre robot humanoïde *Pepper*, a été produit à 27000 exemplaires (Nussey, 2021) et est utilisé aussi bien dans les services d'hôtellerie (Tuomi et al., 2021) que dans des contextes d'apprentissage, par exemple pour permettre à des enfants de développer leurs connaissances sur leur diabète (Zarubica & Bendel, 2024). Le robot *Nao*, avec plus de 19 000 unités déployées dans plus de soixante-dix pays (Aldebaran, 2024) a, quant à lui, été utilisé de nombreuses fois pour soutenir le développement des compétences communicationnelles d'enfant avec troubles du spectre de l'autisme (par exemple, David et al., 2020; Warren et al., 2015; Zhang et al., 2019). Plus récemment, le robot *Miroki* accompagne les enfants lors de séances de radiothérapie pédiatrique (Monceaux, 2024). Pourtant, sur le plan commercial, des entreprises ou des start-up prometteuses s'effondrent : *Aldebaran*, pionnier français de la robotique sociale depuis 2005 et créateur des célèbres robots *Nao* et *Pepper*, a été placée en liquidation judiciaire en juin 2025 après des années de difficultés financières (BODDAC, 2025). La start-up de *Jibo*, salué comme « le premier robot social familial » par les médias, a cessé ses activités deux ans après sa commercialisation (Ackerman, 2018), malgré plusieurs tentatives de relance par de nouveaux acquéreurs (Bartneck, 2024). Enfin, *Anki* (connu pour les robots *Vector* et *Cozmo*) a connu un parcours similaire et a fermé en 2019, puis a été racheté par *Digital Dream Labs*, mais rencontre encore actuellement des difficultés persistantes en matière de commercialisation et de support technique (Clark, 2024).

Au-delà des échecs commerciaux, il semble également que notre cohabitation avec les robots n'est et ne sera pas toujours harmonieuse. Bien qu'ils soient explicitement conçus pour activer nos mécanismes de perception sociale, par le biais de visages expressifs, de voix modulées et de comportements imitant les codes de l'interaction humaine, les interactions humain-robot révèlent parfois une hostilité inattendue. Par exemple, une étude documente des incidents durant lesquels des enfants bloquaient ou agressaient verbalement et physiquement un robot humanoïde dans un centre commercial au Japon (Nomura et al., 2016). En 2015, aux États-Unis, le robot auto-stoppeur *HitchBOT*, qui avait traversé le Canada et plusieurs pays européens grâce à la bienveillance des personnes rencontrées, fut finalement vandalisé et endommagé (Smith & Zeller, 2017). Aussi, bien que les humains perçoivent les agressions envers les robots

humanoïdes comme immorales au même titre que celles dirigées contre des humains, les représailles ne sont jugées morales que lorsqu'elles proviennent d'humains et non de robots. En d'autres termes, un robot n'a pas le droit de se défendre face à un humain (Bartneck & Keijsers, 2020).

Ces échecs commerciaux et ces réactions hostiles rejoignent ce que Henschel et al. (2020) décrivent comme un « *social robotics winter* », une métaphore empruntée aux « *AI winters* », qui décrit une phase de désillusion lorsque la technologie échoue à tenir ses promesses initiales. Ici, les autrices la décrivent comme la désillusion actuelle autour des robots sociaux, les progrès technologiques n'ayant pas été à la hauteur des espoirs et des attentes suscités par les représentations des robots dans les films, à la télévision et dans d'autres médias. Henschel et al. (2020), argumentent que les méthodes des neurosciences cognitives pourraient offrir des perspectives pour comprendre le côté humain de l'interaction humain-robot et ainsi éviter cette impasse.

Ces observations d'échecs commerciaux, de réactions hostiles, d'asymétries morales, préférences paradoxales suggèrent que nos interactions avec les robots sociaux ne suivent pas les modèles initialement anticipés. En effet, comme nous l'aborderons dans ce travail de thèse, nous semblons établir des frontières cognitives spécifiques dans nos rapports aux agents artificiels, frontières qui ne correspondent pas aux catégories binaires (humain/non-humain) des théories classiques. Les cadres classiques dont les grilles de lecture sont héritées de l'interaction humain-humain et des médias interactifs (par exemple, *Media Equation* de Reeves et Nass, 1996) ne suffisent pas toujours, voire même semblent inadéquats pour rendre compte de ce que considère, tolère ou refuse l'humain lorsqu'un artefact adopte des codes sociaux. Ces observations soulèvent plusieurs questions : Comment les humains perçoivent-ils et se connectent-ils aux agents sociaux artificiels comme les robots sociaux ? Quelles sont les spécificités et frontières que nous établissons dans ces interactions par rapport aux relations humain-humain ?

Cette thèse a ainsi pour objectif d'explorer comment les humains réagissent et établissent (ou non) des liens avec des agents sociaux artificiels. Elle adopte une approche résolument interdisciplinaire mobilisant les ressources de la psychologie cognitive, de la robotique sociale et des neurosciences. Elle met également en œuvre une triangulation méthodologique (Denzin, 1970) combinant

méthodes comportementales, psychophysiques et neurophysiologiques. Cette triangulation permet une validation croisée des résultats et une compréhension plus robuste des mécanismes de perception sociale d'agents artificiels. Une telle approche est particulièrement pertinente dans l'étude des interactions humain-robot où les mécanismes peuvent différer de ceux observés en interaction humain-humain (Bethel & Murphy, 2010).

Le manuscrit s'articule en quatre parties. La première établit le cadre théorique et conceptuel nécessaire à la compréhension des mécanismes de perception sociale appliqués aux agents artificiels. Les trois parties suivantes développent autant d'axes d'investigation complémentaires, chacun associé à une contribution empirique.

Le premier axe explore les mécanismes d'émergence de la présence sociale avec des agents artificiels dans un environnement minimaliste. En adaptant l'expérience du *Perceptual Crossing* proposée par Auvray et al. (2009) à l'interaction humain-agent seulement, il s'agit de déterminer si une simple coordination sensorimotrice suffit à faire émerger un sentiment de présence sociale en l'absence d'incitation sociale (via une instruction donnée au participant) explicite, et dans quelle mesure la contingence du comportement de l'agent aux croisements avec le participant influence cette expérience. L'objectif est également de déterminer si l'instruction explicite d'engager une interaction modifie les comportements d'exploration et le sentiment de présence sociale suscité par les agents, ainsi que d'explorer comment les propriétés comportementales des agents affectent cette présence sociale.

Le deuxième axe s'intéresse aux attentes temporelles dans l'échange verbal et les tours de parole avec un robot social, et à leur modulation par les styles de communication. À travers deux études complémentaires, elle vise d'abord à identifier le délai de réponse perçu comme optimal dans une conversation, puis à analyser l'effet du style de communication du robot sur la tolérance aux écarts temporels (par exemple, l'humain sera-t-il plus tolérant face à un robot au style enfantin ou soumis ?). Enfin, l'impact de délais fortement déviants de l'optimum, combinés aux styles de communication, est évalué sur l'appréciation globale du robot.

Enfin, le troisième axe propose un paradigme pour examiner les frontières de l'acceptabilité du contenu des discours lorsqu'ils sont énoncés par des robots.

Deux études mobilisant l'électroencéphalographie (EEG, via la composante N400) analysent la manière dont les humains traitent les énoncés des robots sociaux lorsqu'ils dépassent les limites de ce qui leur est accessible et les limites des thématiques habituellement réservées aux humains ou aux animaux, telles que les émotions.

Ces trois axes convergent vers des objectifs complémentaires : fournir des indicateurs utiles à la conception d'agents sociaux artificiels, proposer des paradigmes transférables à d'autres recherches et documenter les spécificités cognitives de l'interaction humain-robot.

Première partie

Cadre théorique et conceptuel

Chapitre 1

Agents sociaux artificiels et robots sociaux : définitions, typologies et enjeux

1.1 Définitions des Agents Sociaux Artificiels et Robots Sociaux

Les Agents Sociaux Artificiels (ASA) représentent une catégorie d'entités technologiques conçues pour interagir avec les humains selon des modalités sociales (Fitrianie et al., 2019, 2020). Un agent, dans sa conception la plus générale est (traduction) « *tout ce qui peut être considéré comme percevant son environnement à travers des capteurs et agissant sur cet environnement à travers des actionneurs* » (Russell & Norvig, 2021). Ainsi, cette définition générale peut aussi bien englober des entités biologiques que des systèmes artificiels. Il convient de distinguer la notion générale d'agent de celle d'agent social, car tous les agents artificiels ne sont pas nécessairement sociaux. La distinction réside dans la capacité à s'engager dans une interaction sociale.

Selon Thórisson (1996), les agents sociaux sont des (traduction) : « *agents logiciels, robots ou créatures autonomes qui possèdent du savoir-faire en matière d'interaction sociale* [...] *et qui peuvent donc s'engager dans une interaction sociale avec des personnes à un certain niveau.* »

Les ASA, peuvent être définis comme (traduction) « *des entités contrôlées par ordinateur capables d'interagir de manière autonome avec les humains en suivant les règles sociales des interactions humain-humain* » (Fitrianie et al., 2019). De ce fait, un ASA, en tant qu'entité autonome et intelligente, possède des propriétés de base, des traits sociaux et des capacités lui permettant de jouer son rôle

d'agent social dans une interaction (Fitrianie et al., 2020). C'est-à-dire que ces types d'agents peuvent percevoir leur environnement, prendre des décisions et agir de façon relativement indépendante, tout en étant programmés pour adopter des comportements socialement pertinents vis-à-vis de leurs partenaires humains. Ces définitions soulignent des propriétés essentielles : l'autonomie comportementale, la capacité d'interaction et l'adhésion aux règles sociales.

Il existe néanmoins de nombreux ASA différents, tels que les agents conversationnels, les agents virtuels contrôlés par ordinateur ou encore les robots sociaux (Fitrianie et al., 2019, 2020), qui, eux, sont physiquement incarnés (Naneva et al., 2020). Le terme « *robot* » serait apparu pour la première fois dans une pièce de théâtre de science-fiction de Karel Čapek. C'est son frère, Josef Čapek, qui lui aurait suggéré ce mot (Harkins, 1962). La norme ISO 8373 :2021 fournit une définition générale du robot comme un « *mécanisme programmable actionné avec un degré d'autonomie* [...] *pour effectuer des opérations de locomotion, de manipulation ou de positionnement*. » (ISO, 2021). Un robot social se distingue donc des robots industriels ou purement fonctionnels par sa conception centrée sur l'interaction sociale plutôt que sur l'exécution de tâches physiques et techniques (Breazeal, 2003b).

En conclusion, les robots sociaux constituent un type d'ASA. Dans leur travail de compilation de définitions, Hegel et al. (2009) proposent une définition synthétique (traduction) : « *Un robot social est un robot auquel on a ajouté une interface sociale. L'interface sociale est une métaphore qui regroupe l'ensemble des attributs sociaux par lesquels un observateur juge le robot apte à devenir un partenaire d'interaction sociale.* »

1.2 Taxonomies et classifications

Plusieurs taxonomies ont été proposées pour classer les robots qui sont capables d'interaction sociale ou d'évoquer du social. Dans la suite de cette section, les classifications basées sur la sophistication sociale des comportements du robot dans l'interaction seront d'abord présentées, suivies de celles qui se centrent sur la morphologie et le type d'incarnation (*embodiment*).

En ce qui concerne les classifications basées sur le comportement du robot, on peut citer, par exemple, celle de Breazeal (2003a) qui distingue quatre classes

de robots sociaux organisées selon une progression croissante de leur capacité à soutenir le modèle social. Elles vont d'interactions simples à l'environnement social humain complet :

1. évoquant le social (*socially evocative*), qui se limitent à susciter l'anthropomorphisme et des attributions sociales chez l'humain (par exemple, le jouet). Bien qu'une capacité à répondre socialement sera prêtée au robot, son comportement n'est pas réciproque de façon active.
2. interfaces sociales (*social interface*), qui emploient des indices sociaux pour rendre l'interaction plus intuitive (par exemple, un robot guide dans un musée). Le modèle social est superficiel, voire absent. Le comportement est préprogrammé ou réflexe. Le robot n'apprend pas de l'interaction. Les indices sociaux relèvent surtout de l'interface et ne sont qu'apparence.
3. réceptifs socialement (*socially receptive*), qui apprennent passivement dans l'interaction avec l'humain mais n'initient pas d'engagement proactif.
4. sociables, qui participent socialement et sont proactifs. Ces robots ont leurs propres objectifs et motivations internes pour lesquels l'interaction sociale est constitutive : elle bénéficie aux humains mais surtout à eux-mêmes. Ils perçoivent les indices sociaux et modélisent profondément l'humain au niveau cognitif et social.

Ensuite, Fong et al. (2003) reprennent cette grille mais vont l'étendre avec des classes supplémentaires, en s'appuyant sur les travaux de Dautenhahn (1995, 1998) et Dautenhahn et al. (2002). Trois classes viennent alors préciser le degré d'insertion et de raisonnement social au-delà de la seule interaction de surface :

1. situé socialement (*socially situated*) : le robot est dans un environnement social qu'il perçoit et auquel il réagit. Il est capable de distinguer les agents sociaux et les objets. Cette classe introduit ici l'exigence de perception du contexte social.
2. intégré socialement (*socially embedded*) : le robot est non seulement situé, mais il est aussi couplé à l'environnement social. Il va interagir avec les agents en étant partiellement conscient des structures d'interaction entre les humains, comme le tour de parole (turn-taking).
3. intelligent socialement (*socially intelligent*) : à partir de modèles profonds

des compétences sociales et de la cognition humaine, cette classe de robots présente certains aspects de l'intelligence sociale de type humain.

Sans ajouter de nouvelles classes, ils introduisent le terme de *robot socialement interactif* (*socially interactive robot*) pour désigner tout robot dont l'interaction sociale pair-à-pair avec l'humain est le rôle central. Ceux-ci sont donc distincts des robots contrôlés à distance. Cette étiquette de robot est décrite comme ayant des caractéristiques telles que : exprimer ou percevoir des émotions, dialoguer à haut niveau, apprendre ou reconnaître des modèles d'autrui, établir ou maintenir des relations sociales, utiliser des indices naturels comme le regard et les gestes, présenter une personnalité marquée et éventuellement apprendre/développer des compétences sociales (Fong et al., 2003).

D'autres cadres s'intéressent à la morphologie et aux types d'incarnation des robots sociaux. L'apparence, avant même l'interaction, induit et calibre des attentes différentes dans le registre d'indices sociaux que l'humain anticipe (Fong et al., 2003). Quatre catégories d'incarnation de robots sont décrites par Fong et al. (2003) : *anthropomorphes* (humanoïdes, androïdes), *zoomorphes* (inspirés des animaux), *caricaturaux* (formes simplifiées, proportions exagérées et mouvements inspirés des principes de l'animation de personnages) et *fonctionnels* (une conception dictée par les objectifs et les besoins de la tâche). Bartneck et Forlizzi (2004) élargissent cette perspective morphologique et proposent un cadre de conception à cinq propriétés qui articule la forme du robot avec ses capacités comportementales : forme (classée en abstraite, biomorphique, anthropomorphique), modalités, normes sociales, autonomie et interactivité.

Plus récemment, Baraka et al. (2020) proposent un cadre multidimensionnel pour caractériser les robots sociaux selon sept dimensions : apparence, capacités sociales, objectif d'application, rôle relationnel, autonomie et intelligence, proximité et profil temporel. Cette approche multifactorielle reconnaît que la socialité artificielle émerge de l'interaction complexe entre des facteurs techniques, relationnels et contextuels. Concernant la dimension *apparence* en particulier, ils distinguent trois grandes catégories morphologiques en s'appuyant sur de précédents travaux (tels que Fong et al., 2003 et Shibata, 2004) :

1. les robots *bio-inspirés* (*bio-inspired*), qui reproduisent des traits biologiques humains (humanoïdes, androïdes/gynoïdes, geminoïdes, parties du corps) tels que l'humanoïde *Pepper*, ou qui reproduisent des traits biologiques

animaux (réels et imaginaires, familiers et non familiers, ainsi que des fragments anatomiques) tels que *PARO* que l'on peut classer parmi ceux qui sont *réels non-familiers* ;

2. les robots *de forme d'artefact* (*artifact-shaped*), dont l'apparence est empruntée à des formes purement imaginaire ou à des créations humaines. Ces derniers comprennent les robots *inspirés par l'objet* ou *object-inspired* qui ressembleraient par exemple à une valise, les robots *inspirés par le dispositif* ou *apparatus-inspired* qui ressemblerait par exemple à une voiture) ;
3. et les robots fonctionnels, dont l'aspect résulte de l'assemblage des composants mécaniques nécessaires à la réalisation de leur tâche.

D'autres travaux, comme ceux de Mobed et al. (2024), proposent sept morphologies qui sont : (1) anthropomorphisme, (2) zoomorphisme, (3) phytomorphisme (inspiration végétale), (4) artemorphisme (inspiration par objet ou artefact), (5) functiomorphisme (forme dictée par la fonction), (6) amorphisme (forme dépourvue d'ordre fixe ou évolutive, proche des approches en robotique évolutionnaire) et (7) néomorphisme (forme perçue comme nouvelle ou hybride, combinant plusieurs morphologies). Ces morphologies sont croisées avec trois axes perceptifs (naturel-mécanique, familier-étrange, abstrait-réaliste) pouvant servir de base à la catégorisation et offrant un repère pour situer des robots réels et conceptuels au niveau de la morphologie.

En somme, les différentes taxonomies présentées se structurent autour de deux axes principaux pour caractériser les robots sociaux : d'une part, le degré de sophistication des capacités comportementales et sociales (de la simple évocation sociale à des formes d'« intelligence sociale »), et d'autre part, la morphologie et le type d'incarnation (du fonctionnel au bio-inspiré). Plusieurs auteurs soulignent que ces dimensions interagissent : l'apparence calibre les attentes sociales (Fong et al., 2003) et doit correspondre aux capacités réelles ainsi qu'à la tâche pour éviter de fausses attentes (Bartneck & Forlizzi, 2004).

1.3 Enjeux, règles sociales et normes culturelles

Un agent social doit ajuster ses conduites aux règles sociales liées à son rôle et à son contexte. Ces règles sont culturellement situées : attentes, normes et scripts interactionnels varient selon les collectifs, ce qui impose une conception et une évaluation sensibles aux cultures (Lee & Šabanović, 2014 ; Šabanović et al., 2014).

Des différences culturelles robustes sont documentées dans la perception et l'acceptation des robots. Nomura et al. (2008) ont interrogé des étudiants japonais, coréens et américains sur cinq dimensions de perception appliquées à des robots animaux et humanoïdes : autonomie, relation sociale, émotion, rôles et image projetée. Leurs résultats montrent que les Japonais attribuent significativement plus de caractéristiques humaines aux robots humanoïdes et les associent à des fonctions communicationnelles, tandis que les Coréens sont plus méfiants quant aux effets sociaux de ces robots, tout en les jugeant plus appropriés pour des rôles médicaux. Les Américains, eux, présentent des jugements plus contrastés, sans préférence nette pour des fonctions particulières. Contrairement aux idées reçues selon lesquelles les Japonais (ou l'Asie en général) seraient particulièrement enthousiastes à l'égard des robots, Bartneck et al. (2005) n'ont pas observé d'attitudes globalement plus positives chez les Japonais que chez les Chinois ou les Néerlandais. En revanche, les résultats montrent une inquiétude significativement plus importante quant aux effets des robots sur la société chez les Japonais, comparativement aux autres groupes interrogés (Bartneck et al., 2005).

Dans le même objectif, Bartneck et al. (2006) ont comparé les attitudes de participants issus de sept pays face au robot *Aibo* (Sony). Les résultats montrent encore que la culture module fortement les attitudes envers les robots : les participants américains étaient les moins négatifs et exprimaient une plus grande ouverture à l'interaction, tandis que les Japonais se montraient plus ambivalents et manifestaient davantage de préoccupations sociales et émotionnelles. Les participants mexicains présentaient les attitudes les plus négatives, notamment vis-à-vis de l'interaction directe.

Des travaux plus récents confirment, encore aujourd'hui, l'existence de différences culturelles dans la perception des robots. Castelo et Sarvary (2022)

ont démontré que l'augmentation du réalisme physique des robots diminue le confort des participants américains, mais pas celui des participants japonais. En revanche, attribuer des capacités émotionnelles à un robot réduit le confort des Américains mais l'augmente chez les Japonais. Les auteurs suggèrent que, comparativement aux participants américains, les participants japonais perçoivent les robots comme plus animés et dotés d'un esprit ou d'une conscience.

D'autres différences culturelles apparaissent notamment en ce qui concerne la manière dont le robot communique. Au sujet de la crédibilité du discours du robot, Rau et al. (2009) ont démontré que des participants chinois préféraient des robots qui communiquent leurs opinions de manière implicite. Ils les évaluaient alors comme plus crédibles, dignes de confiance et sympathiques. En comparaison, les participants allemands acceptaient moins les recommandations implicites et jugeaient les robots plus négativement sur ces mêmes dimensions. Andrist et al. (2015) ont, quant à eux, montré que la crédibilité accordée au discours de robots-guides varie aussi selon la culture. En manipulant le niveau de connaissances factuelles du robot et la qualité rhétorique de son discours (par exemple, l'organisation du propos ou l'usage de métaphores), ils ont observé que les participants libanais (arabophones) étaient significativement plus sensibles à la rhétorique, tandis que les participants américains (anglophones) préféraient un robot très informé mais peu expressif sur le plan rhétorique. De manière générale, les scores de crédibilité perçue augmentaient avec le niveau de connaissances factuelles du robot et étaient globalement plus élevés chez les participants américains.

De même, l'adaptation culturelle doit prendre en compte non seulement les aspects du discours, mais aussi les codes gestuels propres à chaque culture. À cet égard, Trovato et al. (2013) ont étudié les réactions de participants égyptiens et japonais face à deux robots effectuant des salutations culturellement distinctes : une salutation « à la japonaise » (inclinaison, parole en japonais puis nouvelle inclinaison accompagnée d'une autre parole en japonais) et une salutation version « arabe » (lever la main, parole en arabe, puis main sur le cœur). Dans les deux groupes, les résultats ont révélé une préférence marquée pour le robot correspondant à leur propre culture. Les participants égyptiens préféraient significativement le robot « arabe » et manifestaient des signes d'inconfort (temps de réaction prolongés, expressions faciales négatives) face au robot

« japonais ». Le même schéma a été observé chez les participants japonais, bien que sans allongement significatif des temps de réaction.

Ces différentes observations convergent vers une conclusion claire : la conception d'ASA ne peut reposer sur une approche totalement universaliste et doit prendre en compte les différences culturelles. Ainsi, Šabanović et al. (2014) préconisent une approche de robots « culturellement robustes » fondée sur deux principes complémentaires (traduction) : « *(1) rendre le processus de conception plus réflexif sur le plan culturel et intégrant les points de vue des différentes parties prenantes, et (2) concevoir des robots capables d'être sensibles et adaptables aux facteurs culturels saillants, plutôt que de concevoir des robots pour des cultures spécifiques* ».

En pratique, l'adaptation au contexte d'utilisation peut être efficacement assurée en impliquant directement les usagers locaux dans la conception des interactions, notamment au moyen d'approches participatives. Sun (2012) propose le cadre *CLUE* (Culturally Localized User Experience) qui traite la culture comme une entité flexible plutôt que comme un ensemble de dimensions fixes, et préconise l'implication directe des utilisateurs locaux dans le processus de conception afin d'identifier les besoins spécifiques au contexte d'usage.

Un autre exemple illustrant l'approche d'adaptation contextuelle est le projet *CARESSES* (Culture-Aware Robots and Environmental Sensor Systems for Elderly Support), dans le cadre duquel un système et une architecture logicielle sensibles à la culture ont été développés et intégrés au robot *Pepper*. Ce système incorpore une base de connaissances culturelles (*Cultural Knowledge Base*), permettant au robot d'adapter ses interactions verbales et non-verbales selon l'identité culturelle de l'utilisateur, tout en apprenant progressivement ses préférences individuelles (Papadopoulos et al., 2020). Ce système a démontré qu'il pouvait améliorer le bien-être émotionnel des personnes âgées résidant en maisons de retraite et qu'il avait au moins les mêmes effets que les soins habituels sur la solitude ou la santé physique (Papadopoulos et al., 2022).

En conclusion, cette section a mis en évidence que les comportements et les attentes envers les ASA, ainsi que les rôles sociaux qui leur sont attribués, dépendent du contexte culturel. Les différences concernent aussi la communication du robot, sur des aspects tels que la crédibilité perçue, l'expression d'opinions de façon implicite ou explicite, ou l'utilisation d'une rhétorique per-

suasive. L'ensemble de ces résultats montre que la conception de robots sociaux ne peut être universaliste : elle doit être culturellement sensible et adaptative, fondée sur l'intégration de connaissances culturelles et sur l'apprentissage des préférences individuelles.

Si les variations culturelles façonnent les attentes et les comportements envers les ASA, il existe tout de même des mécanismes cognitifs, relativement universels, qui déterminent la manière dont les humains perçoivent et interprètent ces agents. Le chapitre suivant examine ces processus, tels que l'anthropomorphisme et l'attribution d'intentionnalité, qui sous-tendent notre capacité à reconnaître et à interagir avec des entités artificielles, qu'elles soient présentées comme sociales ou non.

Chapitre 2

Mécanismes de perceptions et d'interprétation appliqués aux agents artificiels

2.1 L'Anthropomorphisme

Ce chapitre porte sur la manière dont se manifestent, face aux agents artificiels, les mécanismes que l'humain mobilise habituellement pour percevoir et interpréter son environnement, ainsi que pour comprendre le comportement d'autrui. Il y sera montré comment ces processus, qui permettent de détecter, d'attribuer du sens et d'inférer des intentions, peuvent également s'activer face à des agents artificiels tels que les robots. Seront également abordées notre tendance à projeter des caractéristiques humaines sur le non-humain (anthropomorphisme), ainsi que plusieurs autres cadres théoriques, tels que la *Posture Intentionnelle*, qui consiste à attribuer des états mentaux afin de prédire et d'expliquer le comportement d'un système, la *perception de l'esprit*, et des théories comme la *Media Equation*, permettant d'explorer les réponses sociales automatiques.

2.1.1 Du biais perceptif à la projection de l'humain

L'être humain manifeste une tendance remarquable à détecter des agents dans son environnement, même là où il n'y en a pas. Avant même l'identification d'un agent à proprement parler, ce biais se manifeste à travers la paréidolie, c'est-à-dire la tendance à percevoir des formes, des motifs ou des objets avec un sens à partir de stimuli ambigus (Flessert, 2022). Parmi ce type de perception illusoire, la paréidolie des visages, dans laquelle des visages ou des motifs

ressemblant à des visages sont reconnus en l'absence de visages réels (Liu et al., 2014), illustre bien la prédisposition perceptive à voir de l'humain (par exemple, des visages) dans des stimuli non humains.

Sur le plan adaptatif, la propension humaine à détecter des agents pourrait trouver ses racines dans des mécanismes évolutifs anciens. Barrett (2000, 2004) fait l'hypothèse d'un dispositif hypersensible de détection d'agents intentionnels (*HADD, Hyperactive/hypersensitive Agent/Agency Detection Device*) qui propose que le système cognitif humain aurait évolué de manière à privilégier les faux positifs plutôt que les faux négatifs plus coûteux : il vaudrait mieux détecter un agent inexistant que manquer un prédateur réel. Des données récentes nuancent toutefois le *HADD* en montrant qu'il n'est pas hypersensible par défaut, puisque des menaces faibles ou modérées n'augmentent pas le taux de faux positifs liés à la détection d'agents ou d'intentionnalité (Maij et al., 2019). Plus largement, Guthrie (1995) suggère que, face aux incertitudes et aux ambiguïtés du monde, le système perceptif humain privilégie un schéma humain pour les interpréter, car il parie prioritairement sur les possibilités les plus importantes : la présence d'êtres vivants, et surtout de ses semblables.

Cette tendance à « projeter de l'humain » afin de faciliter la compréhension et l'interaction avec le monde environnant est appelée anthropomorphisme (Epley et al., 2007 ; Guthrie, 1995). Cette inclination a été largement discutée dans des contextes religieux et philosophiques, où attribuer des caractéristiques humaines à des forces invisibles ou à des entités abstraites constituait une manière privilégiée de rendre le monde intelligible (Duffy, 2003 ; Guthrie, 1995).

2.1.2 Motivations et mécanismes de l'anthropomorphisme : le modèle SEEK

Selon Epley et al. (2007), le terme *anthropomorphisme*, issu du grec *anthropos* (signifiant « humain ») et morphḗ (signifiant « forme »), n'est pas juste de l'animisme, qui attribue la vie à ce qui est inanimé. L'anthropomorphisme désigne plus spécifiquement la tendance de l'être humain à (traduction) : « *attribuer au comportement réel ou imaginé d'agents non-humains, des caractéristiques, motivations, intentions et des émotions semblables à l'humain* ». En d'autres termes, il ne s'agit pas d'un rapport descriptif de comportements observables (ou ima-

ginés) mais d'un processus d'inférence à propos de caractéristiques non observables d'un agent non-humain, ce qui implique l'attribution d'états mentaux à celui-ci. Par exemple, dire « *le chien est affectueux* » revient à décrire des signes observables de comportement (montrer des signes d'affection), tandis que dire « *le chien m'aime* » implique une attribution de sentiment, c'est-à-dire la projection d'une expérience émotionnelle, d'un état mental, comparables à ceux des humains. L'anthropomorphisme opère donc un passage de la simple description à l'inférence d'états mentaux.

Avec le modèle SEEK (pour *Sociality*, *Effectance* et *Elicited agent Knowledge*), Epley et al. (2007) proposent une théorie pour expliquer dans quelles conditions les individus ont tendance à anthropomorphiser ou, au contraire, à ne pas le faire. Ce modèle considère que l'anthropomorphisme repose sur les mêmes mécanismes cognitifs que toute autre inférence, c'est-à-dire l'acquisition, l'activation et l'application de connaissances à une cible donnée, complétées par des tentatives d'ajustements ou de corrections destinées à affiner l'attribution. Ce modèle identifie trois déterminants (facteurs) psychologiques qui interagissent pour moduler l'inférence anthropomorphique : les *Connaissances élicitées* par l'agent (*Elicited agent knowledge*), la *Motivation d'Effectance* (*Effectance motivation*) et la *Motivation à la socialité* (*Sociality motivation*).

Concernant les *Connaissances élicitées* par l'agent, ce facteur stipule que les connaissances que nous possédons sur les humains en général, ou sur nous-mêmes en particulier, servent par défaut de base inductive pour interpréter les agents non humains. La similarité perçue entre l'agent non humain et l'humain module l'accessibilité et l'application de ces connaissances : des indices morphologiques ou de mouvement « humains » renforcent l'activation de schémas fondés sur soi ou sur l'humain. Cette base anthropocentrée s'active comme ancrage par défaut, la connaissance de soi et des humains étant plus accessible, acquise plus tôt et plus richement que les connaissances alternatives. Cependant, ce biais s'atténue à mesure que nous acquérons des connaissances spécifiques sur des entités non humaines.

Ensuite, la *Motivation d'Effectance* renvoie au besoin d'interagir efficacement avec l'environnement, de comprendre, de contrôler et de prédire les systèmes complexes. Cette motivation pousse à recoder le non-humain en termes humains. Face à un système complexe ou imprévisible, l'anthropomorphisme offre

un raccourci interprétatif qui permet de réduire l'incertitude et de rendre le non-humain plus intelligible et maîtrisable. Cependant, lorsque des connaissances non anthropocentriques pertinentes sont disponibles, la *Motivation d'Effectance* tend à les privilégier, ceci pouvant réduire l'anthropomorphisme. À l'inverse, lorsque le comportement d'un agent est perçu comme trop imprévisible ou aléatoire, l'attribution d'un esprit ou d'intentions ne facilite plus la prédiction ni le contrôle : la cible est alors jugée comme dépourvue d'esprit (*« mindless »*).

Enfin, la *Motivation à la socialité* répond au besoin inné fondamental d'établir et de maintenir des connexions sociales. Elle augmente l'accessibilité des indices sociaux et des traits humains, ce qui favorise la perception de « l'humanité » chez des non-humains. Sous privation sociale, elle déclenche également une recherche active de sources de connexion sociale qui va créer des liens de substitution en « fabriquant » du social dans des entités ambiguës.

2.1.3 Anthropomorphisme dans l'interaction humain-robot

Dans le contexte spécifique de l'interaction humain-robot (HRI), Kühne et Peter (2023) ont proposé et développé une conceptualisation multidimensionnelle de l'anthropomorphisme spécifiquement adaptée aux interactions humain-robot. Dans leur conceptualisation, l'anthropomorphisme est considéré comme une forme de cognition humaine qui se concentre sur l'attribution de capacités mentales humaines à un robot. Leur conception distingue (1) les précurseurs de l'anthropomorphisme, (2) le processus lui-même d'anthropomorphisme ainsi que (3) les conséquences de l'anthropomorphisme en HRI. Chacun de ces trois éléments sera détaillé ci-dessous :

1. En amont de tout anthropomorphisme et avant toute inférence d'états mentaux, les précurseurs correspondent aux perceptions liées à des caractéristiques observables du robot telles que les perceptions liées à la forme, au comportement et aux mouvements du robot. Par exemple, les précurseurs perceptifs qui favorisent l'anthropomorphisme peuvent être de simples traits faciaux (par exemple, des yeux, une bouche ou la largeur de tête) qui vont augmenter « l'humanité » perçue d'une tête de robot (DiSalvo et al., 2002). De même, le fait percevoir un mouvement à une vitesse comparable à celle d'un humain accroît l'attribution d'esprit (Morewedge

et al., 2007). L'anthropomorphisme peut être suscité par un large éventail de robots, allant des aspirateurs robots (une conception inspirée par objet) aux conceptions basées organismes (« organism-based ») et humanoïdes (Tan et al., 2018).

2. Ensuite, concernant l'anthropomorphisme lui-même, Kühne et Peter (2023) le définissent comme le fait d'attribuer au robot des capacités mentales humaines inobservables. Selon leur conceptualisation, les dimensions formant son cœur sont structurées en cinq facettes interdépendantes que l'on attribue au robot : la capacité à penser (« *thinking* », les processus cognitifs de haut niveau), à ressentir (« *feeling* », avoir des expériences subjectives), à percevoir (« *perceiving* », accès et interprétation de stimuli externes), à désirer (« *desiring* », des besoins et des préférences qu'il souhaite satisfaire) et à choisir (« *choosing* », choisir librement entre plusieurs lignes de conduite). Chacune peut être évaluée séparément pour étudier comment les utilisateurs perçoivent et interagissent avec des robots aux caractéristiques variées. Afin d'articuler le phénomène et ce qui s'ensuit (c'est-à-dire les conséquences), Kühne et Peter (2023) s'appuient explicitement sur la *Posture Intentionnelle* de Dennett (1987 ; la section 2.2.1 est dédiée à celle-ci) car pour ceux-ci, elle « *correspond en substance à l'anthropomorphisme* » puisque cette posture consiste à traiter l'entité comme un agent rationnel pour prédire son comportement.
3. Enfin, une fois le robot perçu comme doté de capacités mentales, deux types de conséquences internes émergent : l'attribution d'une personnalité et celle d'une valeur morale. La littérature mobilisée par les auteurs montre que, pour qu'une valeur morale et une responsabilité morale soient attribuées à un agent, celui ci doit être perçu comme possédant à la fois de l'agentivité (l'intentionnalité et l'autonomie dans la prise de la décision sont inclus dans ce concept) et de l'expérience (c'est-à dire pouvoir faire l'expérience de la joie et de la douleur). Dès lors, anthropomorphiser augmente la probabilité de considérer le robot comme agent moral.

Après avoir discuté des conceptualisations de l'anthropomorphisme selon Epley et al. (2007) et Kühne et Peter (2023), il convient d'évoquer les impacts de l'anthropomorphisme, hors de ces cadrages. Tout d'abord, une apparence anthropomorphique peut constituer un levier puissant pour faciliter l'interac-

tion humain-robot (Duffy, 2003). Néanmoins, le recours à une conception anthropomorphique doit être manié avec précaution, car un usage excessif ou maladroit peut conduire à des attentes trompeuses (Duffy, 2003). Une méta-analyse de Roesler et al. (2021) rapporte un effet modéré des caractéristiques anthropomorphiques intégrées à la conception des robots sur les interactions humain-robot. Les robots conçus avec des traits humains (par exemple, par le biais de l'apparence, de la communication ou des mouvements) génèrent des perceptions plus favorables chez les utilisateurs : ces robots sont jugés plus sympathiques et intelligents, suscitent davantage de confiance et d'acceptation et favorisent l'engagement social (Roesler et al., 2021). Toutefois, ces caractéristiques de conception anthropomorphique n'améliorent pas la sécurité perçue, l'empathie ressentie envers le robot ou la performance lors de tâches collaboratives (Roesler et al., 2021). L'efficacité de ces éléments de conception varie aussi selon le contexte d'application : les bénéfices sont constants dans les domaines sociaux (par exemple, la thérapie, l'éducation ou le divertissement), inexistants dans les services et variables dans l'industrie (Roesler et al., 2021). Cette hétérogénéité suggère que l'intégration de traits humains dans la conception robotique doit être adaptée aux objectifs spécifiques de l'interaction plutôt qu'appliquée de manière universelle. Par exemple, dans certains contextes spécifiques, une apparence trop humaine peut être contre productive : Kumazaki et al. (2019) ont observé que chez des personnes autistes, la motivation à suivre un entretien sera plus grande avec un robot d'apparence humanoïde sans trop l'être (c'est-à-dire en restant simple) et que plus le robot était considéré comme humain, moins il y avait de motivation.

Dans la même perspective, la revue systématique de Roselli et al. (2025) révèle que la culture influence l'anthropomorphisme envers les robots dans la majorité des études analysées, bien que la direction des effets ne soit pas uniforme. Plusieurs travaux indiquent que les participants issus des cultures d'Asie de l'Est et du Moyen-Orient ont tendance à davantage anthropomorphiser les robots : les participants chinois présentent une propension plus élevée que les Américains (Evers et al., 2008 ; Li et al., 2022), une relation positive entre les valeurs culturelles et l'anthropomorphisme apparaît chez les Japonais mais pas chez les Américains (Ikari et al., 2023), et les participants arabes anthropomorphisent plus volontiers les robots tout en les considérant davantage comme

des membres de leur groupe d'appartenance (Salem et al., 2014). Cependant, d'autres recherches rapportent le schéma inverse. Par exemple, les Canadiens se montrent plus indulgents envers les robots anthropomorphiques que les Indiens (Mehmood et al., 2024). Les voies de l'anthropomorphisme diffèrent selon la culture : les participants allemands et américains apparaissent plus anthropocentriques en évaluant surtout un robot par sa ressemblance au « groupe humain », tandis que les coréens et japonais jugent davantage en termes de « vie mentale » et lui attribuent des capacités mentales (mentalisation) sans passer par la catégorie « humain » (Spatola et al., 2022).

Roselli et al. (2025) avancent deux principales explications à ces variations culturelles. La première concerne les valeurs religieuses : les participants issus de cultures orientales, comme les Chinois ou les Japonais, seraient davantage enclins à l'animisme, ce qui les rendrait plus enclins également à attribuer aux robots des traits anthropomorphiques proches de l'humain (Li et al., 2022). La seconde explication tient à la familiarité avec les robots, qu'elle provienne d'expériences directes d'interaction ou de représentations médiatiques. Selon les auteurs, dans les pays orientaux, cette familiarité est plus fréquente pour les robots androïdes fonctionnels que pour les robots humanoïdes sociaux.

Cette section a présenté nos prédispositions perceptives à détecter des motifs, des agents et des visages dans des contextes ambigus, ainsi que la tendance du système perceptif humain à privilégier un schéma d'interprétation humain pour comprendre son environnement. Il a également été développé deux modèles de l'anthropomorphisme dont un spécifique aux interactions avec les robots, l'impact de ce phénomène et de l'apparence anthropomorphe. Afin d'articuler le phénomène de l'anthropomorphisme et ses conséquences, Kühne et Peter (2023) s'appuient sur la théorie de la *Posture Intentionnelle* proposée par Dennett (1987). La section suivante détaille ainsi la question de l'intentionnalité, de la posture intentionnelle et d'autres cadres associés.

2.2 Attribution d'états mentaux et interprétation des agents artificiels

Cette section aborde les stratégies d'interprétation face à tout système dont nous voulons prédire et expliquer le comportement. Il sera développé les stra-

tégies de Dennett (1987) dont la *Posture Intentionnelle* et son adoption envers les robots, ainsi que ses effets sur l'attention et l'interaction sociale. Il conviendra ensuite de distinguer la *Posture Intentionnelle* de la *Théorie de l'Esprit*, en précisant ce qu'elles ont en commun, notamment l'engagement du réseau cérébral de mentalisation. Pour compléter ces apports relatifs à l'attribution d'états mentaux, nous aborderons la perspective contemporaine de la *Mind Perception* (Gray et al., 2007), qui éclaire sur les facteurs conduisant un humain à prêter un esprit à une entité non-humaine. Nous verrons également que nos stratégies et capacités a faire des interprétations pourraient se baser sur des mécanismes qui émergent très tôt dans le développement. Enfin, nous présenterons la théorie de la *Media Equation* (Reeves & Nass, 1996) et le paradigme *CASA* (Nass et al., 1994), ainsi que leurs limites quant à la tendance à appliquer aux médias et aux ordinateurs les mêmes règles et normes sociales qu'aux humains.

2.2.1 La Posture Intentionnelle

Une fois qu'un agent a été identifié, l'humain cherche à prédire ses actions et à comprendre ses motivations. Pour ce faire, il ne se contente pas d'extrapoler des trajectoires mécaniques mais il recourt spontanément à des stratégies d'interprétation qui permettent d'anticiper le comportement d'autrui en termes d'états mentaux. À l'égard des ASA, l'attitude que l'humain adopte pour les interpréter peut être analysée à travers le prisme de la théorie de la *Posture Intentionnelle* (*Intentional Stance*) de Dennett (1987). Cette posture est une stratégie d'interprétation que nous utilisons spontanément face à tout système dont nous voulons prédire le comportement. Plus précisément, elle consiste à attribuer des états mentaux tels que des croyances, des désirs et des intentions à l'agent, comme s'il était un être rationnel. Cette posture ne présuppose pas nécessairement que l'agent possède réellement ces états mentaux. Dans *The Intentional Stance*, Dennett (1987) défend l'idée que, pour prédire le comportement d'un système, l'humain peut adopter différents « points de vue » explicatifs et donc il distingue trois niveaux d'explication (« *stance* »).

— la *Posture Physique* (« *physical stance* »), d'après laquelle pour prédire le comportement d'un système, nous nous basons sur sa constitution physique, sur la nature physique des contraintes et interactions qui s'exercent

sur lui puis également sur nos connaissances des lois de la physique (par exemple, l'effet de jeter une bouteille en verre sur un sol en béton ou encore l'effet de laisser un glaçon au soleil);

— la *Posture de la Conception* (« *design stance* »), consiste à prédire le comportement d'un système en se basant sur sa conception, en ignorant les détails physiques précis de sa constitution, et en supposant qu'il fonctionnera comme il a été conçu pour le faire dans diverses circonstances (par exemple, sans savoir s'il est à ressort mécanique ou solaire, on suppose qu'un réveil est conçu pour sonner à l'heure programmée et pour fonctionner de façon plus ou moins précise, ce qui suffit à prédire son déclenchement);

— la *Posture Intentionnelle* (« *intentional stance* »), consiste à prédire le comportement d'un système en le traitant comme un agent rationnel doté de croyances et de désirs. Concrètement, cette approche implique d'abord d'attribuer à l'agent les croyances qu'il devrait avoir compte tenu de sa position dans le monde et de ses objectifs, puis de lui attribuer les désirs appropriés selon les mêmes considérations. La prédiction s'effectue ensuite en supposant que cet agent rationnel agira pour satisfaire ses désirs en fonction de ses croyances. Cette posture est adoptée quand elle est le meilleur raccourci prédictif disponible, sans pour autant s'engager sur l'intériorité réelle du système, elle est une heuristique utile mais pas une assertion ontologique ou métaphysique (par exemple, pour prévoir le prochain coup d'un programme d'échec, ce sera dire qu'il croit que prendre mon cavalier ferait perdre sa tour, et dire qu'il veut éviter cette perte; on en déduit donc qu'il ne prendra pas le cavalier).

Pour reprendre l'exemple qui vient d'être cité, si l'on avait utilisé la posture de la conception, on décrirait le programme d'échecs par son fonctionnement prévu : ce serait dire que le programme d'échecs est conçu pour choisir le coup qui maximise une fonction d'évaluation, que le programme d'échecs génère les coups légaux, explore l'arbre de recherche [...] etc. Une telle stratégie serait évidemment très coûteuse et disproportionnée (et nécessiterait de connaître ou supposer correctement l'algorithme), tout comme le serait encore plus la Posture Physique. Autrement dit, face aux agents artificiels comme par exemple les robots sociaux, adopter la *Posture Intentionnelle* consiste à interpréter leurs ac-

tions en termes d'états mentaux supposés plutôt qu'en termes de mécanismes computationnels sous-jacents. Face à un robot social, il devrait être plus économique, cognitivement parlant, de prédire ses actions en l'interprétant comme un « être doté de buts » plutôt qu'en modélisant le détail de ses algorithmes. Par exemple, il est plus simple de choisir de décrire un robot qui se dirige vers un objet comme « voulant » ou « cherchant à » saisir cet objet, plutôt que comme exécutant un algorithme de navigation et de préhension. Comme le précisent Thellman et al. (2017), bien que l'on puisse interpréter les comportements des robots sociaux de manière similaire à ceux des humains avec la *Posture Intentionnelle*, cela reste une stratégie interprétative mais cela ne signifie pas nécessairement que l'on va attribuer réellement, au sens fort, des états mentaux.

Au sein de la recherche en HRI, des travaux ont développé et testé des outils permettant d'évaluer dans quelle mesure les individus adoptent la *Posture Intentionnelle* envers les robots. Le questionnaire développé par Marchesi et al. (2019) permet de le quantifier. Le questionnaire se compose de trente-quatre scénarios fictifs illustrés par des photographies montrant le robot *iCub* dans des actions quotidiennes. Chaque scénario est accompagné de deux descriptions concurrentes : l'une mentaliste (c'est-à-dire une explication formulée en termes de croyances ou de buts tel que « le robot fait X car veut Y ») et l'autre mécaniste (c'est-à-dire décrivant le comportement en termes de processus programmés et d'automatismes). Pour chaque scénario, les participants doivent déplacer un curseur le long d'une échelle bipolaire pour indiquer laquelle des phrases constitue le mieux une description de l'histoire représentée. Après l'analyse des résultats menée à l'aide de ce questionnaire, le score moyen obtenu révèle un biais global en faveur des explications mécanistes, donc de l'adoption d'une *posture de la conception*. Cependant, les résultats montrent qu'il est possible d'induire l'adoption de la *Posture Intentionnelle* envers un robot humanoïde, puisque certains scénarios ont suscité des réponses nettement plus mentalistes.

Dans le prolongement de ces travaux, Spatola et al. (2021b) ont validé une version abrégée en 12 items du même questionnaire (*IST-2, Intentional Stance Test-2*). Une analyse factorielle a permis d'identifier une structure en deux facteurs dans lesquels se répartissent les items : (1) le facteur « social robot », regroupant les scénarios où le robot interagit avec un humain et (2) le facteur « isolated robot », comprenant ceux où il agit seul. Des scores moyens plus éle-

vés de *Posture Intentionnelle* ont été trouvés en contextes sociaux, indiquant que la présence d'une interaction humaine favorise son adoption. L'analyse a pu également révéler que plus les scores à l'échelle augmentent, plus les participants attribuent au robot des scores élevés aux dimensions d'*Agentivité*, de *Sociabilité* et d'*Animéité* de l'échelle *Human-Robot Interaction Evaluation Scale* (HRIES, Spatola et al., 2021a). Ceci reflète le fait qu'adopter la *Posture Intentionnelle* se retrouve dans la tendance à attribuer des propriétés mentales. En comparant ces résultats à ceux qu'ils ont obtenus en utilisant l'échelle *NARS* (*Negative Attitude towards Robots scale*, de Nomura et al., 2006), ils ont également observé que plus les attitudes des individus envers les robots sont négatives, plus leur propension à adopter la *Posture Intentionnelle* est faible. À l'inverse, les personnes qui apprécient les activités cognitives exigeantes et éprouvent un fort besoin de réduire l'ambiguïté (en structurant et en ordonnant l'information) ont tendance à davantage mentaliser les actions du robot et, par conséquent, à adopter plus facilement la *Posture Intentionnelle* envers lui. Le facteur « *social robot* » prédisait significativement les attributions anthropomorphiques à d'autres robots (*NAO, Pepper*), tandis que le facteur « *isolated robot* » n'était associé qu'à la dimension de sociabilité. En somme, ce corpus de résultats indique qu'il existe des différences interindividuelles et contextuelles dans la propension à attribuer de l'intentionnalité aux robots, mais lorsque les conditions sont réunies, l'humain peut tout à fait adopter envers un robot la même posture interprétative que celle qu'il adopte face à un partenaire humain.

2.2.2 Réseau cérébral de mentalisation et attribution d'intentionnalité

Il convient maintenant de distinguer la *Théorie de l'Esprit* (*Theory of Mind, ToM*) de la *Posture Intentionnelle*. En effet, ces concepts sont étroitement liés et souvent confondus dans la littérature (Griffin & Baron-Cohen, 2002 ; Thellman et al., 2017). La *ToM* désigne la capacité à attribuer des états mentaux (croyances, désirs, intentions) à autrui et à comprendre que ces états jouent un rôle causal dans le comportement (Griffin & Baron-Cohen, 2002). Elle est typiquement mesurée par des tâches de fausses croyances qui testent notre capacité à comprendre qu'autrui peut avoir des croyances erronées sur la réalité (Baron-Cohen,

2001). La *Posture Intentionnelle* de Dennett (1987), pour rappel, est une stratégie pour prédire le comportement en attribuant des états mentaux, mais sans se préoccuper de savoir si ces états existent vraiment. Ainsi, la *ToM* renvoie à la capacité de comprendre l'esprit d'autrui, tandis que la *Posture Intentionnelle* relève d'une stratégie d'interprétation. Au niveau neuroanatomique, la *ToM* recrute de façon robuste le réseau de mentalisation, soit la jonction temporo-pariétale (TPJ) et le cortex préfrontal médian (mPFC, incluant l'aire paracingulaire antérieure, aPCC), avec un appui du précunéus et, selon les tâches, du pôle temporal (TP). En somme, la *ToM* et la *Posture Intentionnelle* ne sont pas identiques, mais se recouvrent souvent : dès qu'une tâche pousse les participants à adopter la *Posture Intentionnelle* envers un agent, on observe typiquement l'engagement du même réseau de mentalisation (Frith & Frith, 2006).

Lorsqu'on attribue de l'intentionnalité à un stimulus abstrait, le mPFC, la TPJ, ainsi que le sillon temporal supérieur (STS), des régions temporales basales et le cortex visuel sont sollicités (Castelli et al., 2000). Les zones du STS et de la TPJ sont associées à la *ToM* et plus un stimulus est perçu comme intentionnel, plus le débit sanguin augmente dans ces zones (Castelli et al., 2000). Ces mêmes régions s'activent lors de la perception des agents humains et artificiels (De Castro Martins et al., 2022) mais le degré de réalisme de l'agent module fortement l'intensité de l'activation (Mar et al., 2007).

Abu-Akel et al. (2020) ont fait jouer des participants à « pierre-feuille-ciseaux » dans quatre conditions croisées. Ils ont utilisé l'IRMf (Imagerie par résonance magnétique fonctionnelle, signal BOLD) et ont manipulé orthogonalement la perception de l'intentionnalité de l'adversaire (il répondait activement ou suivant un script) et son incarnation (humain ou ordinateur). Ainsi, leurs résultats montrent que le simple fait de considérer un adversaire (qu'il soit humain ou un ordinateur) comme intentionnel activait le réseau cérébral de mentalisation (TPJ bilatéral, aPCC/mPFC, précunéus, et TP droit). En revanche, jouer contre un adversaire humain, comparé à un ordinateur, activait un sous-réseau latéralisé à droite plus restreint (TPJ droit et aPCC). Enfin, l'analyse de l'interaction entre intentionnalité et type d'adversaire a révélé une activation du pôle frontal gauche, interprété comme un coût de contrôle lorsque l'on doit aller contre la posture qu'on activerait par défaut, c'est-à-dire traiter un ordinateur répondant activement comme intentionnel ou un humain scripté comme non-intentionnel. En

somme, c'est la croyance d'intentionnalité attribuée à l'adversaire (qu'il soit humain ou ordinateur) qui déclenche le réseau de mentalisation et, à intentionnalité égale, une augmentation plus localisée à droite apparaît pour l'humain (Abu-Akel et al., 2020). Ces résultats montrent ainsi que l'engagement du réseau de mentalisation dépend avant tout de l'attribution d'intentionnalité, indépendamment de l'incarnation humaine.

Cependant, dans une tâche analogue de « pierre-feuille-ciseaux », Chaminade et al. (2012) montrent que jouer contre un humain active clairement le réseau de mentalisation (mPFC, TPJ droit), alors que jouer contre une intelligence artificielle n'engendre pas d'activation supplémentaire dans ces régions par rapport à un adversaire répondant de façon aléatoire. L'activité est surtout accrue dans des régions liées à l'effort attentionnel (précunéus, sillon intrapariétal postérieur, cortex préfrontal latéral) et, plus légèrement, à la résonance motrice (prémoteur/sillon intrapariétal antérieur gauches). Les auteurs en concluent que, dans ce protocole, les participants n'ont pas adopté la *Posture Intentionnelle* envers l'agent artificiel et qu'en conséquence, les régions de mentalisation (mPFC, TPJ droit) ne montrent pas d'activation supérieure à celle observée face à l'adversaire aléatoire.

Bossi et al. (2020) ont examiné si l'activité neuronale de repos pouvait prédire les postures, c'est à dire l'usage de la *Posture Intentionnelle* ou de la *Posture de Conception*, envers les robots humanoïdes. Pour cela, ils ont enregistré l'activité électroencéphalographique (EEG) de participants avant qu'ils ne réalisent la tâche d'attribution d'intentionnalité de Marchesi et al. (2019) présentée plus tôt dans ce chapitre. Les résultats ont révélé que l'activité bêta au repos (13-27 Hz) différenciait significativement les participants selon leur tendance ultérieure à adopter la *Posture Intentionnelle* ou la *Posture de Conception* envers le robot *iCub*. Spécifiquement, les participants du groupe *Posture de Conception* présentaient une activité bêta plus élevée au niveau d'un *cluster* temporo-pariétal gauche et fronto-temporal droit, comparativement au groupe *Posture Intentionnelle*. Les auteurs interprètent cette activité bêta comme un corrélat du réseau du « mode par défaut » (*Default Mode Network*) qui est traditionnellement associé aux processus de mentalisation. Paradoxalement, une plus grande activation des processus de mentalisation au repos prédisait une tendance à traiter les robots selon la *Posture de Conception* plutôt que selon la *Posture Intentionnelle*,

suggérant que l'engagement préalable dans la réflexion sur les états mentaux humains pourrait accentuer la distinction catégorielle entre agents naturels et artificiels. Durant la tâche elle-même, une analyse temps-fréquence a révélé que le groupe *Posture de Conception* présentait une désynchronisation gamma (28-45 Hz) plus marquée dans une région occipito-temporale avant l'émission de la réponse, confirmant les travaux sur l'activité gamma comme marqueur des processus de mentalisation.

Dans le prolongement de ces travaux, Roselli et al. (2024) ont exploré l'influence du domaine de formation et l'exercice professionnel de personnes sur l'adoption de la *Posture Intentionnelle* en comparant l'activité EEG au repos chez des roboticiens et des psychothérapeutes. Les thérapeutes, dont la formation encourage le développement de compétences de mentalisation, présentaient des scores à l'*IST* significativement plus élevés que les roboticiens. Cette différence comportementale était accompagnée d'une activité gamma plus marquée au repos dans un *cluster* postérieur droit, cohérente avec l'implication de la jonction temporo-pariétale droite dans les processus de mentalisation. Contrairement aux résultats de Bossi et al. (2020), aucune différence significative n'a été observée dans la bande β. Ceci suggère que les différences individuelles liées au domaine de formation et à l'exercice professionnel se manifestent différemment des biais généraux observés en population générale.

En somme, la *ToM* et la *Posture Intentionnelle* sont proches et mobilisent le même réseau cérébral de mentalisation, mais diffèrent dans leur nature. Le fait d'attribuer de l'intentionnalité à un agent (qu'il soit humain ou artificiel) peut activer ce réseau indépendamment de son incarnation, mais un humain est spontanément perçu comme intentionnel, tandis qu'un agent artificiel ne l'est que si le contexte pousse à l'interpréter ainsi. Enfin, son engagement semble dépendre également du domaine professionnel des individus.

2.2.3 Mécanismes d'attention sociale et attribution d'intentionnalité

L'adoption de la *Posture Intentionnelle* envers un agent ne modifie pas seulement la manière dont nous interprétons son comportement, mais influence aussi nos mécanismes d'attention sociale. Plusieurs études ont montré que la

simple croyance en l'intentionnalité d'un agent module aussi bien les performances comportementales que les réponses cérébrales précoces. Une série d'expériences l'a démontré à travers le paradigme du *gaze-cueing*, elles sont détaillées dans la suite du manuscrit.

Wiese et al. (2012) ont manipulé l'adoption de la *Posture Intentionnelle* dans une tâche de *gaze-cueing*. Dans cette tâche, les participants doivent détecter des cibles qui apparaissent soit (1) là où un visage regarde (ce sont les essais dits « valides »), soit (2) du côté opposé (ce sont les essais dits « invalides »). Dans leur expérience, les auteurs ont directement influencé la croyance des participants par des instructions : les mêmes visages (humains ou robots) étaient présentés, mais on disait aux participants qu'ils observaient soit (1) un agent intentionnel (humain ou robot contrôlé par un humain), soit (2) un système mécanique (mannequin, ou robot pré-programmé). Ainsi, quand les participants croyaient observer un comportement intentionnel, ils étaient bien plus rapides à détecter des cibles apparaissant là où le visage regardait (en comparaison aux essais invalides où les cibles apparaissaient du côté opposé). Cet effet d'orientation de l'attention (*cueing*), mesuré par la différence entre essais valides et invalides, était deux fois plus important que lorsqu'ils pensaient observer un système mécanique, et ce indépendamment de l'apparence physique de l'agent. De même, Wykowska et al. (2014) ont démontré que cette modulation comportementale par la *Posture Intentionnelle* peut être mesurée dans des réactions cérébrales précoces : la composante P1 (apparaissant 100-140 ms après l'apparition de la cible), qui reflète l'amplification du signal visuel dans le cerveau, était significativement plus importante pour les cibles validement orientées par le regard uniquement lorsque les participants croyaient que ce regard émanait d'un agent intentionnel.

Cependant, d'autres recherches ont affiné ces résultats en testant les conditions et les limites de l'effet. Morillo-Mendez et al. (2023) ont investigué si l'effet de *gaze-cueing* serait réduit lorsque le robot (ici *Nao*) ne peut pas « voir » les cibles. Dans leur protocole, le robot *NAO* avait soit (1) la ligne de vue sur les écrans alternativement dégagée (condition de base), soit (2) obstruée. Il a été observé que l'effet de *gaze-cueing* (qui se manifeste, pour rappel, par des temps de réaction réduits quand la cible apparaît du côté vers lequel le robot tourne la tête) était présent dans les deux conditions, mais significativement réduit en

condition d'occlusion.

Enfin, d'autres travaux ont exploré l'impact de la répétitivité des essais de *gaze-cueing* avec un robot sur l'adoption de la *Posture Intentionnelle* et sur les effets comportementaux du *gaze-cueing*. Abubshait et Wykowska (2020) ont comparé trois durées d'exposition (96, 176 ou 256 essais) en utilisant le robot *iCub* et ont mesuré les scores au questionnaire *IST* (Marchesi et al., 2019) ainsi que l'effet de *gaze-cueing* (c'est-à-dire le temps de réaction) avant et après interaction. Il en ressort que l'exposition courte augmente significativement les scores d'attribution d'intention : un effet qui s'annule après exposition prolongée, tandis que les effets de *gaze-cueing* s'amplifient avec la durée d'exposition. Autrement dit, même si l'attention suit de plus en plus le regard du robot avec la répétition, les participants cessent d'attribuer des intentions au robot lorsque l'exposition devient trop longue.

En résumé, l'ensemble de ces travaux montre que l'adoption de la *Posture Intentionnelle* semble influencer les mécanismes d'attention sociale. Les individus orientent plus efficacement leur attention lorsqu'ils croient que l'agent observé agit intentionnellement, et ce, indépendamment de son apparence. Cette croyance amplifie également les réponses cérébrales précoces associées au traitement visuel. Cependant, cet effet dépend du contexte : il diminue lorsque les indices suggèrent que l'agent ne perçoit pas la scène et il évolue dans le temps. De même, une exposition prolongée à un agent renforce les réponses attentionnelles automatiques mais réduit progressivement l'attribution d'intentionnalité. En somme, l'adoption de la *Posture Intentionnelle* module à la fois l'attention sociale, la perception et les dynamiques selon la croyance en l'intention de l'agent et les conditions d'interaction.

2.2.4 Perception de l'esprit et agents artificiels

Une autre approche complémentaire pour comprendre l'attribution d'intentionnalité aux agents artificiels consiste à examiner la manière dont les individus attribuent globalement un « esprit » aux entités non humaines. C'est précisément l'objet du cadre de *Mind Perception* (perception de l'esprit) proposé par Gray et al. (2007).

Dans leur étude, les participants devaient comparer soixante-dix-huit paires d'entités sur la base de capacités mentales (par exemple, être capable de res-

sentir la douleur, être capable de mémoriser, etc.) ou sur la base de jugements personnels (par exemple, lequel ils aiment le plus, lequel est le plus susceptible d'avoir une âme, etc.). Les entités évaluées étaient : une grenouille, un chien, un chimpanzé, un fœtus humain, un bébé, une petite fille, une femme, un homme, la personne interrogée elle-même, un homme en état végétatif, une femme décédée, Dieu et un robot. Les résultats de cette étude révèlent que, contrairement aux approches qui considéraient la perception de l'esprit comme étant unidimensionnelle (on attribuerait plus ou moins d'esprit), l'attribution d'états mentaux s'organise finalement selon deux dimensions orthogonales : *Expérience* et *Agentivité* (« *Agency* »). L'*Expérience* englobe des capacités subjectives telles que la faim, la peur, la douleur, le plaisir, la rage, le désir, la personnalité, la conscience, la fierté, l'embarras et la joie. L'*Agentivité* couvre des fonctions d'action et de raisonnement telles que l'autocontrôle, la moralité, la mémoire, la reconnaissance des émotions, la planification, la communication et la pensée. Ainsi, le *Mind Perception* offre un cadre descriptif et mesurable sur ce qu'on attribue comme esprit à une entité cible.

Dans la continuité du modèle proposé par Gray et al. (2007), Waytz et al. (2010) examinent les causes et les conséquences de l'attribution d'un esprit, selon qu'elles relèvent de l'entité perçue ou de la personne qui lui prête un esprit. S'agissant d'abord des causes, certaines tiennent donc à l'entité elle-même. Ses caractéristiques peuvent favoriser l'attribution d'un esprit, par exemple lorsqu'elle présente des indices rappelant l'humain ou une similarité avec la personne qui lui prête un esprit. De même, le fait que l'entité soit imprévisible ou qu'elle produise des conséquences négatives favorise le fait qu'on lui attribue un esprit. D'autres causes tiennent à la personne qui attribue un esprit. Ses motivations peuvent découler d'un besoin de comprendre et de reprendre le contrôle face à l'incertitude ou d'un besoin de connexion sociale qui l'amène à anthropomorphiser pour recréer du lien social. Ces causes s'inscrivent dans le modèle *SEEK* d'Epley et al. (2007) présenté plus tôt dans ce chapitre, qui identifiait la *Motivation d'Effectance* et le besoin de lien social comme déterminants de l'anthropomorphisme.

Concernant ensuite les conséquences de l'attribution d'un esprit selon Waytz et al. (2010), elles peuvent également être envisagées du point de vue de l'entité ou de celui de la personne qui réalise l'attribution. Pour l'entité, être perçue

comme dotée d'un esprit lui confère un statut moral : elle peut alors être considérée comme un être envers lequel on reconnaît des droits ou, selon le cas, comme un agent moral responsable de ses actes. Pour la personne qui attribue un esprit, cette attribution favorise la création de sens et renforce aussi la perception d'une surveillance, ce qui peut modifier ses comportements. Ainsi, une faible attribution d'un esprit conduit à la déshumanisation et peut légitimer des traitements inhumains. En somme, Waytz et al. (2010) explique que l'attribution d'un esprit n'est pas automatique et qu'elle peut s'appliquer aussi bien à des humains qu'à des entités non humaines (comme des dieux, des animaux ou des objets), ou même être refusée à certains humains (notamment dans le cadre de la déshumanisation), car elle dépend des motivations de la personne et des caractéristiques perçues de l'entité.

Ce cadre d'attribution d'états mentaux selon deux dimensions (Gray et al., 2007) offre un éclairage particulièrement intéressant pour comprendre le malaise lié à la perception des agents artificiels et des robots sociaux, ressenti par certains individus. En effet, les travaux de Gray et Wegner (2012) montrent que ce malaise est surtout associé à de l'attribution d'*Expérience* plutôt qu'à l'attribution d'*Agentivité*. Autrement dit, les individus ont tendance à éprouver un certain inconfort face à des agents perçus comme capables de ressentir sans pouvoir agir, alors qu'un agent perçu comme purement fonctionnel ne suscite généralement pas ce type de réaction. Cela suggère que, dans la conception des robots sociaux, la mise en avant de comportements orientés vers l'action et la planification pourrait contribuer à réduire le malaise perçu, tandis qu'une simulation trop marquée ou réaliste d'états émotionnels internes risque de renforcer cette impression d'inconfort.

En conclusion, cette section montre que l'attribution d'intentionnalité aux agents artificiels peut s'inscrire dans un cadre plus large de perception d'un esprit. Les individus attribuent des états mentaux selon deux dimensions distinctes : l'*Expérience* (ressentir) et l'*Agentivité* (agir). Cette attribution dépend à la fois des motivations de la personne (besoin de contrôle ou de lien social) et des caractéristiques de l'entité (ressemblance humaine). Ces mécanismes ont des effets psychologiques et moraux, par exemple sur l'empathie, la protection ou la responsabilité. Enfin, ce cadre éclaire le malaise suscité par certains robots qui est surtout associé à l'attribution d'*Expérience* plutôt qu'à l'*Agentivité*.

2.2.5 Modèle Like-me et Accordage Social

Les stratégies et capacités d'interprétation pourraient se baser sur des mécanismes qui émergent très tôt dans le développement de l'humain. Le modèle *Like-Me* de Meltzoff (2007) propose que la reconnaissance des équivalences entre soi et autrui constitue le fondement de la cognition sociale. Selon cette approche, les nourrissons possèdent un système de représentation qui leur permet de faire correspondre les actions qu'ils voient chez autrui avec leurs propres actions corporelles. Cette capacité repose sur un code commun qui intègre les informations visuelles, motrices et proprioceptives. Les résultats empiriques appuient cette proposition : dès quatorze mois, les nourrissons préfèrent les adultes qui les imitent plutôt que ceux qui agissent de manière synchronisée mais différente (Meltzoff, 2007). Cette préférence indique que les nourrissons détectent la correspondance entre leurs mouvements et ceux d'autrui, au-delà de la simple synchronisation temporelle. Cette reconnaissance précoce des similarités constituerait la base développementale pour attribuer des états psychologiques aux autres agents. Dans le contexte de l'interaction humain-robot, Meltzoff et al. (2010) ont testé si ces mécanismes s'appliquent aussi aux agents artificiels. Leur étude montre que des nourrissons de dix-huit mois suivent davantage le regard d'un robot humanoïde après avoir observé ce dernier engagé dans des échanges communicatifs et imitatifs avec un adulte. Ces résultats suggèrent que l'observation de comportements sociaux familiers peut amener les nourrissons à traiter les entités artificielles comme des agents psychologiques.

Dans une perspective de clarification conceptuelle pour l'HRI, Perez-Osorio et Wykowska (2020) proposent de nommer et unifier sous un terme unique l'ensemble des mécanismes de cognition sociale activés en interaction, en convergeant vers le concept d'*Accordage Social* (« *social attunement* »). Les auteurs le définissent comme un concept parapluie qui englobe tous les mécanismes de cognition sociale comme le *regard mutuel* (« *mutual gaze* »), l'*attention conjointe* (« *joint attention* ») et la *prise de perspective spatiale* (« *spatial perspective taking* »), qui sont activés au cours des interactions sociales. Ils suggèrent que l'adoption de la *Posture Intentionnelle* constitue potentiellement un levier pivot facilitant cet *Accordage Social* avec les agents artificiels (Perez-Osorio & Wykowska, 2020). Les données présentées précédemment soutiennent cette proposition, puisque l'attribution d'états mentaux aux robots module effective-

ment l'engagement des mécanismes attentionnels (Wiese et al., 2012), perceptifs (Wykowska et al., 2014) et sociaux précoces (Meltzoff et al., 2010). Cette perspective suggère que, sous certaines conditions, les mêmes mécanismes sociocognitifs qui permettent l'accordage avec les partenaires humains peuvent être activés dans l'interaction humain-robot.

2.2.6 Théorie de la Media Equation et le paradigme CASA

La théorie de la *Media Equation* (Reeves & Nass, 1996) propose une perspective plus générale à l'ensemble des théories développées plus tôt dans ce manuscrit : les humains tendent à traiter les ordinateurs, télévisions et autres médias comme des entités sociales. Selon cette approche, nous appliquons inconsciemment aux échanges médiatiques les mêmes règles et normes sociales qu'aux interactions entre humains. Cette réaction sociale « automatique » s'expliquerait par des mécanismes cognitifs profondément enracinés : notre cerveau, façonné par l'évolution dans un monde où seuls les êtres vivants manifestent des comportements sociaux riches, n'aurait pas développé de stratégie distincte pour répondre aux signaux sociaux émis par des entités non vivantes. Ainsi, dès lors qu'un média ou un artefact technologique présente des indices sociaux (par exemple, le langage, la voix, le regard, l'interactivité), il enclenche en nous les mêmes heuristiques et schémas de réponse que ceux dédiés aux interactions humaines. Reeves et Nass (1996) soulignent d'ailleurs que cette réponse est largement inconsciente (« *mindless* », c'est-à-dire sans délibération active) : nous traitons automatiquement ce qui semble social comme réellement social, sans nécessairement croire que la machine a des intentions ou des sentiments. Ils distinguent en ce sens ce phénomène de notre tendance à faire de l'anthropomorphisme.

Le paradigme *CASA* (*Computers Are Social Actors*), initié dès 1994 (Nass et al., 1994) et consolidé plus tard (Nass & Moon, 2000 ; Nass et al., 1996), découle directement de la *Media Equation* et s'intéresse plus spécifiquement aux ordinateurs. Dans les faits, le *CASA* est donc une spécialisation de la *Media Equation*. Ce paradigme avance que les humains appliquent automatiquement les mêmes règles sociales humaines aux ordinateurs : par exemple, les participants évaluent plus favorablement un ordinateur lorsque l'évaluation se déroule sur le même ordinateur plutôt que sur un ordinateur différent (ils appliqueraient une norme

de politesse) ou encore vont appliquer des stéréotypes de genre à une voix de synthèse féminine ou masculine en percevant une voix masculine comme plus dominante et assertive (Nass et al., 1994).

Aujourd'hui, la *Media Equation* et le *CASA* devraient, en principe, s'appliquer à nos robots sociaux modernes, notamment en raison de leur apparence et de leur comportement. Cependant, la pertinence de ces théories, tant en général que pour les robots sociaux, pose question. D'une part, les travaux fondateurs datent d'une époque où les technologies interactives étaient différentes et relativement nouvelles pour le public. D'autre part, plusieurs études ont clairement nuancé et précisé les limites des idées postulées à la base de la *Media Equation*, tandis que d'autres ont échoué à répliquer les travaux princeps. Plusieurs de ces études sont présentées dans la suite de cette section.

Une étude de Bartneck et al. (2005), voulant investiguer les limites de la *Media Equation* offre une mise en perspective critique de celle-ci en montrant que l'attribution spontanée de caractéristiques sociales aux artefacts n'est pas inconditionnelle. En effet, en reproduisant le protocole de Milgram avec un robot « victime », Bartneck et al. (2005) observent que 100% des participants allaient jusqu'au choc maximal, contre seulement 40% dans la condition humaine, soit une réticence largement amoindrie à infliger de la « souffrance » à un agent perçu comme non-sentient. Le modèle *CASA* décrit efficacement l'inclinaison générale des utilisateurs à traiter les interfaces comme des entités sociales, mais cette propension semble s'estomper dès lors que l'objet de l'interaction ne peut avoir de sentience ou de capacité à éprouver une douleur.

En testant l'hypothèse selon laquelle les utilisateurs appliqueraient inconsciemment ou machinalement (« *mindlessly* », comme supposé par la théorie) les mêmes règles sociales aux ordinateurs qu'aux humains, Johnson et Gardner (2007) ont obtenu des résultats pouvant également nuancer le paradigme *CASA*. En effet, ils ont fait intervenir des étudiants (distingués selon leur faible ou forte expérience en informatique) qui ont réalisé deux tâches informatiques : un *Text Rating Task* consistant à évaluer six adjectifs décrivant un passage de fiction, et un *Desert Survival Task* consistant à classer douze objets selon leur importance pour survivre dans un désert. Chaque participant travaillait soit seul, soit au sein d'une équipe humaine, soit au sein d'une équipe mixte intégrant humains et ordinateurs. Or, chez les participants très expérimentés, la présence de l'ordi-

nateur comme co-équipier n'a pas induit de conformité aux recommandations de l'ordinateur : ils ont jugé l'information fournie par la machine de moindre qualité, déclaré être moins influencés par elle et ajusté leurs propres notes et classements à l'inverse des recommandations informatiques.

Une autre preuve des limites du paradigme *CASA*, bien qu'elle ne soit pas formulée explicitement comme telle, peut être trouvée dans une étude longitudinale de Croes et Antheunis (2021). Elles montrent, dans le cadre de sept interactions avec l'agent conversationnel *Mitsuku*, réparties sur trois semaines, une diminution linéaire et significative des indicateurs de réponses sociales (tels que l'attraction sociale ou l'empathie), à mesure que l'exposition se répète avec l'agent conversationnel.

Enfin, la mise en échec la plus significative du paradigme *CASA* est l'étude de Heyselaar (2023) qui est une réplication directe de l'expérience fondatrice du paradigme sur la politesse. Cette réplication menée avec 132 participants et selon le protocole original de 1994 ne révèle aucune différence significative entre les conditions. Cette absence de réplication remet en question la validité temporelle du *CASA*. L'autrice suggère que l'effet retrouvé dans l'étude princeps tenait principalement à la nouveauté technologique que représentaient les ordinateurs dans les années 1990, plutôt qu'à un phénomène psychologique stable, et que l'effet *CASA* ne s'appliquerait plus aux ordinateurs de bureau, mais se limiterait aux technologies émergentes ou socialement nouvelles.

Cependant, en dépit de certaines limites, la *Media Equation* et le *CASA* demeurent des cadres précieux pour comprendre pourquoi et comment les humains réagissent aux technologies : ils montrent que la présence de simples indices sociaux (par exemple, le langage, la voix, le regard ou l'interactivité) suffit à déclencher nos heuristiques sociales et à orienter nos comportements. Dans le cas des robots sociaux, ces cadres aident à expliquer pourquoi les signaux sociaux peuvent déclencher les mêmes comportements sociaux que ceux retrouvés dans les interactions entre humains, tout en soulignant que ces effets ne sont ni stables ni universels.

En résumé, cette section sur l'intentionnalité a montré que la *Posture Intentionnelle* est une stratégie utile pour prédire le comportement d'un système en lui prêtant des croyances, des désirs et des buts, sans pour autant lui accorder des intentions ou sentiments réels. Elle se distingue de la *ToM*, qui est une

capacité, tout en recrutant le même réseau de mentalisation lorsque l'agent est tenu pour intentionnel. Cette croyance module l'attention sociale, avec des effets dépendants du contexte et du temps d'exposition. L'attribution d'esprit se structure selon deux dimensions, *Expérience* et *Agentivité*, elle varie avec les motivations de l'individu et les indices portés par l'entité, et elle a des conséquences morales pour l'individu comme pour l'entité perçue. Enfin, la *Media Equation* et le paradigme *CASA* expliquent pourquoi de simples indices sociaux déclenchent des réponses sociales, tout en rappelant des limites liées à la technologie, au contexte et à l'expérience des utilisateurs.

2.3 Similitudes et différences des mécanismes face à un humain

Après avoir développé différents mécanismes et modèles que nous adoptons face aux agents artificiels, cette section cherche à les mettre en perspective pour apporter un début de réponse à la question qui reste centrale à cette thèse : réagissons-nous aux robots comme à nos semblables humains, ou bien mobilisons-nous des processus distincts ? En effet, les données actuelles suggèrent que de nombreux mécanismes sociaux fondamentaux s'enclenchent automatiquement face aux robots, indiquant ainsi une potentielle continuité avec l'interaction entre humains. Cependant, des divergences significatives et des asymétries persistent qui révèlent les limites de cette analogie.

2.3.1 Convergences des mécanismes

Nous l'avons abordé, il semblerait que des mécanismes sociaux de base s'enclenchent face aux agents artificiels, suggérant une continuité avec l'interaction humain-humain : plusieurs cadres théoriques anciens appuient cette idée comme la *Media Equation* et le *CASA* (Nass et al., 1994 ; Reeves & Nass, 1996). Dans cette continuité, l'anthropomorphisme, c'est-à-dire notre tendance à projeter de l'humain sur du non-humain, révèle comment nous utilisons des cadres humains pour combler les lacunes relationnelles et réduire l'incertitude (Epley et al., 2007). D'un point de vue évolutionniste et à un niveau perceptuel, l'humain semble être câblé pour manifester une sensibilité à la détection d'agents inten-

tionnels (Barrett, 2000 ; Guthrie, 1995 ; Maij et al., 2019) : il réagit de façon excessive à des informations ambiguës susceptibles de signaler la présence d'autres agents. Ce biais s'applique aux agents tels que les humains et les autres animaux (Barrett, 2000) mais il pourrait potentiellement être mobilisé avec les ASA. En effet, Castelli et al. (2000) montrent que la perception d'intention, même dans des formes abstraites, active les zones associées à la théorie de l'esprit et de la mentalisation (les régions TPJ et STS notamment). Ces dernières s'activent dans la perception des agents humains et artificiels (De Castro Martins et al., 2022) avec le degré de réalisme de l'agent modulant fortement l'intensité de l'activation (Mar et al., 2007). Il est à noter que, dans certaines conditions, les nourrissons humains traitent les robots comme des agents psychologiques (c'est-à-dire comme des entités perçues comme capables de perception et d'interaction sociale ; Meltzoff et al., 2010). La théorie de la *Posture Intentionnelle* de Dennett (1987) propose que les humains expliquent et prédisent le comportement d'un système en attribuant des états mentaux tels que des croyances, des désirs et des intentions à l'agent, comme s'il était un être rationnel, afin de prédire et d'expliquer son comportement. Des études récentes montrent que les humains pourraient adopter la *Posture Intentionnelle* envers des agents artificiels (Abu-Akel et al., 2020) comme les robots sociaux (Marchesi et al., 2019 ; Spatola et al., 2021b). Le modèle de la perception de l'esprit, qui est un cadre descriptif sur ce qu'on attribue comme esprit à une cible, montre que les dimensions selon lesquelles sont évalués les humains sont également mobilisées lors du jugement sur les robots (Gray et al., 2007 ; Gray & Wegner, 2012).

Un autre élément de convergence se retrouve dans l'activation des neurones miroirs lors de l'observation d'une action, qu'elle soit réalisée par un humain ou par un robot. En effet, le système des neurones miroirs (situé dans les cortex prémoteur dorsal et ventral, pariétal inférieur et le gyrus temporal moyen), qui est un réseau s'activant à la fois lors de l'exécution d'une action et de l'observation d'autrui la réalisant, permet de simuler mentalement l'action perçue afin d'en comprendre l'intention (Gazzola et al., 2007). Plus précisément, il a été démontré que l'observation d'actions à finalité claire (par exemple, saisir un verre ou verser un liquide), même lorsqu'elles sont effectuées par un bras robotique, mobilise les mêmes circuits neuronaux que lorsque ces actions sont accomplies par un humain (Gazzola et al., 2007).

De même, Urgen et al. (2013) ont montré par EEG que l'atténuation de la puissance mu (8-13 Hz), indice de l'activité du système miroir, est équivalente lors de la perception d'actions effectuées par un humain, un robot androïde ou un robot d'apparence mécanique, tandis que l'augmentation des oscillations thêta frontales (4-8 Hz), associée aux processus mnésiques, est significativement plus marquée lors de l'observation de l'agent artificiel à l'apparence la plus robotique comparativement à un robot androïde et à un humain. Ceci suggère un effort accru de récupération et d'encodage mnésique pour traiter des actions mécaniquement moins familières.

Les circuits neuronaux relatifs aux émotions et à l'empathie présentent également certains schémas d'activation similaires face à un agent artificiel et à un humain (Rosenthal-von Der Pütten et al., 2014). Ces circuits se situent au niveau de l'amygdale, l'insula et le gyrus frontal inférieur. Il a été observé qu'ils s'activent lors de l'observation de la souffrance d'un humain, mais aussi d'un robot (ici le robot dinosaure *Pleo*). Néanmoins, la violence dirigée contre un humain provoque une réponse significativement plus forte dans le putamen droit, traduisant une empathie accrue envers la victime humaine. De la même manière, observer les mains d'un robot ou d'un humain dans des situations douloureuses déclenche des réponses d'empathie automatique, bien que la réponse émotionnelle soit plus tardive dans le cas des mains robotiques. Ces résultats suggèrent que si l'émotion générée par l'observation de la douleur d'agents artificiels est comparable à celle suscitée par l'observation de la douleur humaine, le traitement cognitif de l'empathie diffère selon la nature de l'agent (Suzuki et al., 2015).

En termes de réponses physiologiques, Li et al. (2017) ont montré que le fait de toucher des zones de faible accessibilité (c'est-à-dire les parties du corps que les individus laissent rarement ou difficilement toucher par autrui, comme l'intérieur des cuisses, les fesses ou l'entrejambe) chez un robot produit des effets mesurables sur la perception et la réaction des participants, comparables à ceux observés par d'autres études lorsque ces mêmes zones sont touchées chez un humain. En effet, toucher ces zones chez un robot humanoïde (*NAO*) induit une excitation physiologique (observée par l'augmentation de la conductance cutanée et du temps de réaction) significativement plus élevée que pour des zones de haute accessibilité (pour lesquelles les contacts sont plus souvent acceptés, comme les mains ou les pieds). Ceci suggère que les humains tendent à trans-

férer et à activer involontairement, lors de leurs interactions avec un agent artificiel, des normes sociales similaires à celles qu'ils appliquent aux interactions entre humains.

2.3.2 Divergences et limites des mécanismes

Si de nombreux mécanismes observés chez les individus, qu'ils soient exposés à un humain ou à un agent artificiel, semblent converger, la littérature met également en évidence des ruptures, voire de fortes asymétries, qui remettent en question l'idée d'une application stricte et uniforme des mêmes modèles aux deux contextes.

Pour rappel, l'une des premières limites empiriques aux modèles de la *Media Equation* ou du *CASA*, et donc une divergence importante, concerne l'extension automatique des normes et règles sociales humaines aux robots : Bartneck et al. (2005) montrent que, les individus se montrent beaucoup moins réticents à infliger des chocs électriques à un robot qu'à un humain. Cette moindre inhibition à provoquer une « souffrance » met en évidence une frontière nette dans la projection de droits ou de statut moral. D'autres travaux, comme ceux de Johnson et Gardner (2007), nuancent l'idée d'une application automatique ou non consciente des règles sociales : les usagers experts en informatique jugent l'information fournie par un ordinateur de moindre qualité et sont moins influencés par lui, montrant que la familiarité, l'expertise ou la redondance de l'exposition atténuent la propension à généraliser les règles sociales humaines à la machine. Cette observation est également soutenue par Croes et Antheunis (2021), qui mettent en évidence une diminution linéaire et significative de l'empathie, de l'attraction sociale et des réponses conformes attendues par le paradigme *CASA* lors d'interactions répétées avec un agent conversationnel. Heyselaar (2023) rapporte un échec de réplication de l'effet de politesse chez l'ordinateur, suggérant que ces effets étaient sans doute en partie liés à la nouveauté technologique à la période étudiée initialement.

Ensuite, les réponses cérébrales suggèrent que le traitement émotionnel varie selon l'agent. Chaminade et al. (2010) ont montré, par IRMf, que l'observation de gestes émotionnels (de joie, de colère ou de dégoût) chez un robot humanoïde (*WE-4RII*) se traduit par une augmentation de l'activation des cortex occipitaux et temporaux postérieurs par rapport un humain adoptant ces même gestes

émotionnels. Ceci suggère une mobilisation différente de ces circuits et un traitement visuel supplémentaire lorsqu'on perçoit un agent mécanique anthropomorphe. De manière contrastée, on observe une diminution de la résonance dans des régions clés des réseaux moteurs et émotionnels. Ainsi, le cerveau mobilise davantage ses aires visuelles associatives pour analyser les expressions produites par un robot (Chaminade et al., 2010).

Si la *Posture Intentionnelle* semble être adoptée envers un robot social, elle n'est ni systématique, ni stable, contrairement à ce qui peut être retrouvé pour l'humain. Les travaux de Marchesi et al. (2019) révèlent une préférence globale pour les explications mécanistes (via la *Posture de Conception*) lorsque les participants doivent expliquer le comportement d'un robot, sauf dans des contextes explicitement sociaux ou scénarisés, qui peuvent induire une posture plus mentaliste. La version abrégée du questionnaire *IST-2* (Spatola et al., 2021b) montre que l'attribution d'intentionnalité est significativement plus forte lorsque le robot est impliqué dans une interaction avec un humain. De même, Abu-Akel et al. (2020) montrent que c'est l'attribution d'intentionnalité, plus que l'incarnation humaine de l'agent, qui détermine l'activation du réseau de mentalisation. Toutefois, cette activation est plus constante avec un agent humain, spontanément perçu comme intentionnel, alors qu'elle ne se manifeste pour un agent artificiel que lorsque le contexte pousse à l'interpréter ainsi (Abu-Akel et al., 2020). Les asymétries morales constituent aussi un autre marqueur de divergence significative par rapport à l'humain. Bartneck et Keijsers (2020) montrent que, si l'agression envers les robots est jugée aussi immorale qu'envers les humains, la réciprocité morale ne s'applique pas : il n'est pas jugé acceptable pour un robot de se défendre contre un humain.

2.3.3 Entre similitudes et différences

En définitive, les humains appliquent-ils les mêmes modèles lorsqu'ils sont exposés à un agent artificiel ou à un humain ? L'interaction avec un robot mobilise-t-elle les mêmes mécanismes cognitifs que l'interaction entre humains, ou bien repose-t-elle sur des processus distincts ? Les données actuelles invitent à une réponse nuancée : de nombreux indicateurs neuronaux, physiologiques, attentionnels et comportementaux mettent en évidence des schémas similaires selon que l'agent soit humain ou artificiel. Néanmoins, des frontières importantes per-

sistent entre les traitements que nous appliquons à nos semblables humains et à des entités artificielles.

Sur le plan des convergences, de nombreux mécanismes sociaux s'activent automatiquement face aux agents artificiels. Par exemple, le système des neurones miroirs s'active de manière comparable lors de l'observation d'actions réalisées par un humain ou par un robot, bien qu'une mobilisation accrue des processus mnésiques soit observée pour les agents à l'apparence plus mécanique. De même, les circuits neuronaux liés aux émotions et à l'empathie présentent des schémas d'activation similaires, quoique atténués pour les agents artificiels. Les réponses physiologiques révèlent également que les normes sociales humaines relatives aux zones corporelles de faible accessibilité se transposent involontairement aux robots humanoïdes. Ces convergences pourraient suggérer que les humains traitent les robots suffisamment anthropomorphes comme des congénères humains.

Toutefois, cette conclusion serait prématurée car parallèlement à ces réactions automatiques similaires, des différences importantes subsistent. Les règles et les normes sociales ne s'appliquent pas systématiquement aux robots. Par exemple, la réticence à infliger une souffrance à un robot est bien plus faible que celle observée lorsque la victime potentielle est un humain, ce qui signale une frontière nette dans la projection de droits ou de statut moral. De même, bien que l'agression envers un robot soit jugée immorale, la réciprocité morale ne s'applique pas : un robot n'est pas perçu comme ayant le droit de se défendre.

Ainsi, les humains semblent, dans certains contextes spécifiques, réagir aux robots comme à des humains, sans toutefois les considérer de manière véritablement équivalente. L'ensemble des données converge vers l'hypothèse selon laquelle les humains ne mobilisent pas de systèmes cognitifs spécifiques à l'interaction avec les agents artificiels. Les mêmes systèmes que ceux impliqués dans les interactions entre humains seraient mobilisés, mais leurs paramètres seraient modulés lorsqu'il s'agit d'interagir avec des agents artificiels.

Chapitre 3

Présence sociale : le sentiment d'être avec un autre

3.1 Le concept de présence sociale

3.1.1 Définitions de la présence sociale

Le concept de *Présence Sociale* trouve ses origines dans les recherches pionnières de Short et al. (1976) sur la psychologie sociale des télécommunications. Ils la définissent comme (traduction) : « [...] *le degré de saillance de l'autre personne dans l'interaction et la saillance conséquente aux relations interpersonnelles* [...] » et la considèrent comme une qualité subjective (c'est-à-dire perçue) d'un médium. Ce serait une dimension perceptuelle dépendante de la capacité du médium à transmettre des indices non-verbaux (par exemple, les expressions faciales, le regard, la posture, les indices vocaux non verbaux) et à susciter chez l'utilisateur l'impression d'être en présence d'un autre. À l'époque où les auteurs proposent ce concept, la question centrale porte sur la capacité des différents médias (téléphone, vidéoconférence, etc.) à transmettre suffisamment de signaux sociaux pour que l'interlocuteur soit perçu comme « présent » : les médias qui transmettent davantage d'indices non verbaux tendent à être jugés plus « présents » que les médias audio ou écrits (Short et al., 1976). Short et al. (1976) établissent une relation entre la *Présence Sociale* et deux autres concepts qui sont l'*Intimité* (« intimacy ») et l'*Immédiateté* (« immediacy ») :

— L'*Intimité*. Selon le modèle d'Argyle et Dean (1965) elle résulte d'un équilibre entre proximité, regard mutuel, sourire et thème de conversation personnel, auquel s'ajoute selon Short et al. (1976) la présence sociale du médium de communication ;

— L'*Immédiateté*, qui est définie par Wiener et Mehrabian (1968) comme une mesure de la distance psychologique qu'un communicateur instaure avec son interlocuteur et qui se manifeste par des comportements verbaux et non verbaux d'approche. Short et al. (1976) précisent à ce sujet que certains auteurs parlent de la notion « immédiateté technologique » qui correspond au sentiment de proximité produit par un médium donné, sans que cela ne se confonde avec la *Présence Sociale*. Le premier peut varier sans que le second ne change, même si les deux évoluent ensemble.

Depuis les années 1990, l'essor des environnements virtuels, des mondes immersifs et des agents artificiels a conduit à repenser la notion de *Présence Sociale* : elle est définie comme le sentiment, pour un utilisateur, qu'un agent (humain ou artificiel) est mutuellement accessible et attentif au cours de l'interaction (Biocca et al., 2003). Elle émerge notamment via l'attribution d'intentionnalité et la mobilisation de notre capacité à inférer des états mentaux (comme la *ToM*), autrement dit à « lire dans l'esprit » de l'autre, ce qui revient dans une certaine mesure selon Biocca et al. (2003) à adopter la *Posture Intentionnelle*.

La *Présence Sociale* ne se réduit pas au sentiment d'être *là* (« sense of the place ») de la téléprésence, mais elle implique un « *sentiment d'être avec un autre* » dans un environnement (Biocca et al., 2003). Le sentiment de *Présence Sociale* dépend de la facilité avec laquelle on perçoit l'accès à l'intelligence, aux intentions et aux impressions sensorielles de l'autre (Biocca, 1997). En interaction médiée (par une technologie), le médium filtre les indices sociaux disponibles (sensorimoteurs, cognitifs, affectifs), ce qui façonne la représentation de l'autre et la *Présence Sociale* perçue. Au-delà de ce filtre, la *Présence Sociale* dépend aussi des propriétés de l'échange (par exemple, la tâche ou les objectifs relationnels) et de l'allocation attentionnelle.

Une mesure possible de la *Présence Sociale* est celle du questionnaire *Networked Minds Social Presence* (Harms & Biocca, 2004) composé de trente-six items couvrant six dimensions : coprésence, allocation attentionnelle, compréhension perçue du message, compréhension affective perçue, interdépendance émotionnelle perçue et interdépendance comportementale perçue. Les participants évaluent chaque énoncé, par exemple, « *J'ai remarqué mon partenaire* » sur une échelle de Likert en sept points (1 = « *Pas du tout d'accord* », *7* = « *Tout à fait d'accord* »).

3.1.2 Présence sociale dans l'interaction humain-robot

Dans l'interaction humain-robot, Fiore et al. (2013) ont étudié comment deux types de signaux sociaux non verbaux, à savoir, le comportement proxémique du robot et le regard (congruent, orienté vers l'humain ou variable) influencent la *Présence Sociale* perçue et les états émotionnels attribués à un robot (*iRobot, Ava mobile robotics platform*) lors d'une situation de navigation partagée en couloir. Les participants répartis selon le type de regards, effectuaient des essais où le robot alternait les comportements proxémiques (passif ou assertif) et renseignaient à deux reprises le questionnaire du *Networked Minds Social Presence* (Harms & Biocca, 2004) ainsi que la *Circular Mood Scale* (Jacob et al., 1989) pour évaluer respectivement la présence sociale ainsi que la valence et l'éveil émotionnels (*arousal*).

Les deux comportements proxémiques consistaient (1) dans la condition passive, à ce que le robot ralentisse et décentre sa trajectoire vers un côté du couloir pour laisser davantage d'espace au participant et lui céder le passage lors du croisement de trajectoires ; (2) dans la condition assertive, à ce que le robot accélère et coupe le virage pour passer devant le participant, réduisant ainsi l'espace que celui-ci occupe et affirmant sa priorité dans le cheminement.

Les résultats de cette étude montrent que le style de comportement proxémique du robot influence significativement la perception de sa *Présence Sociale* et les états émotionnels qui lui sont attribués : Le robot adoptant le comportement passif était perçu comme plus socialement présent et recevait des attributions émotionnelles de valence plus positive, tandis qu'un robot assertif suscitait des niveaux d'éveil émotionnel plus élevés. Par ailleurs, les perceptions de *Présence Sociale* et les attributions émotionnelles augmentaient au fil des interactions répétées. En revanche, la manipulation du regard n'a pas eu d'effet sur la *Présence Sociale* et n'a modulé l'intensité des émotions qu'à la marge, par le biais des interactions avec le temps et/ou la proxémie dans ce scénario précis.

Également dans le domaine des HRI, Chen et al. (2023) ont développé un cadre théorique complet spécifiquement pour la *Présence Sociale* robotique, fournissant un questionnaire en dix-sept items et cinq dimensions : la présence, la distribution de l'attention, l'expression interactive et la compréhension de l'information, l'interdépendance émotionnelle perçue et enfin la perception du comportement d'interaction. Ils proposent même une définition d'un concept

de *Présence Sociale* appliquée au robot, qu'ils nomment *Présence Sociale robotique* (traduction) : « *Dans le cas où des individus interagissent directement avec un robot social doté d'intelligence artificielle, si ces individus perçoivent des sensations identiques ou similaires à celles éprouvées lors d'une interaction avec un être humain réel, alors le robot est considéré comme ayant une présence sociale.* »

Pour ce qui est de la conception de l'apparence des robots, Damiano et Dumouchel (2018) proposent de concevoir délibérément des robots pour susciter chez l'utilisateur des projections anthropomorphiques et ainsi générer une *Présence Sociale* par le biais de deux dimensions : l'apparence humanoïde et le réalisme comportemental. Chacune de ces dimensions, lorsqu'elle atteint un niveau de réalisme élevé, suffirait à déclencher chez l'utilisateur ce sentiment (Damiano & Dumouchel, 2018). Heerink et al. (2010) montrent que la *Présence Sociale* perçue augmente lorsqu'un robot (*iCat*) adopte un comportement socialement expressif (par exemple, en employant le nom du participant, en maintenant le contact visuel et en s'excusant en cas d'erreur). En revanche, elle ne diffère pas entre une version du robot qui initie proactivement des rappels et des services selon les besoins et une version où il attend explicitement la demande de l'utilisateur.

En termes de caractéristiques anthropomorphiques, au sein de contextes simples et peu interactifs, un robot avec des yeux peut susciter une plus forte *Présence Sociale* qu'un même robot sans cette caractéristique (Kim et al., 2014 ; Kwak, 2014). Cette caractéristique chez un robot peut même accroître la propension des participants à effectuer un don (Kwak, 2014).

Chez les enfants, une étude exploratoire de Barco et al. (2020) suggère que l'interaction avec un robot à apparence anthropomorphe (*NAO*) contrôlé en mode *Wizard of Oz* induit un sentiment de *Présence Sociale* plus élevé qu'avec un robot à apparence animale (*Pleo*), mais ne diffère pas de celle suscitée par un robot à apparence caricaturale (*Cozmo*) ou anthropomorphe. Cependant, comme le soulignent les auteurs, cette perception pourrait avoir été influencée par la différence de mode d'interaction (contrôle en *Wizard of Oz* vs. une interaction libre).

Le choix de concevoir préférentiellement des robots à apparence anthropomorphe afin de renforcer la *Présence Sociale* mérite toutefois d'être nuancé. Nowak et Biocca (2003) ont observé que dans un environnement virtuel, les par-

ticipants interagissant avec une représentation virtuelle peu anthropomorphe ressentaient une *Présence Sociale* significativement plus élevée que ceux exposés à une image fortement anthropomorphe ou à l'absence d'image. Également dans un environnement virtuel et avec des agents artificiels, le réalisme comportemental élevé accroît la *Présence Sociale* (Von Der Pütten et al., 2010) tandis que la croyance d'interagir avec un avatar contrôlé par un humain plutôt qu'un agent renforce encore plus cet effet lorsque l'apparence est féminine (Guadagno et al., 2007).

Pour résumer cette section sur la *Présence Sociale*, Xu et al. (2023) montrent dans leur méta-analyse que la manipulation des indices sociaux d'un robot peut exercer un effet global de taille faible à modérée sur la *Présence Sociale*. Même si les indices sociaux, pris dans leur ensemble, semblent peu contribuer au sentiment que le robot est présent, leurs analyses par sous-groupes montrent que certains types de signaux sont très puissants pour renforcer la *Présence Sociale* perçue. En effet, la *Présence Sociale* peut être particulièrement renforcée lorsque le robot affiche des expressions faciales, notamment animées (par exemple, par l'ouverture et la fermeture des yeux ou par les mouvements de bouche) mais aussi des indices de mouvements (par exemple, des mouvements porteurs de sens communicatif, fluides et sans à-coups). Ensuite, l'utilisation d'indices haptiques (par exemple, pouvoir toucher le robot à l'épaule ou toucher sa peau étant à une température chaude) produit un effet modéré. Enfin, et à un degré moindre, la voix (par exemple, extravertie ou humaine) et le langage (par exemple, un langage littéral, un langage plus ou moins chaleureux, l'usage de *fillers* ou la prosodie) peuvent aussi avoir des effets. En revanche, contrairement à ce que l'on pourrait penser, l'apparence (par exemple, anthropomorphe ou zoomorphe) a un effet très limité sur la *Présence Sociale*.

3.2 De la reconnaissance au sentiment d'être avec un autre

La conceptualisation de la *Présence Sociale* développée dans la section précédente s'inscrit dans une tradition cognitiviste et attributionnelle de la cognition sociale : le sentiment subjectif d'être avec un autre, qu'il soit humain ou artificiel, passe, en partie, par la construction de modèles mentaux de l'autre.

Biocca et al. (2003) rattachent ce processus à la *ToM* et à la *Posture Intentionnelle*, que nous avons discutée plus tôt, et formulent l'hypothèse que la *Présence Sociale* peut être le sous-produit de la lecture et de la simulation des états mentaux de l'autre. Rappelons brièvement que, si le médium filtre les indices sociaux, la présence perçue dépend aussi de la manière dont l'échange s'organise (Harms & Biocca, 2004). C'est précisément cette dynamique, le couplage et la co-régulation entre partenaires qui nous intéresse ici.

En contraste avec cette perspective, d'autres courants proposent une reconceptualisation de la cognition sociale, en remettant en question l'idée que la perception de l'autre passe nécessairement par une modélisation interne de ses états mentaux. C'est le cas des approches interactionnistes et énactives (Varela et al., 1991) de la cognition sociale, défendues notamment par De Jaegher et al. (2010). Ce type d'approche part du processus d'interaction : quand l'ajustement réciproque (c'est-à-dire une coordination dynamique et un couplage sensorimoteur réciproque) entre deux agents s'installe et se stabilise, l'interaction forme une dynamique propre qui guide et structure ce que chacun perçoit et fait ; c'est le « *participatory sense-making* » (De Jaegher & Di Paolo, 2007). Autrement dit, une part de la cognition sociale peut se jouer dans la dynamique interactionnelle elle-même et ne réside pas uniquement « dans la tête » des individus, mais entre les individus (De Jaegher et al., 2010). Cela n'implique toutefois pas qu'une telle approche se confonde avec la *Présence Sociale* vue plus tôt dans ce manuscrit : même si certaines expériences dans l'interaction (par exemple, un sentiment de connexion ressenti pendant la coordination) peuvent contribuer à ce sentiment, l'approche énactive n'en décrit pas le vécu subjectif mais s'intéresse aux conditions dynamiques interactionnelles qui rendent possible et qui orientent la perception mutuelle.

Les propositions énactives doivent cependant être nuancées. En particulier, les arguments selon lesquels la cognition sociale s'explique en partie par la dynamique d'échange plutôt que par des mécanismes d'attribution interne ne démontrent pas que l'interaction constitue la cognition sociale : ils restent compatibles avec des mécanismes intra-individuels d'attribution et d'inférence (Overgaard & Michael, 2015). D'autre part, il faut éviter le glissement entre couplage et constitution : le fait que l'échange soit nécessaire ou structurant n'implique pas qu'il fasse partie du mécanisme cognitif proprement dit (Herschbach, 2012). De

même, l'énactivisme rend bien compte des coordinations en direct (c'est-à-dire quand on interagit), mais il semble mal traiter des usages « hors-ligne » (c'est-à-dire, des raisonnements contrefactuels et explications narratives), qui exigent des processus de représentation ; une approche qu'on pourrait qualifier d'approche « intégrative » pourrait être donc préférable (De Bruin & Kästner, 2012). Au regard de la *Présence Sociale*, il conviendrait plutôt de traiter la dynamique de l'interaction comme un facteur parmi tant d'autres (par exemple, l'attention, la tâche, la relation et le médium) plutôt que comme une explication suffisante.

En résumé, l'approche énactive spécifie les conditions dynamiques de la rencontre : une interaction est un couplage co-régulé entre agents qui peut acquérir une autonomie propre et jouer un rôle contextuel, facilitateur ou constitutif dans la cognition sociale (*participatory sense-making*). La *Présence Sociale* telle que mesurée par le questionnaire de Harms et Biocca (2004), cible le niveau subjectif et opérationnel, c'est-à-dire le sentiment d'être avec un autre, ainsi que ses dimensions qui permettent d'opérationnaliser le vécu et de le mesurer.

Pour investiguer dans quelle mesure la cognition sociale se joue en direct, dans l'interaction même, au sein de dynamiques de couplage sensorimoteur réciproques, comme le suggèrent les approches énactives et interactionnistes (De Jaegher & Di Paolo, 2007), un paradigme offrant un certain cadre expérimental minimaliste est particulièrement intéressant. Il s'agit de l'expérience de *Perceptual Crossing* développée par Auvray et al. (2009) qui examine si certains mécanismes sous-jacents à la reconnaissance d'autrui sont intrinsèques à l'interaction (qui est une activité perceptuelle partagée) elle-même.

3.2.1 Une formalisation expérimentale du Perceptual Crossing

Dans la formalisation expérimentale du *Perceptual Crossing* proposée par Auvray et al. (2009), des paires de participants (rendus aveugles quant à la position de l'un et de l'autre) interagissent à distance dans un espace unidimensionnel de 600 px et reçoivent un feedback sur l'index via un dispositif tactile de matrice braille nommé *Tactos*. Cet espace est en forme de tore (c'est-à-dire circulaire comme un anneau) pour éviter les effets liés à la présence de bords. Chaque participant contrôle un avatar de 4 px (c'est son champ récepteur) via des déplacements latéraux de la souris dans cet espace virtuel partagé. Lorsque l'avatar du participant chevauche au moins un pixel d'un quelconque « objet » dans l'envi-

ronnement, le dispositif *Tactos* fournit un feedback tactile qui persiste tant que l'avatar reste sur l'objet (c'est un feedback en tout-ou-rien). En d'autres termes, chaque participant ne perçoit qu'un seul *bit* d'information sensorielle : contact ou aucun contact. Trois types d'objets peuvent être rencontrés dans cet espace :

1. l'avatar du partenaire (qui est également un champ récepteur de 4 px). Tout chevauchement d'au moins $\geq$1 px entre les deux avatars déclenche simultanément un feedback tactile chez les deux participants ;
2. un objet fixe (large de 4 px), assigné à chaque participant et détectable uniquement par celui-ci. Ainsi, lorsque l'avatar du participant recouvre de cet objet au moins $\geq$1 px, seul ce même participant reçoit une stimulation tactile. Chaque objet fixe est placé à des coordonnées prédéterminées et diamétralement opposées sur le tore afin de favoriser l'exploration.
3. un leurre mobile (de 4 px également), attaché à une distance fixe du centre de l'avatar de chaque participant (précisément placé à une distance de 48 à 52 px de son centre), reproduisant exactement les mêmes mouvements que son avatar. Il est en quelque sorte comme une ombre.

L'expérience se divise en deux temps et commence par une phase d'entraînement avec le dispositif, puis est suivie d'une phase principale d'évaluation. Avant l'entraînement, le participant reçoit une explication sur le dispositif, au sujet du déplacement de l'avatar avec la souris, du chevauchement avec un objet et du déclenchement d'une stimulation tactile. Cependant, aucun détail n'est donné sur l'attache du leurre aux mouvements de l'avatar. La phase d'entraînement avec le dispositif comprend trois séquences de 1 min, au cours desquelles la consigne est de maintenir le contact avec la cible : (1) une première séquence avec un objet fixe de 4 px, (2) une deuxième avec un objet mobile se déplaçant à 15 px/s, et (3) une troisième avec un objet mobile se déplaçant à 30 px/s.

Ensuite, la phase d'évaluation est une séquence d'une durée de 15 min avec des pauses toutes les 5 min. Les participants, bien qu'en paire, ne sont pas dans une même pièce. Dans cette phase, la consigne est qu'ils peuvent explorer librement l'environnement et qu'ils peuvent rencontrer trois types d'objets : l'avatar de leur partenaire, des objets fixes et des objets mobiles (sans pour autant les informer sur le lien entre leurre et avatar). Il leur est demandé de cliquer (sur le bouton gauche de la souris) uniquement lorsqu'ils pensaient avoir croisé l'avatar de l'autre et de le faire chaque fois que cela se produisait. Dans cette phase,

cliquer n'envoie aucune stimulation à qui que ce soit, les clics servent seulement de mesure pour les chercheurs. Auvray et al. (2009) ont, en effet, quantifié les clics des participants pour déterminer s'ils survenaient plutôt lors de rencontres avec le partenaire (avatar) plutôt qu'avec le leurre mobile ou l'objet fixe. Il a ainsi été calculé, pour chaque type d'entité, la proportion de stimulations reçues et la proportion de clics effectués.

Les résultats d'Auvray et al. (2009) ont mis en évidence que les participants distinguaient bien les objets fixes des objets mobiles (que ce soient des leurres et des avatars), mais ils ne parvenaient pas à différencier l'avatar du leurre : la probabilité de cliquer après une stimulation liée au chevauchement avec l'un ou l'autre était équivalente. Pourtant, le nombre total de clics sur l'avatar était nettement plus élevé. Cette différence provient d'une dynamique d'interaction émergente : lorsque deux avatars se croisaient, chacun recevait une stimulation et avait tendance à inverser sa trajectoire, ce qui, sans intention consciente de coordination, créait un attracteur dynamique favorisant des rencontres répétées. Le leurre, qui ne renvoyait pas de stimulation à l'autre participant, échappait à cette boucle et était donc rencontré moins souvent. Cette dynamique se reflète dans la distribution des distances entre avatars : un pic prononcé apparaît à 0 px (c'est-à-dire quand les avatars sont superposés), correspondant aux situations de contact mutuel, indiquant que les participants passaient proportionnellement beaucoup plus de temps en perception réciproque qu'en interaction avec le leurre. Un pic secondaire, bien moins marqué, est observé autour de 50 px et correspond aux rencontres avec le leurre. Le leurre mobile, lui, ne réagit pas aux contacts et ne tendait pas à « retenir » les participants de la même manière. Ces données montrent que le surplus de clics sur l'avatar s'explique surtout par le fait que les trajectoires des deux participants s'influencent mutuellement, ce qui les amène à revenir fréquemment et à rester plus longtemps dans une position où il y a contact mutuel.

En résumé, Auvray et al. (2009) se sont interrogés sur les conditions minimales permettant de reconnaître un autre sujet percevant, en opposant l'approche individualiste et l'approche interactionniste. Ils concluent que leur dispositif minimaliste montre une sensibilité aux interactions mutuelles compatible avec l'interactionnisme, tout en laissant ouvert le rôle exact des processus individuels. Ce résultat avait conduit Auvray et al. (2009) à conclure que « *la perception d'un*

autre sujet intentionnel n'était pas basée sur une forme particulière ou sur des trajectoires objectives ; elle était basée sur des propriétés intrinsèques à l'activité perceptive conjointe elle-même ».

3.2.2 Modélisation du Perceptual Crossing par robotique évolutionnaire

Di Paolo et al. (2008) ont proposé l'une des premières modélisations du paradigme de *Perceptual Crossing* avec des agents artificiels. Au lieu de participants humains, ils ont utilisé des agents virtuels contrôlés par des réseaux de neurones récurrents à temps continu, évolués par un algorithme génétique générationnel. Leurs simulations montrent que des agents simples peuvent développer des coordinations stables pour se croiser mutuellement, étant analogues aux allers-retours observés entre humains. En plus de cette condition d'interaction en direct, ils ont simulé une interaction unilatérale dans laquelle un agent interagit avec un partenaire non contingent rejouant simplement des mouvements enregistrés lors de l'interaction en direct. Ils montrent que les agents peuvent différencier une interaction avec un partenaire réactif d'une interaction avec un partenaire non contingent simplement parce que la coordination mutuelle, lorsqu'elle est réciproque, reste stable malgré le bruit et de petits décalages temporels, alors qu'elle se dégrade rapidement avec un partenaire non réactif. Cette distinction émerge donc de la dynamique de l'interaction elle-même, sans recourir à un mécanisme interne de représentation de l'autre. Même résultat avec un second modèle où deux populations d'agents sont coévoluées de manière à faire émerger la discrimination entre l'interaction en direct et une condition rejouée : la stabilité du couplage réciproque fournit le signal pertinent, alors que l'absence de contingence fait s'effondrer la coordination.

Ces résultats ouvrent directement sur la question des conditions dynamiques, telles que la contingence, qui rendent possible la reconnaissance d'un autre comme agent percevant. Si cette différenciation repose sur la stabilité du couplage réciproque, elle doit laisser des traces mesurables dans la dynamique temporelle de l'interaction et, corrélativement, dans le vécu subjectif d'être avec un autre.

3.2.3 Organisation multi-échelle de l'interaction dans le Perceptual Crossing

Récemment, Bedia et al. (2014) ont proposé une approche quantitative pour analyser les différences entre les interactions humain-humain et humain-agent dans le paradigme du *Perceptual Crossing*. Ceux-ci partent d'une critique méthodologique des analyses classiques du paradigme et indiquent que les études antérieures (notamment Auvray et al., 2009) se concentrent sur des indices à court terme (comme la probabilité de clic dans une fenêtre de 2 s et la fréquence des stimulations) en partant implicitement du principe que l'engagement social peut être réduit à une échelle temporelle dominante ou à une seule durée. Ils estiment que restreindre l'analyse à une seule grille de temps risque de manquer la structure et la complexité multi-niveaux de l'interaction sociale. Ils avancent ainsi que les interactions humain-humain se caractérisent par une organisation multi-échelles, où les comportements des participants s'imbriquent sur plusieurs niveaux temporels : micro-ajustements rapides, régularités à moyen terme et motifs de coordination à long terme. Une telle imbrication générerait des corrélations à long terme observables dans les dynamiques des deux agents.

Ils ont donc développé une version du paradigme de *Perceptual Crossing* qui, contrairement au dispositif d'Auvray et al. (2009) contenant trois types d'objets (avatar partenaire, objet fixe, leurre mobile reproduisant les mouvements en temps réel), ne présente ici qu'un seul type de partenaire à chaque essai. Dans l'étude de Bedia et al. (2014), un participant pouvait donc être confronté soit à un autre humain, soit à un agent oscillatoire dont le déplacement suivait une trajectoire sinusoïdale prédéterminée (à une fréquence de 0.5 Hz et une amplitude de 200 px) ou soit un agent dit « *shadow* » reproduisant exactement les mouvements du participant lui-même, mais en asynchrone avec un délai de 400 ms et décalé de 125 px. Cet agent n'est pas du tout l'équivalent à un leurre mobile. L'environnement reste un espace virtuel unidimensionnel fermé de 800 px (donc légèrement plus grand que chez Auvray et al., 2009), en forme de tore et sur lequel l'avatar est contrôlé via la souris d'ordinateur. Dans l'expérience d'Auvray et al. (2009), la stimulation était tactile et durait le temps du chevauchement mais Bedia et al. (2014) utilisent quant à eux, une stimulation auditive sous la forme d'un bip sonore durant 500 ms et retentissant à chaque chevauchement.

L'expérience comprend deux phases. Tout d'abord, une phase d'entraînement composée de trois essais de 1 min avec des agents de complexité croissante (statique, mobile lent et mobile plus rapide). Ensuite, une phase d'évaluation composée de dix essais de 40 s. À chaque nouvel essai, un agent est aléatoirement attribué et le participant doit indiquer si son partenaire était humain ou non à la fin de chaque essai.

Les positions des avatars et des agents sont enregistrées en continu. A partir de ces données, la distance entre eux a pu être calculée à chaque instant, puis la vélocité relative peut être obtenue en extrayant la dérivée de cette distance. La vélocité relative correspond à la vitesse à laquelle les participants se rapprochent ou s'éloignent l'un de l'autre. Cette variable de relation capture la dynamique émergente de l'interaction et non pas les comportements individuels.

Ensuite, pour caractériser la structure des corrélations à travers différentes échelles temporelles, ces séries temporelles de vélocité relative sont soumises à : (1) une *Detrended Fluctuation Analysis* (*DFA*) pour estimer l'exposant β, indicateur de la présence de corrélations à long terme de type bruit *1/f*, et à (2) une *Multifractal Detrended Fluctuation Analysis* (*MDFA*) pour mesurer la largeur du spectre Δh, indicatrice de la diversité des échelles temporelles impliquées. Cette méthode permet de caractériser la structure fractale des séries temporelles et de détecter les motifs *1/f* caractéristiques des systèmes complexes où différentes échelles temporelles s'influencent mutuellement. Le spectre fractal devient ainsi un indicateur quantitatif de l'organisation multi-échelles : sa présence signale l'existence de corrélations s'étendant à travers différentes échelles temporelles. L'hypothèse était que seules les interactions humain-humain présenteraient cette organisation multi-échelles, contrairement aux interactions avec des agents artificiels qui devraient présenter des motifs temporels plus simples.

Les résultats de Bedia et al. (2014) révèlent des signatures temporelles distinctes selon le type d'interaction. En effet, les interactions humain-humain présentent une structure fractale $1/f$ (β proche de 1) dans la vélocité relative : les fluctuations à court terme restent corrélées avec celles à plus long terme, suggérant une organisation multi-échelles des dynamiques d'interaction. En revanche, les interactions avec l'agent oscillatoire présentent un motif de type

bruit brun (β proche de 1.5), reflétant une dynamique plus rigide et contrainte par le mouvement cyclique préprogrammé. Les interactions avec l'agent « *shadow* » montrent un motif proche du bruit blanc (β proche de 0), indiquant une décorrélation temporelle plus marquée.

L'analyse multifractale confirme ces différences : seules les interactions entre humains présentent un spectre multifractal large, indiquant une riche variabilité dans l'organisation temporelle à différentes échelles. Toutefois, dans les interactions avec l'agent « *shadow* », le spectre est d'une complexité intermédiaire, plutôt proche de celle de l'interaction humain-humain. Ceci étant probablement lié à la dépendance directe aux mouvements du participant. Cette signature fractale discriminante est observable principalement dans la vélocité relative et non dans les variables individuelles (les mouvements isolés du participant ou du partenaire) confirmant que l'organisation multi-échelles émerge spécifiquement de la dynamique conjointe de l'interaction. Ces observations soutiennent l'idée que les interactions entre humains dans cet espace unidimensionnel se distinguent par des corrélations temporelles complexes fournissant un indicateur quantitatif pour différencier ces dynamiques de celles induites par des agents artificiels. Leur analyse suggère que se baser uniquement sur des mesures « à court terme » peut manquer des différences plus subtiles, qui se révèlent à plus longue échelle temporelle et dans la structure multi-échelles de l'activité conjointe. Néanmoins, ces analyses comportementales et de dynamiques, ne renseignent qu'indirectement sur l'expérience vécue des participants.

3.2.4 Clarté de la présence de l'autre et réussite conjointe de l'interaction

Froese et al. (2020) ont mené une expérience utilisant une version modifiée du *Perceptual Crossing* de Auvray et al. (2009) dans laquelle la dynamique d'échange est reliée à une évaluation explicite de l'expérience subjective. Cette évaluation concerne la clarté avec laquelle le participant perçoit la présence de l'autre. Durant vingt essais de 60 s (au lieu d'un unique bloc continu de 15 min), chaque participant devait cliquer une seule fois par essai au moment où il estimait interagir avec l'autre, puis évaluer, après l'essai, la clarté de cette expérience sur une échelle (*Perceptual Awareness Scale, PAS*) en quatre niveaux (de

« aucune expérience » à « expérience claire ») adaptée à la perception sociale. Ce protocole comportant un clic unique par essai et l'échelle *Perceptual Awareness Scale*, avait été introduit pour la première fois dans une étude antérieure (Froese et al., 2014). Il permet de relier la dynamique d'échange à l'expérience subjective d'être avec l'autre, tout en conservant les mêmes types d'objets (avatar, leurre, objet fixe) que chez Auvray et al. (2009). D'autre part, pour renforcer l'engagement des participants et standardiser la prise de décision (qui correspond à un clic unique par essai), une consigne de performance a également été ajoutée. Les participants étaient informés qu'ils formaient une équipe et qu'ils gagneraient ou perdraient des points en fonction de la justesse de leurs clics, sans toutefois recevoir de retour en direct.

Ainsi, les résultats montrent que la clarté de la perception sociale (mesurée par le score à l'échelle *Perceptual Awareness Scale*) n'augmente pas quand un participant fait un clic correct isolé (c'est-à-dire quand il reconnaît l'autre), mais elle augmente lorsque les deux partenaires cliquent correctement dans le même essai (cette situation marque un succès conjoint à se reconnaître l'un et l'autre), souvent dans une fenêtre temporelle de quelques secondes (moins de 3 s). Froese et al. (2014) en concluent que l'augmentation de la clarté de la perception, de l'expérience subjective, ne peut pas être expliquée par l'activité d'un seul individu, mais doit être comprise comme le résultat d'un processus d'interaction qui est distribué : le sentiment d'être avec l'autre dépend de la réussite conjointe de l'interaction, et non de la performance individuelle seule.

Cette conclusion prolonge directement les résultats d'Auvray et al. (2009), qui montraient que la réussite au niveau comportemental émergeait d'un couplage stable, en ajoutant ici que la perception subjective de cette présence est elle aussi sensible à la dynamique dyadique. En d'autres termes, les changements observés dans l'expérience subjective, ici une perception plus ou moins claire de la présence de l'autre, ne peuvent être expliqués au niveau individuel. Ces changements dans l'expérience vécue sont mieux comprises au niveau dyadique, ce qui rejoint les prédictions interactionnistes et énactives : certaines dimensions de l'expérience sociale semblent émerger de la coordination entre les agents, et ne sont pas réductibles à un traitement cognitif localisé dans un seul individu. De plus, les moments de reconnaissance mutuelle correcte tendent à se synchroniser dans une fenêtre temporelle de l'ordre de 3 s, ce qui est cohérent

avec les échelles de temps proposées par Varela et al. (1991) pour l'intégration neuronale sous-tendant la conscience du moment présent. En effet, Varela et al. (1991) soulignent le moment conscient du « maintenant » comme une fenêtre vécue d'environ 2 à 3 s, liée à une dynamique d'intégration-relaxation neurale. C'est le temps durant lequel des assemblées de neurones se synchronisent pour former un acte conscient complet.

3.2.5 Reconnaissance de l'humain au travers des contingences réciproques

Par ailleurs, dans le prolongement des travaux de Bedia et al. (2014), Barone et al. (2020) ont mené deux études visant à examiner si les participants peuvent recourir aux contingences sensorimotrices réciproques (c'est-à-dire quand les actions d'un individu modifient ce qu'il perçoit et que l'autre individu adapte aussi ses actions à celles du premier en retour) pour détecter une interaction et distinguer les agents humains des agents artificiels sur cette base.

Dans ce protocole, les participants faisaient face à trois types d'agents : un autre participant humain, un agent oscillatoire, et un agent humain (condition « offline ») rejouant l'enregistrement d'une interaction humain-humain précédente où les participants se sont correctement reconnus comme humains.

L'expérience était divisée en deux phases : une phase d'entraînement et une phase expérimentale. La phase d'entraînement était une phase de familiarisation permettant aux participants de se repérer dans l'interface. Elle comprend quatre essais de 15 s chacun. Tandis que la phase expérimentale était composée de trois blocs de neuf essais de 30 s chacun. À la fin de chaque essai, les participants devaient dire avec qui ils avaient interagi : « machine » ou « personne » et recevaient un retour leur indiquant si la réponse donnée était correcte ou incorrecte.

Les deux types de feedback lors d'un croisement étaient soit audio (via un bip de 500 ms, comme chez Bedia et al., 2014) ou audiovisuel (via un bip, des avatars visibles et la ligne de l'environnement visible). Chaque bloc faisait varier le type de feedback reçu par le participant lors d'un croisement (bloc 1 : audio, bloc 2 : audio et visuel, bloc 3 : audio). À la fin de la procédure, les participants répondent à trois questions : ils devaient décrire comment ils avaient « joué »,

comment ils avaient décidé qu'ils avaient interagi avec un humain et comment ils avaient décidé qu'ils avaient interagi avec une machine. Les auteurs ont ensuite classé ces réponses en trois catégories : basée sur la réciprocité, partiellement basée sur la réciprocité, ou non réciproques.

Les résultats montrent que les participants n'ont pas détecté la contingence réciproque. L'agent oscillatoire était correctement identifié comme « machine » dans tous les blocs. Le partenaire humain est jugé comme une « personne » en blocs 1 et 2, mais le taux de réponse correcte ne dépasse pas le hasard en bloc 3. Cependant, l'agent « *offline* » est majoritairement classé comme « personne » même dans les blocs où le feedback est audiovisuel.

Concernant l'analyse de la structure des corrélations, contrairement à Bedia et al. (2014) qui rapportaient un $\beta \approx 0.86$ en interaction humaine « en direct », $\beta \approx 0.29$ en condition « *shadow* », et $\beta \approx 1.4$ pour l'agent oscillatoire, Barone et al. (2020) ne retrouvent pas de valeurs similaires, à l'exception de la condition « offline » ($\beta \approx 0.20$) relativement similaire à la condition « *shadow* » de Bedia et al. (2014). En effet, les β, tous très proches de 0, sont extrêmement similaires entre les agents, et ce, surtout dans les conditions « en direct » et « *offline* », donc sans signature $1/f$ discriminante. Pour les indices fractals spécifiquement, les auteurs suggèrent que la variation de feedback entre blocs pourrait avoir perturbé l'espace partagé nécessaire à l'émergence de la signature $1/f$. Ajouté à cela, les autres mesures implicites n'isolent aucune signature propre dans la comparaison entre la condition « en direct » et « *offline* » : les trajectoires et les nombres de collisions ne permettent pas de discriminer les conditions. Toutefois, les analyses de corrélations révèlent que les contingences semblent opérer dans des fenêtres de ≈500 ms à 1 s, suggérant des échelles temporelles courtes pour la détection d'interaction et cohérentes avec les échelles temporelles de la conscience du moment présent. Le débriefing avec les participants révèle que les décisions semblent guidées par l'heuristique selon laquelle un mouvement périodique indique une machine et un mouvement non-périodique indique une personne. Les auteurs attribuent ce profil au cadrage de la tâche et au feedback essai-par-essai, qui ont encouragé une stratégie observationnelle ou heuristique plutôt qu'une détection active de réciprocité.

Face à ces résultats, Barone et al. (2020) ont conçu une seconde étude pour tester l'hypothèse selon laquelle les participants se basaient sur l'observation

des motifs de mouvement plutôt que sur la détection active des réponses mutuelles. Ils ont apporté plusieurs modifications méthodologiques :

- réduction du nombre d'essais, pour passer de neuf à six essais par bloc ;
- suppression de l'agent oscillatoire, pour ne garder que l'agent humain « en direct » et « *offline* » ;
- reformulation de la question posée au participant, pour plutôt demander s'il avait perçu que l'interaction avait eu lieu « en direct » ou « *offline* » ;
- remplacement du choix binaire de réponse à cette question par une échelle de Likert à 7 points indiquant le degré de certitude (« [...] complètement sûr que c'était en direct » ou « [...] complètement sûr que c'était *offline* »),
- et enfin, suppression du feedback suite à leur réponse à cette question.

Cette conception visait à forcer les participants à se concentrer sur la détection des contingences réciproques (c'est-à-dire des réponses mutuelles). Les deux types de trajectoires étant équivalentes en complexité et en imprévisibilité d'un point de vue observationnel.

Les résultats de cette étude variaient selon le type de feedback (bloc). En effet, en condition audiovisuelle (bloc 2), les participants ont correctement identifié l'agent « en direct » tandis que l'agent « *offline* » était reconnu non significativement différent du hasard. En revanche, dans les blocs audio uniquement (les blocs 1 et 3), la performance était similaire au hasard pour les deux types d'agents. L'analyse des réponses pour lesquelles les participants se déclaraient « complètement sûrs » renforce ces conclusions : la reconnaissance correcte des agents « en direct » et « *offline* » n'émergeait qu'en bloc 2 audiovisuel. Seul le bloc prédisait la justesse de leur réponse, mais ni le type d'agent, ni les corrélations ou collisions ne la discriminent.

Les modifications de méthodologie ont effectivement induit un changement de stratégie chez les participants : 65% ont rapporté des décisions basées sur la réciprocité (contre 11% dans la première étude), avec 25% des décisions partiellement réciproques et 10% des décisions non-réciproques. Cependant, les indices fractals demeuraient proches de 0 pour tous les agents, sans reproduction de la signature $1/f$ de Bedia et al. (2014). Les auteurs concluent que la détection des contingences sensorimotrices réciproques semble possible mais nécessite une information multimodale suffisante, ici audiovisuelle. L'information

auditive seule est insuffisante pour distinguer de manière fiable une interaction contingente d'une reproduction non-contingente d'interaction préalablement enregistrée.

3.3 Émergence du sentiment de présence dans la relation

Comme cela a été évoqué dans les sections précédentes de ce chapitre, la *présence sociale* est le sentiment d'être avec un autre. Elle est fondée sur l'accessibilité mutuelle, l'attention conjointe et l'attribution d'intentions (Biocca, 1997 ; Biocca et al., 2003). Elle est contrainte par le médium qui filtre les indices sensorimoteurs, cognitifs et affectifs disponibles (Biocca, 1997). Ce sentiment d'être avec un autre semble dépendre davantage du réalisme et de la contingence des comportements que de l'apparence seule (Damiano & Dumouchel, 2018 ; Nowak & Biocca, 2003 ; Xu et al., 2023). La proxémie du robot modulerait ce sentiment et les émotions attribuées (Fiore et al., 2013). Les leviers les plus efficaces pour susciter ce sentiment semblent être les expressions faciales animées, le mouvement expressif fluide et l'haptique plutôt que la seule voix ou l'apparence (Xu et al., 2023).

Les travaux présentés dans cette section convergent pour suggérer que la reconnaissance d'un autre comme agent percevant ne repose pas sur des caractéristiques formelles ou des trajectoires objectives, mais sur la maîtrise de contingences sensorimotrices réciproques qui ne peuvent être pleinement saisies qu'au niveau de la dyade qui interagit. Cette convergence entre analyses dynamiques, mesures de la clarté de la perception et capacités de discrimination explicite renforce l'hypothèse que la cognition sociale, du moins dans ses manifestations les plus élémentaires, ne peut être entièrement comprise en se limitant aux mécanismes intra-individuels d'attribution et d'inférence. Elle suggère que certaines propriétés de l'expérience sociale émergent de la coordination elle-même.

En résumé, le Chapitre 3 établit que la cognition sociale et être avec un autre découlent d'abord de la contingence et de la coordination réciproques, plus que de l'apparence. Ce vécu, s'inscrivant fondamentalement dans la dyade avec le partenaire et étant multi-échelles, est sensible à des fenêtres de moins d'une

seconde pour détecter la réciprocité et à des fenêtres de quelques secondes pour la reconnaissance mutuelle, avec des signatures d'interaction de type *1/f* rapportées, mais encore débattues. Dans la continuité de l'investigation des aspects temporaux des interactions, le chapitre suivant les abordera plus précisément dans le cadre du dialogue et de la communication verbale humain-robot.

Chapitre 4

Discours et temporalité dans l'interaction humain-robot

4.1 Temporalité, tour de parole entre humains

4.1.1 Fluidité universelle du tour de parole

Les interactions verbales entre humains sont caractérisées par une étonnante fluidité temporelle où deux interlocuteurs se parlent en général sans se couper excessivement la parole ni laisser de longs blancs entre leurs tours de parole : cette alternance des tours (*turn-taking*) dans la conversation humaine avec des transitions rapides entre locuteurs semble être presque universelle (Levinson & Torreira, 2015 ; Stivers et al., 2009). Une étude de Stivers et al. (2009) couvrant dix langues a montré que la plupart des conversations à travers le monde tendent à éviter les silences prolongés et les chevauchements, avec des écarts moyens d'environ 200 ms (avec de légères variations culturelles) (Stivers et al., 2009). Autrement dit, le délai entre la fin du tour d'un locuteur et le début de la réponse de l'autre est d'environ 200 ms. C'est-à-dire l'équivalent de la durée d'un clignement d'œil (Cruz et al., 2011 ; Gordon, 1951).

Cette synchronisation est d'autant plus remarquable que la planification d'un énoncé (c'est-à-dire trouver ses mots ou structurer sa phrase) prend typiquement au moins 600 ms (Stivers et al., 2009). Comment parvenons-nous alors à répondre si vite, souvent quasiment en synchronie avec la fin de la phrase de l'autre ? La clef réside dans des mécanismes prédictifs sophistiqués : le cerveau anticipe la fin du tour de l'autre pendant l'écoute même du discours en utilisant des indices syntaxiques, comme la structure de la phrase (qui laisse deviner quand elle va se terminer), des indices sémantiques-pragmatiques comme

le sens probable de la fin de l'énoncé et des indices prosodiques tels que l'intonation de fin de phrase (Ekstedt & Skantze, 2020 ; Garrod & Pickering, 2015). D'autres indices, tels que l'intonation descendante ou la complétude syntaxique, indiquent une possible fin de tour (Gravano & Hirschberg, 2011). En d'autres termes, nous commençons souvent à préparer notre réponse avant même que l'autre ait complètement fini de parler, si bien que nous sommes prêts à enchaîner presque instantanément lorsqu'il achève son propos.

4.1.2 Variabilité et signification sociale des délais

La structure du dialogue joue néanmoins un rôle : certains contextes admettent des délais un peu plus longs. Par exemple, selon Strömbergsson et al. (2013), les questions ouvertes qui demandent une réflexion (par exemple « *Pourquoi… ?* ») entraînent des réponses avec un temps de latence plus élevé (≈300-450 ms en moyenne), alors que les questions fermées oui/non ou à choix (qui préstructurent la réponse) peuvent obtenir des réponses plus rapides (≈100-180 ms)

De même, cette orchestration temporelle de la conversation humaine repose sur un ensemble d'indices multimodaux, en plus du langage lui-même. Par exemple, lorsqu'un locuteur regarde son interlocuteur en finissant sa phrase, il signifie qu'il s'agit du tour de ce dernier. Des études classiques (Duncan, 1972 ; Kendon, 1967) ont décrit ces signaux de régulation du tour de parole tels que le regard soutenu, un hochement de tête, ou au contraire l'évitement du regard si l'on veut maintenir son tour.

Au-delà de leur rôle structurel, les délais de réponse dans la conversation portent une signification sociale et peuvent influencer la perception au sein des conversations. Par exemple, des délais de réponse rapides (moins de 250 ms) sont fortement associés à des sentiments accrus de connexion sociale (Templeton et al., 2022). En revanche, un long délai avant de répondre oriente l'auditeur à inférer une issue défavorable (par exemple, un refus, une critique ou un désaccord) et à choisir l'interprétation la plus négative d'un énoncé poli (Bonnefon et al., 2015). Aussi, des pauses plus longues avant de répondre à une question de connaissance factuelle diminuent la compétence perçue du locuteur (Matzinger et al., 2023).

Les humains semblent s'adapter aux caractéristiques perçues de leurs inter-

locuteurs. Casillas et al. (2016) ont constaté que dans des interactions de type question-réponse entre adultes et enfants, lorsque les enfants posent des questions, les adultes affichent des latences de réponse médianes d'environ ≈371 ms (c'est-à-dire nettement plus longues que dans les conversations entre adultes ; Stivers et al., 2009), tandis que lorsque les adultes posent des questions, les enfants affichent des latences médianes plus longues d'environ 625 ms. De même, Ervin-Tripp (1979) a observé que malgré les délais importants des enfants pour prendre la parole, les conversations restent fluides car les adultes s'adaptent inconsciemment à ces temps de réponse plus longs. Il se pourrait donc que les humains développent des attentes temporelles spécifiques pour différentes catégories d'interlocuteurs, telles que les enfants ou potentiellement les robots sociaux en tant que catégorie générale.

4.2 Systèmes de dialogues et limites dans l'interaction humain-robot

4.2.1 Contraintes techniques et modèles prédictifs

Cette orchestration temporelle précise entre humains est susceptible de poser des défis majeurs pour la robotique sociale puisque l'alternance des tours est un aspect fondamental du dialogue. Un tel défi temporel devient encore plus prononcé dans les interactions où les robots doivent coordonner la parole et les actions motrices simultanément (Hough & Schlangen, 2016). De nombreux systèmes de dialogue s'appuient sur des seuils de silence fixes (par exemple, 600 ms dans Cuijpers et Van Den Goor, 2017) et peinent à reproduire la fluidité naturelle des échanges humains (Skantze, 2021). Ces limites des systèmes de dialogue vocal peuvent mener à des pauses artificielles ou à des interruptions, ce qui perturbe considérablement le flux de l'interaction (Edlund & Heldner, 2005 ; Raux & Eskenazi, 2008). Pour répondre aux limites de tels systèmes, des travaux tels que Skantze et Irfan (2025) combinent des modèles de *turn-taking* entrainés sur des données humain-humain, notamment *TurnGPT* (Ekstedt & Skantze, 2020) qui modélise la complétude syntaxico-pragmatique à partir du texte ou encore *VAP* (Ekstedt & Skantze, 2022) qui projette l'activité vocale future à partir du signal audio et capte des indices prosodiques. Une telle approche leur a

permis d'atteindre des délais relativement rapides et un taux d'interruption réduit comparé à un système à seuil fixe (mode de 600 ms vs. 2600 ms, moyenne de 1500 ms vs. 2700 ms ; taux d'interruptions de 6.9% vs. 16.6%).

4.2.2 Incarnation, signaux visuels et attentes temporelles

Un facteur important influençant la coordination du *turn-taking* est l'incarnation physique du robot. Dans le domaine de la HRI, la présence de signaux visuels naturels, tels que l'orientation de la tête et les expressions faciales, peut permettre des interactions plus fluides par rapport à des systèmes uniquement vocaux, comme le souligne la revue de Skantze (2021). Si des études sur l'interaction entre humains ont montré que des indices tels que l'intonation descendante et la fermeture syntaxique constituent des indicateurs fiables de la fin d'un tour de parole (Gravano & Hirschberg, 2011), l'intégration d'indices visuels devient particulièrement importante dans l'interaction avec des robots. Par exemple, Cuijpers et Van Den Goor (2017) ont constaté que dans les interactions humain-robot, les signaux naturels de tour de parole, tels que les mouvements de la tête et des gestes des bras entraînaient des temps de réponse plus rapides que les signaux artificiels (comme des clignotements de LED) ou un simple silence. Cependant, cet avantage ne se produisait que lorsque les signaux étaient fournis immédiatement à la fin de l'énoncé du robot (c'est-à-dire 0.6 s avant le seuil de pause naturel établi dans leur étude). Les signaux retardés au-delà de la pause naturelle perturbaient la fluidité de l'interaction, avec un effet plus prononcé pour les signaux naturels saillants. L'absence ou la qualité dégradée de certains indices visuels, en particulier la direction fine du regard en trois dimensions, nuit considérablement à la précision de la prise de parole et augmente les temps de réponse, avec des différences moyennes de 470 ms entre les conditions deux dimensions et trois dimensions (Al Moubayed & Skantze, 2011).

Shiwa et al. (2009) ont exploré la vitesse à laquelle un robot doit nous répondre (sans varier de cadrage social comme le statut du robot, son registre de politesse, son style de communication, etc.) en s'interessant au temps de réponses du système (*SRT; System Response Time*). Ils ont comparé la préférence des utilisateurs pour quatre délais (0, 1, 2 et 3 s) dans une discussion avec un robot humanoïde (*Robovie*) et celle dans des opérations sur une interface utilisateur graphique (GUI). Pour la GUI, les scores de préférence ont diminué de

façon monotone à mesure que le délai augmentait : 0 s (instantané) étant le temps de réponse le plus apprécié. Tandis qu'avec le robot, la préférence était élevée également à 0 s, mais atteignait un maximum net à 1 s avant de chuter drastiquement dès 2 s. Les auteurs ont également testé des délais plus longs (de 3 à 9 s), avec l'ajout de *fillers* conversationnels et sans cet ajout (verbaux pour le robot et un symbole de sablier pour la GUI) afin de voir si leur présence allaient atténuer la baisse de préférence associée aux longs délais. Shiwa et al. (2009) ont ainsi pu déterminer que les impressions négatives des utilisateurs concernant un SRT long pouvaient être atténuées par la présence de *fillers* (cet effet est surtout prononcé pour le robot), mais les préférences continuaient quand même à diminuer à mesure que le délai augmentait. Ils ont également examiné l'effet de l'habituation à un délai (0 ou 1 s) sur les préférences et ont montré que cette exposition modifiait le SRT préféré avec le robot. En effet, une habituation à un délai d'1 s conduit à préférer ce délai et à moins apprécier la réponse instantanée (0 s), tandis qu'une habituation à 0 s maintient une préférence similaire pour 0 s et 1 s, avec toutefois une baisse marquée dès 2 s.

Ces résultats établissent une idée de référence en matière de rapidité de réponse d'un robot (différente de l'humain-humain), tout en laissant ouverte la question de savoir si la façon dont le robot s'exprime modifie ce qui est considéré comme un délai de réponse optimal ou acceptable. En effet, La position d'autorité comme stratégie de communication d'un robot influence notre perception de celui-ci (notamment à travers la manière dont le rôle est verbalement mis en scène). Par exemple, Saunderson et Nejat (2021) ont constaté que, contrairement aux attentes dérivant des interactions entre humains où une plus grande autorité est généralement associée à une plus grande force de persuasion, les robots occupant des rôles d'autorité étaient nettement moins persuasifs et suscitaient davantage d'attitudes négatives que les robots occupant des rôles de pairs.

Les humains tendent à projeter spontanément des traits de personnalité sur les agents robotiques (Fong et al., 2003) mais l'impact de ces attributions sur les attentes temporelles reste mal compris. La personnalité projetée sur un robot pourrait influencer davantage la dynamique temporelle de deux manières : (1) soit en modulant ce qui constitue le délai de réponse optimal lui-même (par exemple, on peut s'attendre à ce qu'un robot autoritaire réponde plus rapide-

ment qu'un robot enfantin), (2) soit en affectant la tolérance des utilisateurs à l'égard des écarts par rapport à l'optimum (par exemple, un robot soumis ou enfantin pourrait susciter une plus grande acceptation des réponses retardées). Ainsi, la capacité d'adaptation temporelle observée entre humains à différentes catégories d'interlocuteurs (voir sous-section 4.1.2) suggère que les humains pourraient également développer des attentes temporelles spécifiques aux robots sociaux. Pour rappel, les délais de réponse ne sont en général pas « neutres » et sont interprétés avec des significations sociales (Bonnefon et al., 2015; Matzinger et al., 2023; Templeton et al., 2022). Ainsi, dans l'interaction avec un robot, la sensibilité humaine aux dynamiques temporelles de la conversation, où des variations de l'ordre de 100 à 200 ms peuvent déjà influencer la perception de fluidité (Stivers et al., 2009), suggère qu'une modulation subtile des délais de réponse, combinée à différentes « personnalités » de robots (c'est-à-dire à leurs caractéristiques et à leur classe d'interlocuteur), pourrait affecter la perception de la fluidité de l'interaction.

4.3 Réaction cérébrale au discours des robots

4.3.1 Vigilance et confiance

Les humains, lorsqu'ils communiquent entre eux, disposent d'une suite de mécanismes cognitifs visant à se prémunir contre le risque d'être mal informés, que ce soit par erreur ou par tromperie (Sperber et al., 2010). Ces mécanismes dits de vigilance épistémique portent à la fois sur la source, en évaluant la compétence ainsi que la bienveillance de l'informateur, et ces mécanismes portent également sur le contenu de l'information en évaluant la crédibilité intrinsèque de l'énoncé et sa cohérence (Sperber et al., 2010). Dès le jeune âge, c'est-à-dire vers quatre ans, les enfants peuvent identifier qui sont les informateurs fiables et s'y fient en l'absence d'information vérifiable. Ils peuvent néanmoins s'en tenir à leur propre perception en cas de contradiction (Clément et al., 2004). De même, ils tendent à savoir repérer la personne pertinente selon l'expertise (Aguiar et al., 2012; Vanderborght & Jaswal, 2009) et pour apprendre de nouveaux mots, ils vont privilégier l'informateur le plus fiable, même si c'est un pair plutôt qu'un adulte (Jaswal & Neely, 2006). Dès seize à dix-huit mois, les bébés portent davan-

tage d'attention aux adultes qui nomment des objets correctement qu'à ceux qui les étiquettent de manière erronée (Koenig & Echols, 2003).

En principe, face à un robot social qui nous adresse la parole, nous devrions mobiliser les mêmes mécanismes que ceux utilisés dans l'échange avec un humain. Autrement dit, nous interprétons d'abord ses propos sous le régime d'une confiance provisoire (Sperber et al., 2010). Parallèlement, nous évaluons à la fois la source et le contenu du message avant de l'accepter, de le rejeter ou de suspendre notre jugement, tout en ajustant en conséquence notre confiance envers l'interlocuteur (Sperber et al., 2010). Il est possible que les mécanismes et réactions évoqués précédemment, tels que l'anthropomorphisme, les heuristiques sociales (*CASA*) ou la mentalisation, puissent ainsi orienter la manière dont nous considérons à la fois le discours et le robot qui le produit.

Cependant, il semble que l'on puisse accorder foi à une machine dans certains domaines ou pour certaines tâches, parfois même davantage qu'à un humain. Par exemple, à qualité égale, des participants tendent à accorder plus de crédit à un conseil présenté comme émanant d'un algorithme qu'à celui attribué à une personne (Logg et al., 2019). De la même manière, les individus semblent plus enclins à communiquer leurs informations bancaires à une machine qu'à un agent humain (Sundar & Kim, 2019). Cependant, cette confiance accordée aux algorithmes semble dépendre du type de tâche : les individus ont tendance à éviter leurs conseils pour les tâches perçues comme subjectives. En revanche, présenter la tâche comme objective ou amener les participants à percevoir les algorithmes comme plus « humains » accroît leur confiance et leur propension à les utiliser (Castelo et al., 2019). D'autre part, voir un algorithme faire des erreurs fait chuter la confiance et l'usage, au point que les gens préfèrent un humain moins précis même après avoir constaté que l'algorithme le surpasse (Dietvorst et al., 2015).

Du côté des interactions avec les robots sociaux, il est observé des tendances similaires à celles retrouvées entre humains mais variant plus spécifiquement selon le contexte d'interaction et l'âge. En effet, les humains, en particulier les enfants, peuvent se laisser influencer considérablement par les propos d'un robot, parfois même autant que par un humain. En effet, en reproduisant l'expérience de Asch (1956) avec des robots, Vollmer et al. (2018) ont montré que les adultes résistent à la pression de se conformer exercée par un groupe de robots

(comparé à des humains), contrairement aux enfants de sept à neuf ans qui s'y conforment. Qin et al. (2022) ont voulu essayer une autre configuration de cette expérience, plus crédible socialement, où les complices sont un groupe principalement d'humains mais avec un seul robot (donc en minorité) comparativement à un groupe composé seulement d'humains : les adultes se conformaient bien au groupe avec un robot et lorsque le robot était dissident en brisant l'unanimité du groupe, il pouvait réduire la conformité mais moins que si le dissident complice était humain. Dans une tâche où il n'y a pas vérité objective, Salomons et al. (2018) ont montré que les adultes se conforment aux jugements d'un groupe de robots environ un tiers du temps tant qu'ils les jugent compétents, mais cessent de le faire quand leur tendance à être inexacte est démontrée. Pour ce qui est de la confiance chez les enfants, surtout les plus âgés, celle-ci chute plus vite que chez les adultes après des erreurs du partenaire (humain ou robot), et seule l'excuse du robot (pas de l'humain) ralentit cette baisse (Flanagan et al., 2024). Dans une tâche où il est demandé de localiser un jouet, même si les jeunes enfants choisissent de demander de l'aide plus souvent à un robot (qui a démontré sa compétence) qu'à un adulte imprécis, ils n'adopteront pas pour autant plus la réponse du robot (Baumann et al., 2024). Dans une tâche d'apprentissage de mots, les enfants de trois à cinq ans font confiance aux informateurs ayant été exacts par le passé, qu'ils soient humains ou robots (Li et al., 2023). Cependant, face à des informateurs ayant été inexacts par le passé, les plus jeunes font davantage confiance à l'humain inexact qu'au robot inexact, tandis que les plus âgés se méfient de l'informateur inexact qu'il soit robot ou humain. (Li et al., 2023).

En somme, face à l'information délivrée par les robots, nos mécanismes cognitifs tendent d'abord à les considérer d'une façon similaire aux humains, en leur accordant une confiance provisoire que l'on ajuste ensuite selon leur compétence perçue, leur historique d'exactitude et la tâche. Toutefois, cette confiance reste plus fragile et contextuelle qu'avec des humains : l'erreur pénalisera les robots plus fortement, particulièrement chez les enfants.

Au Chapitre 2, nous avons vu que les mécanismes tels que l'anthropomorphisme ou l'adoption de la *Posture Intentionnelle* guident nos prédictions et interprétations face aux robots sociaux. Cependant, que se passe-t-il ou comment réagirions-nous si le robot parle de lui, s'il affirme être en colère ou rap-

porte qu'il a déjà fait telle ou telle chose lui étant impossible et inaccessible ? Est-ce que nous traiterions ce discours comme nous le ferions chez un humain ? Serions-nous surpris par l'incongruence entre ce qui est dit et ce que l'on perçoit de ses capacités ? Dirions-nous qu'un robot ne peut pas être en colère ou bien qu'il se pourrait qu'il peut être programmé pour l'être ? Autrement dit, comment l'humain traite-t-il un discours qui, bien que grammaticalement correct et formulé d'une façon compréhensible, semble incompatible avec l'agent qui le prononce ?

4.3.2 Marqueurs neurocognitifs

Pour investiguer la façon dont les humains réagissent au niveau cérébral à de tels énoncés, les neurosciences cognitives offrent des outils comme les potentiels évoqués (ERP). Les ERP sont des potentiels électriques du cerveau liés à des événements spécifiques, extraits de l'électroencéphalographie (EEG) par moyennage des segments alignés sur ces événements (Luck, 2005). En d'autres termes, en mesurant l'activité électrique du cerveau à l'aide de l'EEG, il est possible de détecter des composants spécifiques associées à différents processus cognitifs.

L'un des marqueurs les plus connus en psycholinguistique est la composante N400, découverte par Kutas et Hillyard (1980). Ceux-ci décrivaient pour la première fois la N400 en montrant que les phrases contenant un mot cible incongruent avec le contexte général de la phrase provoquent une amplitude négative plus marquée que les phrases qui se terminent par un mot cible congruent. Par exemple, dans la phrase « *Il a tartiné le pain chaud avec des chaussettes* » (« *He spread the warm bread with socks* ») l'emploi du mot « chaussettes » (« *socks* »), qui induit une incongruence sémantique, est susceptible de déclencher une N400 plus grande par rapport à un élément attendu contextuellement (Kutas & Hillyard, 1980) tel que le mot « beurre » (« *butter* »).

Pour résumer, la N400 est une composante des ERP qui apparaît généralement environ 400 ms après la présentation d'un mot ou d'un stimulus qui est sémantiquement incongruent avec son contexte. Depuis sa découverte, la N400 a été largement étudiée (Luck, 2005) et elle ne se limite pas au langage écrit, elle est aussi sensible à la congruence de l'action (Reid & Striano, 2008 ; van Elk et al., 2008). Elle a également été trouvée dans un contexte où des images étaient sé-

mantiquement incongruentes par rapport à nom d'objet présenté auparavant (Friedrich & Friederici, 2004 ; Hamm et al., 2002). L'amplitude de la N400 augmente généralement avec le degré d'incohérence entre le stimulus cible et le contexte (Kutas & Federmeier, 2011). De même, il a été démontré que le traitement sémantique des mots cibles ne se fait pas seulement au niveau local de la phrase dont ils font partie, mais également et simultanément, au niveau du contexte global dans lequel ils sont présentés (Hagoort & van Berkum, 2007 ; van Berkum et al., 1999, 2003). Ainsi, ce sera le cas d'un contexte où l'incongruence est créée par le locuteur : une phrase cohérente telle que « Chaque soir, je bois un peu de vin avant d'aller me coucher » (« *Every evening I drink some wine before I go to sleep* ») suscite une réponse N400 plus forte lorsqu'elle est prononcée par la voix d'un enfant que lorsqu'elle est prononcée par une personne adulte (van Berkum et al., 2008).

Dès lors, les humains présentent-ils une réponse N400 typique lorsqu'ils entendent un robot évoquer des expériences qui dépassent ses capacités, comme parler de ses émotions ? Par exemple, la phrase « *Hier, j'ai regardé un film dramatique et cela m'a rendu triste* » paraît parfaitement cohérente lorsqu'elle est prononcée par un humain, car ressentir de la tristesse devant un film dramatique est une expérience attendue. En revanche, dans le cas où la phase est prononcée par un robot, qui est censé ne pas pouvoir éprouver d'émotions (et qui plus est ne peut pas « voir » un film au sens humain du terme) une telle affirmation paraît incongrue. L'inadéquation entre le locuteur et le contenu de son discours pourrait alors se traduire par une réponse N400 plus marquée en réaction à cette phrase que si celle-ci était prononcée par un humain.

Nos connaissances du monde nous indiquent qu'en général seuls les êtres vivants, notamment les animaux, possèdent de véritables états affectifs. Une réponse N400 pourrait ainsi apparaître lorsqu'un robot évoque le sujet des émotions, étant donné qu'il fait référence à une expérience qu'il ne peut réellement connaître, avec une amplitude qui serait encore plus marquée lorsqu'il parlerait de ses propres émotions. Toutefois, il est également possible que l'humain, surtout lorsqu'il est fortement engagé dans l'interaction, en vienne à accepter partiellement l'idée que le robot possède une vie mentale, ce qui pourrait atténuer l'amplitude de la N400. Enfin, si aucune différence n'était observée entre la condition « robot dit X » et « humain dit X », cela suggérerait que, pour l'audi-

teur, le robot est intégré à son système de connaissances presque au même titre qu'un humain. Si au contraire la différence est marquée, c'est le signe qu'une barrière conceptuelle subsiste. Ainsi, jusqu'où le cerveau humain va-t-il pour accorder aux entités artificielles le bénéfice d'être des « êtres sociaux » comme les autres, et où s'arrêtera-t-il en opposant un « veto » ?

En synthèse de cette Partie I, celle-ci a établi le cadre conceptuel pour aborder comment l'humain perçoit l'être social dans les agents artificiels. Elle a permis de poser quelques fondements nécessaires à la compréhension des mécanismes par lesquels les humains perçoivent et interagissent avec les ASA. L'examen des définitions et taxonomies (Chapitre 1) a établi la diversité des ASA et l'importance des variations culturelles dans leur perception, remettant en question l'idée d'une perception ou de réactions universelles et uniformes. Ensuite, il a été abordé la forte prédisposition de l'humain à projeter de l'humain autour de lui ou à s'engager socialement avec ce qui semble interactif notamment via des mécanismes tels que l'anthropomorphisme ou encore via l'adoption de la *Posture Intentionnelle* (Chapitre 2), tout en soulignant que les mécanismes cognitifs humains s'activent partiellement face aux robots, que des modulations spécifiques et des frontières persistantes semblent distinguer nos réactions face aux robots de celles face aux humains. Dans un troisième temps (Chapitre 3), le sentiment d'être avec un autre et la cognition sociale ont été précisés à travers plusieurs perspectives. Deux conceptions principales de la cognition sociale ont été distinguées : la conception traditionnelle, qui l'ancre dans des processus internes d'attribution (tels que la *ToM* ou la *Posture Intentionnelle*), et les approches interactionnistes et énactives, qui la considèrent comme émergeant des dynamiques d'interaction elles-mêmes, au sein du couplage entre agents. Le paradigme du *Perceptual Crossing* illustre cette idée en montrant que la reconnaissance d'autrui et la perception d'un autre sujet intentionnel ne reposent ni sur une forme particulière ni sur des trajectoires objectives, mais sur des propriétés intrinsèques à l'activité de perception conjointe elle-même.

Enfin, il a été souligné (Chapitre 4) que la temporalité et l'interaction verbale posent des défis au champ de recherche de l'interaction humain-robot. Les conversations humaines minimisent les silences, avec des transitions très courtes entre locuteurs se situant dans les 200 ms en général. Les systèmes actuels, qui incluent les robots, sont limités par des contraintes techniques, voire

peut-être par les attentes temporelles de l'humain : les préférences temporelles pour les robots semblent ne pas se calquer sur les échanges humain-humain. De même, il se pourrait que des frontières persistent dans ce que les humains acceptent qu'un robot dise. Ainsi, la perception de l'échange avec un robot pourrait être modulée par des facteurs à la fois temporels et sémantiques.

La partie suivante s'appuie sur ces bases théoriques pour examiner expérimentalement comment les propriétés des agents et les instructions données aux participants influencent la présence sociale ressentie et l'interaction.

Deuxième partie

Émergence de l’interaction et Présence sociale

Émergence de l'interaction et Présence sociale

Après avoir établi le cadre théorique et conceptuel nécessaire à la compréhension des mécanismes de perception sociale appliqués aux agents artificiels dans la première partie de cette thèse, cette seconde partie examine la possibilité qu'un sentiment de présence sociale émerge dans un contexte d'interaction minimaliste fondé sur une coordination sensorimotrice face à un agent artificiel. En particulier, nous cherchons à comprendre comment deux facteurs, (1) la présence ou l'absence d'une instruction sociale explicite en amont de l'interaction (c'est-à-dire, inviter le participant à rechercher activement un contact social) et (2) les propriétés dynamiques de l'agent rencontré, influencent à la fois le sentiment d'être avec un « autre » présent dans l'environnement et la manière dont les participants explorent et interagissent avec cet environnement.

Le Chapitre 5 présente en détail notre adaptation méthodologique de l'expérience d'Auvray et al. (2009) relative au paradigme du *Perceptual Crossing*. Ce chapitre inclut la conception de l'environnement virtuel, l'implémentation des différents agents artificiels, la procédure expérimentale de nos deux études ainsi que les méthodes d'analyse que nous utiliserons. Le Chapitre 6 rapporte les résultats des deux études et propose une discussion générale articulant les contributions de ces travaux à la compréhension de la présence sociale dans le cadre d'une interaction minimaliste entre humains et agents artificiels.

Chapitre 5

Adaptation du paradigme du perceptual crossing

5.1 Introduction

Comme exposé au Chapitre 3, le paradigme du *Perceptual Crossing* proposé par Auvray et al. (2009) constitue un cadre expérimental minimaliste particulièrement pertinent pour étudier l'émergence de l'interaction sociale en temps réel. Pour rappel, cette approche repose sur l'idée que la reconnaissance d'autrui et le sentiment d'être avec un autre peuvent émerger directement de la dynamique de couplage sensorimoteur entre deux agents, sans recours à des indices visuels, sémantiques ou communicationnels explicites.

Pour rappel également, dans l'expérience d'Auvray et al. (2009), deux participants interagissaient dans un environnement virtuel minimaliste, constitué d'un espace unidimensionnel circulaire où ils se déplaçaient à l'aide d'un avatar. Chaque contact avec un objet déclenchait un retour tactile binaire. Trois types d'objets pouvaient être rencontrés : l'avatar du partenaire, un leurre mobile reproduisant ses mouvements et un objet fixe. Les participants devaient indiquer, par un clic, lorsqu'ils pensaient être en contact avec l'autre personne. Les résultats de cette étude montraient l'émergence d'une dynamique d'exploration conjointe : les deux participants développent des schémas de mouvement mutuellement ajustés (par exemple, des oscillations en antiphase autour d'un point de contact) qui leur permettent de maintenir le contact avec l'avatar du partenaire significativement plus longtemps qu'avec le leurre ou l'objet fixe. Autrement dit, la contingence mutuelle (le fait pour chacun d'être simultanément affecté par et affecter le comportement de l'autre) permet l'émergence d'une coordination spécifique avec le partenaire humain, irréductible aux seuls mé-

canismes de détection individuelle. Ces travaux suggèrent que la contingence mutuelle pourrait être un paramètre important dans l'émergence de la reconnaissance de l'autre et du sentiment de présence sociale.

Toutefois et comme nous l'avons relevé au Chapitre 3, l'expérience originale d'Auvray et al. (2009) ne mesurait pas directement l'expérience subjective de présence sociale des participants mais portait exclusivement sur des indices comportementaux objectifs et sur la performance explicite de reconnaissance évaluée par les clics.

Nous avons alors présenté, toujours au Chapitre 3, les travaux de Bedia et al. (2014) qui ont montré que les dyades humaines présentant une coordination mutuelle manifestaient des corrélations à long terme et des motifs multifractals caractéristiques de dynamiques auto-organisées, révélant une structure temporelle fractale $1/f$ (avec une pente spectrale β proche de 1) absente lors d'interactions avec des agents non contingents. Ces signatures indiquent que la coordination humaine repose sur des couplages dynamiques multi-échelles. Cependant, Barone et al. (2020) n'ont pas répliqué ces résultats : leurs participants distinguaient mal les agents humains des agents artificiels, surtout lorsque la tâche imposait une classification binaire avec retour de performance. En revanche, lorsque le retour était supprimé et l'évaluation rendue graduelle, la reconnaissance correcte n'apparaissait que lorsque le feedback était audiovisuel, l'information auditive seule étant insuffisante.

Dans la continuité du paradigme de Auvray et al. (2009), Froese et al. (2014, 2020) ont étudié le lien entre la dynamique d'échange et l'expérience subjective de présence. En remplaçant l'unique essai continu par vingt courts essais de 60 s durant lesquels le participant devait cliquer une seule fois au moment où il estimait interagir avec l'autre puis évaluer après chaque essai la clarté de cette expérience, ils ont montré que la clarté de la perception sociale n'augmente pas après une détection correcte isolée, mais seulement lorsque les deux partenaires se reconnaissent mutuellement dans un même essai. L'expérience subjective de présence dépend de la réussite conjointe de l'interaction et non de la performance individuelle seule.

Ces travaux convergent vers l'idée que la coordination sensorimotrice module la reconnaissance d'un autre, mais suggèrent également que sa détection et son influence sur l'expérience subjective dépendent des modalités sen-

sorielles disponibles, du cadrage de la tâche expérimentale et des instructions données aux participants.

Cependant, plusieurs limitations méthodologiques et conceptuelles justifient une adaptation du paradigme du *Perceptual Crossing*.

Premièrement, aucun des travaux antérieurs n'a mesuré directement la *Présence Sociale* dans ce paradigme, telle que définie au Chapitre 3. Les travaux précédents se sont principalement focalisés sur des indices comportementaux (tels que la fréquence des contacts ou la distribution des clics) ou sur la clarté subjective de l'interaction. Ils n'ont pas évalué dans quelle mesure une telle configuration minimaliste est susceptible de générer un sentiment de présence sociale.

Notre premier objectif cherche à répondre à cette limitation en intégrant une mesure explicite du sentiment d'être avec un autre qu'est la *Présence Sociale*, et ce, à l'aide du questionnaire *Networked Minds Social Presence Inventory* (Harms & Biocca, 2004).

Nous nous concentrons spécifiquement sur deux dimensions : la coprésence et l'interdépendance comportementale perçue. Cela permet d'examiner dans quelle mesure la coordination sensorimotrice, dans un environnement minimaliste, peut générer une expérience subjective de présence sociale et comment ce sentiment varie selon les propriétés dynamiques de l'agent.

Deuxièmement, les travaux antérieurs fournissaient systématiquement une instruction sociale explicite aux participants, dès le début de l'expérience, instaurant d'emblée une attente de nature sociale. Or, cette instruction initiale peut orienter les stratégies d'exploration et influencer l'interprétation des événements sensoriels. Il reste donc à déterminer si les participants manifestent une propension spontanée à rechercher « l'autre » en l'absence de toute instruction sociale, ou si cette orientation nécessite une incitation explicite.

Dans ce sens, notre second objectif est d'isoler l'effet de l'instruction sociale sur l'exploration et la présence sociale perçue. Notre protocole manipule ce facteur : dans l'Étude 1, les participants reçoivent une instruction neutre et sans incitation sociale (« Explorez l'environnement »). Celle-ci ne mentionne pas la présence éventuelle d'un autre et permet d'observer si une propension à rechercher des contacts ou générer une interaction peut émerger. Dans l'Étude 2, l'instruction est socialement orientée (« Explorez l'environnement et essayez d'interagir avec l'autre »), instaurant une attente explicite d'interaction avec un

« autre », sans toutefois préciser la nature de celui-ci ni fournir de stratégie particulière. Ces deux études permettent de comparer l'impact de l'instruction sur les stratégies d'exploration, la *Présence Sociale* perçue, et les signatures des dynamiques temporelles de l'interaction. En d'autres termes, d'évaluer si l'incitation sociale est nécessaire pour orienter l'exploration vers une dynamique de recherche d'interaction.

Troisièmement, bien que certains travaux aient comparé des interactions humain-humain à des interactions avec des agents non-contingents, il se trouve que peu d'études ont manipulé la contingence du comportement de l'agent aux actions du participant dans le but d'évaluer leur influence respective sur la *Présence Sociale* perçue et les dynamiques. Pour répondre à cette limite, nous investiguons l'impact du comportement de l'agent et de sa contingence au participant en implémentant quatre types d'agent : deux agents non-contingents au comportement oscillatoire (l'un parfaitement régulier et périodique, l'autre avec une amplitude variable d'oscillation), un agent contingent explorant tout l'environnement qui répond aux collisions avec le participant, et un agent au comportement exploratoire similaire à l'agent contingent mais entièrement indépendant du comportement du participant. Cette manipulation permet d'évaluer dans quelle mesure la contingence comportementale de l'agent peut moduler l'émergence de la présence sociale.

Enfin, nous cherchons à caractériser les signatures temporelles multi-échelles de l'interaction. En reprenant certaines approches d'analyse utilisées par Bedia et al. (2014), nous appliquons l'analyse des fluctuations redressées et l'analyse multifractale des fluctuations redressées aux séries temporelles de vélocité individuelle (participant, agent) et collective (vélocité relative). Ces analyses permettent de quantifier la présence de corrélations à long terme (exposant β) et l'étendue de l'hétérogénéité temporelle multi-échelles (largeur du spectre multifractal Δh). Nous cherchons également à relier les mesures subjectives (questionnaire), comportementales (nombre de croisements) et dynamiques (analyses fractales) pour examiner les relations entre différentes facettes de l'interaction.

5.2 Méthode

5.2.1 Conception

Nos deux études partagent une structure expérimentale commune (même environnement, mêmes agents artificiels, mêmes mesures) mais diffèrent par leur plan expérimental et l'instruction fournie aux participants. L'Étude 1 adopte un plan inter-sujets avec quatre conditions expérimentales correspondant aux quatre types d'agents. Chaque participant est aléatoirement assigné à un type d'agent et réalise un bloc de deux essais de 60 s chacun face au même agent. Aucune mention d'un partenaire n'est faite dans les instructions : la consigne est uniquement « *Explorez l'environnement* ». À l'issue de l'exploration, les participants complètent un questionnaire. Cette conception permet notamment d'évaluer si un sentiment de présence sociale peut émerger de la coordination sensorimotrice en l'absence de potentielles attentes sociales préalables.

Pour l'Étude 2, un plan intra-sujets dans lequel chaque participant rencontre un seul type d'agent différent à chacun des 4 blocs a été utilisé (l'ordre des 4 blocs est aléatoirement déterminé à l'aide d'un algorithme Fisher-Yates en début d'expérience). Pour chacun des 4 blocs (soit à chaque agent), le participant réalise deux essais de 60 s puis complète le questionnaire de présence sociale. L'instruction fournie est explicitement orientée vers la recherche sociale : « *Explorez l'environnement et essayez d'interagir avec l'autre* ». Les participants savent dès le départ qu'ils doivent rechercher et tenter d'interagir avec un « autre », bien qu'ils ignorent la nature (humaine ou artificielle) de cet autre ou l'existence de différents types d'agents et leurs caractéristiques comportementales. Cette conception permet d'évaluer l'effet de l'incitation sociale explicite sur les stratégies d'exploration et la présence sociale perçue.

5.2.2 Participants

Étude 1

Deux cents participants ($N = 200$, soit cinquante participants par condition d'agent) ont été recrutés via la plateforme *Prolific* selon les critères suivants : un âge entre 18 et 65 ans, l'anglais comme première langue et langue maternelle,

l'absence de déficiences auditives, l'accès à un ordinateur (pas de tablette ou smartphone), être en plein écran sur son navigateur durant l'expérience, avoir un clavier, porter un casque filaire, utiliser de préférence une souris filaire et bénéficier d'une connexion internet stable.

L'étude durait environ 10 min et les participants étaient rémunérés 8 £ par heure (soit ≈1.35 £). Après exclusion des participants ayant rencontré des problèmes techniques ou fourni des données incomplètes, l'échantillon final ($n = 190$) était composé de 109 hommes (57.4%), 80 femmes (42.1%) et de 1 personne (0.5%) qui a préféré ne pas le préciser. L'âge des participants variait de 19 à 65 ans (Mdn $= 39.0$, $M = 40.1$, SD $= 11.2$). La majorité des participants résidaient au Royaume-Uni (56.8%), suivis des États-Unis (33.7%), du Canada (5.8%), de l'Irlande (2.6%), puis de l'Australie (0.5%), et de la Nouvelle-Zélande (0.5%).

Étude 2

Un total de quatre-vingt-douze participants a été recruté via la plateforme *Prolific* selon les mêmes critères que l'Étude 1. L'étude durait environ ≈20 min et les participants étaient rémunérés 8 £ par heure (soit ≈2.00 £). Concernant le genre, l'échantillon final après exclusion ($n = 89$), était composé de 46 hommes (51.7%), 42 femmes (47.2%) et de 1 personne (1.1%) non-binaire. L'âge des participants variait de 20 à 62 ans (Mdn $= 39.0$, $M = 40.2$, SD $= 11.8$). La majorité des participants résidaient au Royaume-Uni (55.1%), suivis des États-Unis (33.7%), du Canada (5.6%), de l'Irlande (4.5%) et de la Nouvelle-Zélande (1.1%).

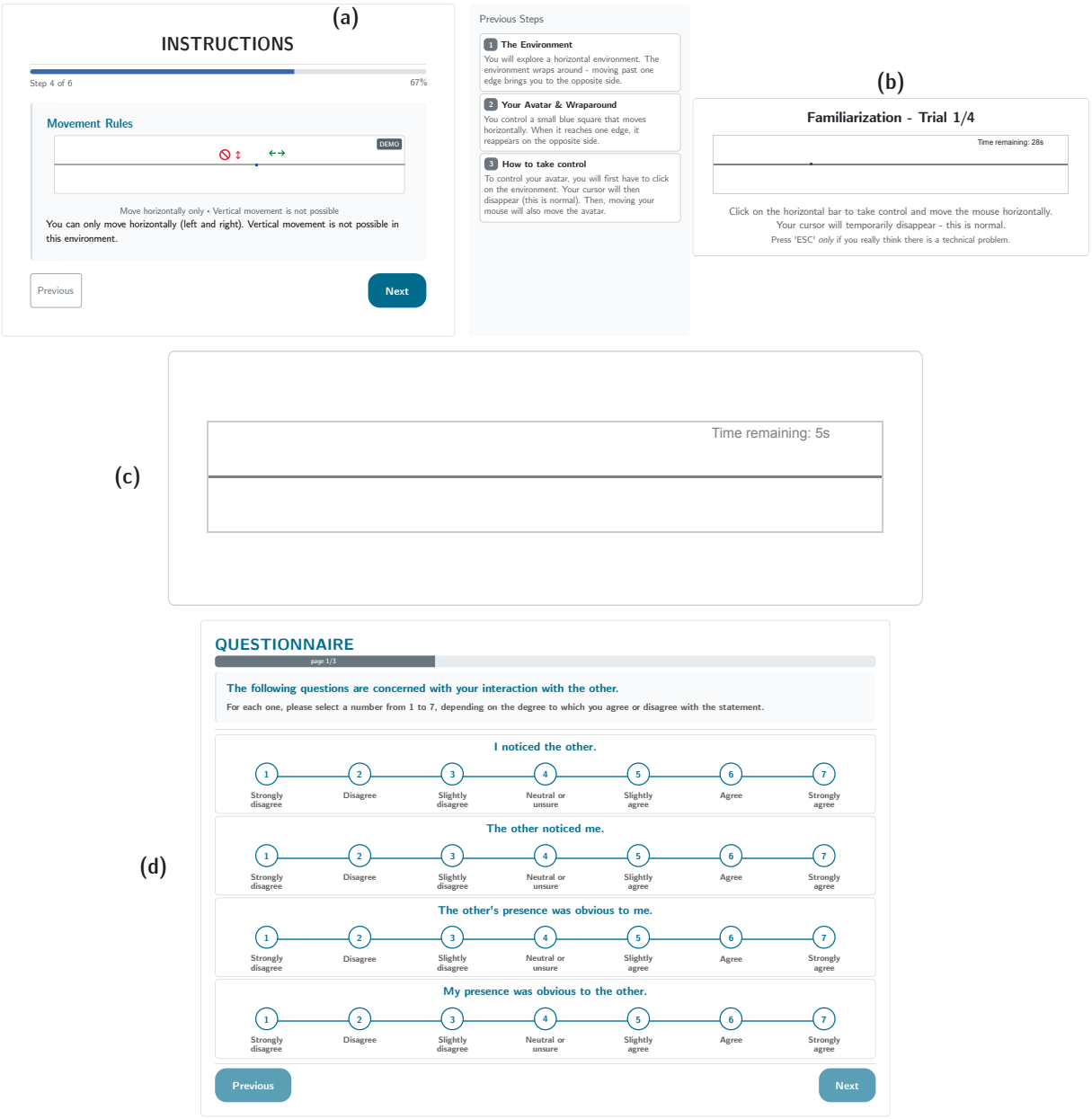
5.2.3 Matériel et Apparatus

Plateforme expérimentale

L'expérience et son interface étaient accessibles aux participants via le navigateur web de leur ordinateur (voir la Figure 5.1 pour un aperçu de l'interface). L'implémentation technique repose sur une application web développée en *JavaScript* à l'aide de la bibliothèque *React* (avec un environnement de développement optimisé par l'outil *Vite*).

Figure 5.1

Aperçu de l'interface pour différentes phases



Note. Capture d'écran : **(a)** d'une des étapes de la phase d'introduction où les participants recevaient un tutoriel visuel animé et structuré présentant l'environnement et les principes pour interagir avec. **(b)** de l'environnement lors de la phase de familiarisation durant laquelle les participants s'exerçaient et se familiarisaient avec l'environnement, notamment à contrôler leur avatar. **(c)** de l'environnement minimaliste de la phase principale. **(d)** de l'interface du questionnaire évaluant la Présence Sociale après avoir interagit dans l'environnement.

Le système intègre plusieurs mécanismes de contrôle. Il était conçu pour bloquer tout accès initial depuis un smartphone ou une tablette et affichait une fenêtre modale informant l'utilisateur qu'il devait utiliser un ordinateur. Le système était également conçu pour bloquer l'expérience et demander explicitement d'entrer en mode plein écran si le participant quittait le mode plein écran ou réduisait la fenêtre.

Pendant les essais, le verrouillage du pointeur de la souris s'active au clic du participant sur le canevas de l'environnement, faisant disparaître le curseur visuel et capturant les mouvements relatifs de la souris pour contrôler la position du curseur virtuel du participant dans l'espace toroïdal.

Le système d'enregistrement capturait les positions (du participant et de son partenaire) avec un intervalle de 16 ms et chaque collision était enregistrée avec son temps et ses positions exactes (en px). Ces données étaient transmises à la base de données à la fin de chaque essai. Les calculs physiques du mouvement de l'agent s'effectuent par pas de temps de 16 ms, indépendamment du taux de rafraîchissement du navigateur, garantissant une dynamique temporelle stable malgré les variations potentielles de performance selon les ordinateurs.

Stimuli

L'environnement d'interaction est un espace unidimensionnel circulaire de 600 px (en tore), c'est-à-dire une ligne dont les extrémités se rejoignent (lorsque l'agent ou le participant sort d'un côté, il réapparaît immédiatement de l'autre). Une ligne horizontale de 2 px de couleur grise était affichée par dessus cet environnement durant les essais : elle servait de référence visuelle et de support au feedback d'interaction (voir Figure 5.1c). Les avatars (participants et agents) sont des carrés de 4 px. Ils demeurent invisibles pendant toute la durée de la phase expérimentale, à l'exception de la phase de familiarisation, où l'avatar du participant était visible sous la forme d'un carré bleu durant deux des quatre essais, afin qu'il puisse se représenter la correspondance entre les mouvements de souris et le déplacement de l'avatar dans l'environnement.

Lors d'un chevauchement spatial ($\geq$1 px de superposition entre les deux avatars), c'est-à-dire un croisement ou une collision, l'événement est signalé selon deux modalités (audio et visuelle) en mode tout-ou-rien. En effet, tant qu'il y a un chevauchement, la ligne horizontale s'épaissit et est accompagnée d'un bip

sonore (note sinusoïdale à fréquence fondamentale de 300 Hz). Il est à noter qu'il a été fixé un temps minimum de 16 ms durant lequel l'audio est joué afin de lisser les artefacts audio (un phénomène de « claquement ») qui pourraient être provoqués par de très courts chevauchements.

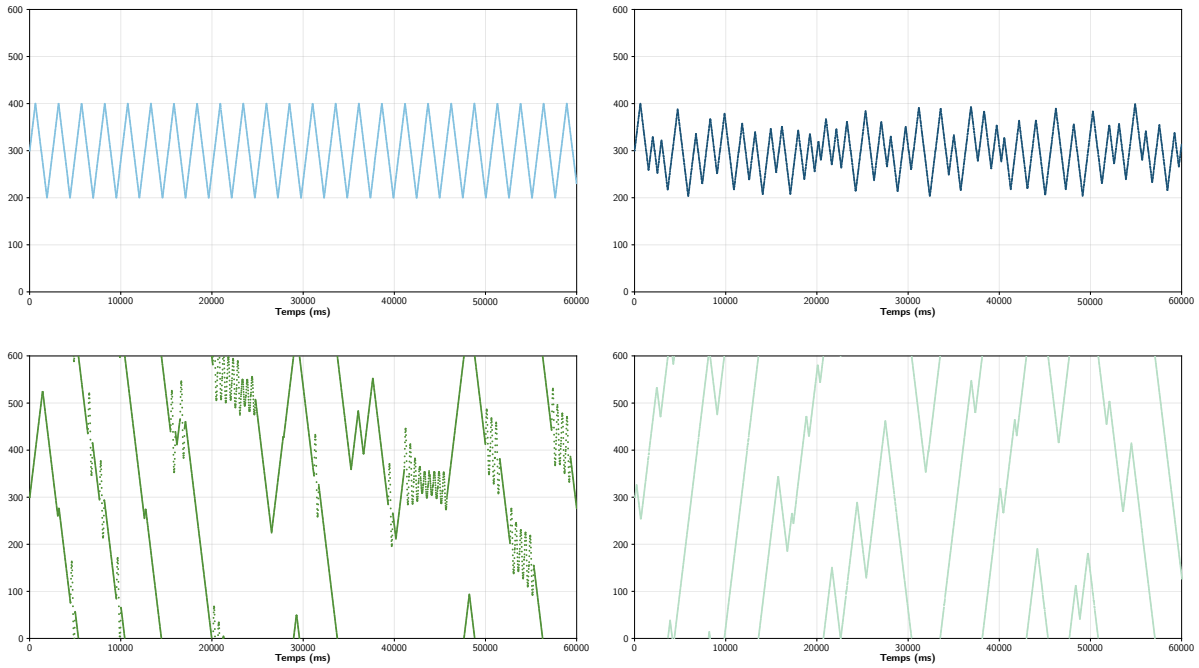
Notre approche contraste avec celles d'autres travaux comme Bedia et al. (2014) ou Barone et al. (2020) qui utilisaient un feedback avec un bip sonore d'une durée fixe minimale de 500 ms à chaque collision. Une telle durée aurait introduit plusieurs biais méthodologiques qui peuvent contaminer l'analyse de la dynamique sociale : elle masque les micro-dynamiques rapides en fusionnant, dans la perception, en un seul événement continu, différentes collisions espacées de moins de 500 ms. Si trois croisements surviennent en moins de 500 ms, le participant n'entendra qu'un seul bip continu dès le premier croisement. Cela peut imposer un rythme artificiel où le participant doit percevoir le monde par blocs de 500 ms, biaisant potentiellement l'analyse des schémas multi-échelles. C'est donc pour cela que nous avons adopté une approche où le retour sensoriel est délivré en tout-ou-rien et ne dure que tant qu'il y a chevauchement. Concernant les différents agents auxquels le participant peut faire face, les quatre types de comportements des agents implémentés diffèrent par leurs dynamiques de mouvement et leur réactivité aux interactions avec le participant. Dans toutes les conditions, la position est calculée modulo 600 px et chaque agent se déplace à $\approx$160 px/s. La description de chaque agent est fournie ci-dessous. La Figure 5.2 illustre les trajectoires pour chaque type d'agent.

L'agent dit « Périodique » a une trajectoire purement oscillatoire, avec une amplitude fixe de 100 px autour du centre [200 ; 400] px. Il effectue des va-et-vient réguliers à vitesse constante ($\approx$160 px/s) et alterne entre se déplacer vers une extrémité (droite ou gauche) et revenir au centre, sans jamais changer d'amplitude ou de vitesse. Il ne réagit pas au participant et ignore complètement ses actions ou sa position.

L'agent dit « Fluctuant » suit une trajectoire similaire à celle de l'agent Périodique (des oscillations d'aller-retours à vitesse constante $\approx$160 px/s) à la différence qu'il ajoute une part d'imprévisible : l'amplitude des oscillations autour du centre est renouvelée à chaque cycle complet via un tirage uniforme continu entre 20 et 100 px. Il ne réagit pas au participant et maintient ses oscillations indépendamment de la position ou des actions de ce dernier.

Figure 5.2

Exemples de trajectoires d'agents sur un essai



Note. Les tracés représentent la position (en pixels) en fonction du temps (ms) pour chacun des quatre types d'agents implémentés dans notre environnement de *Perceptual Crossing* dans un essai de 60 s. En haut à gauche, en bleu clair, l'agent Périodique. En haut à droite en bleu foncé, l'agent Fluctuant. En bas à gauche, en vert foncé, l'agent Réactif. En bas à droite, vert clair, l'agent indépendant. Ces exemples illustrent les caractéristiques dynamiques distinctives de chaque agent : des oscillations strictement régulières ou oscillations d'amplitude variable centrés sur le milieu de l'environnement sans réaction aux croisements avec le participant, une exploration de tout l'environnement et des réponses contingentes aux collisions avec le participant, et une exploration de l'environnement sans réagir au participant. Les trajectoires sont issues d'un participant de ces figures correspondent aux données données expérimentales d'un des participants.

Concernant l'agent dit « Réactif », celui-ci se déplace dans tout l'environnement à une vitesse de base (≈160 px/s) mais de façon relativement imprévisible car sa trajectoire est perturbée à chaque instant (à chaque 16 ms) par un bruit aléatoire corrélé dans le temps (bruit rose). Ainsi, plutôt que d'avancer en ligne parfaitement droite, l'agent avance (il zigzague) avec de petits écarts plus ou moins réguliers à cause de légères fluctuations dans la direction du mouvement. De plus, à chaque pas de temps de 16 ms, l'agent a une faible probabilité ($p = \frac{9}{1000} \approx 0.009 = 0.9\%$) de changer spontanément de direction.

La caractéristique distinctive de cet agent est sa contingence aux actions du participant. Lorsque le système lui indique qu'il y a collision, l'agent programme une oscillation qui dure 500 ms et qui est centrée sur sa position du moment : celle-ci est programmée avec une amplitude adaptée selon les événements passés (via des compteurs internes) tout en notant l'événement pour ajuster son comportement futur. C'est-à-dire qu'il effectue son va-et-vient d'amplitude initiale de 100 px mais adaptée (diminuée avec les interactions récentes, augmentée avec les oscillations sans nouvelle collision) et centrée sur la position de départ de cette oscillation spécifique. Si une nouvelle collision survient pendant qu'une oscillation est en cours, elle déclenche la programmation d'une nouvelle oscillation (avec un nouveau point central), mais l'oscillation actuelle continue jusqu'à sa fin. À chaque fin d'oscillation sans nouvelle collision, le compteur spécifique aux collisions manquées augmente et l'amplitude des futures oscillations augmente. Si aucune interaction n'est enregistrée pendant un certain laps de temps (fixé à 3 fois la durée d'une oscillation), l'agent change d'état en revenant automatiquement à son mode de déplacement initial (exploration). En somme, il simule de façon rudimentaire un agent qui agit comme s'il « cherchait » quelque chose sans savoir où. Lorsqu'il le « trouve », c'est comme s'il « voulait » maintenir le contact. Ce contact qu'il « veut » maintenir via des mouvements de balayage oscillatoires rappelle la dynamique d'antiphase décrite par Di Paolo et al. (2008), selon laquelle une forme stable de coordination entre agents peut résider dans une telle alternance rythmée autour d'un point de rencontre. De même, les oscillations d'une durée de 500 ms sont un parallèle à la concentration des croisements et des corrélations de trajectoires maximales autour de 500 ms dans des conditions de contingence mutuelle (Barone et al., 2020 ; Bedia et al., 2014) et au succès conjoint de se reconnaître l'un et l'autre

(Froese et al., 2014).

Pour l'agent dit « Indépendant », il possède le même comportement d'exploration dans l'environnement que l'agent Réactif mais ne réagit jamais au participant. Il simule un agent en quelque sorte libre et indifférent à la présence du participant.

5.2.4 Procédure

Les deux études suivent les mêmes phases : phase de vérifications techniques et de consentement, phase d'introduction, phase principale et phase de questionnaire.

Vérifications techniques et consentement

L'accès à l'expérience est soumis à des vérifications automatiques : le système bloque toute tentative d'accès depuis un autre appareil qu'un ordinateur et affiche une fenêtre modale invitant à utiliser un ordinateur. De même, il bloquait l'expérience et demandait explicitement d'entrer en mode plein écran si le participant quittait ce mode ou réduisait la fenêtre. La page d'accueil présente l'aperçu général de l'étude et sa durée estimée ($\approx$10 min pour l'Étude 1 et $\approx$20 min pour l'Étude 2), les prérequis techniques tels que la nécessité d'une connexion internet stable, l'utilisation d'un casque audio filaire et d'une souris (de préférence filaire), ainsi que la consigne de maintenir l'onglet du navigateur actif. Après lecture, la personne accède à la page de consentement.

Ensuite, la partie avec le formulaire de consentement décrit la nature des tâches à effectuer dans l'expérimentation en termes généraux (explorer un environnement minimaliste et remplir des questionnaires) sans révéler la nature des agents, ni les hypothèses de l'étude. Une fois le consentement obtenu, une phase de test audio permettait de vérifier le bon fonctionnement du système. Les participants devaient cliquer sur un bouton pour jouer un fichier audio contenant un mot (« *evidence* ») qu'ils devaient retranscrire. En cas de réponse incorrecte, un message d'erreur apparaît et l'interface propose de réécouter et de tenter à nouveau. Après cinq tentatives infructueuses, une option apparaissait et permettait d'indiquer l'impossibilité d'entendre, ce qui déclenchait une fenêtre modale de confirmation : si le participant confirmait ne pas pouvoir en-

tendre, il était redirigé vers *Prolific* (ce participant était noté comme n'ayant pas pu participer suite à un échec au test audio).

Phase d'introduction

Après le test audio, le participant passait par une phase de familiarisation. Cette phase se décompose en deux séquences : la première consiste en un tutoriel visuel animé et structuré en six étapes présentant successivement les éléments de l'environnement et les principes pour interagir avec ce dernier (voir l'exemple Figure 5.1a). Un système de pagination permettait de naviguer librement entre les étapes.

1. *L'environnement* : les participants apprennent qu'ils exploreront un espace horizontal qui est « circulaire » et que lorsqu'ils atteignent un bord, ils réapparaissent de l'autre côté.
2. *Leur avatar dans l'environnement* : il leur est montré un carré qui bouge sur une ligne et il leur est indiqué qu'ils contrôlent un petit carré qui ne se déplace que horizontalement. Il est également rappelé que traverser un bord fera réapparaître leur avatar de l'autre côté.
3. *Prise de contrôle* : une animation montre comment contrôler l'avatar en indiquant qu'il faut tout d'abord cliquer sur l'environnement pour commencer, que le curseur de la souris disparaîtra alors, et que cela est tout à fait normal. Ensuite, il leur est montré qu'une fois qu'ils déplacent leur souris, l'avatar lui aussi se déplacera.
4. *Règles de déplacement* : sur cet écran, il leur est indiqué qu'ils ne peuvent se déplacer qu'horizontalement (à gauche et à droite) et que les déplacements verticaux ne sont pas possibles dans cet environnement. Une animation illustre les mouvements possibles (un carré bougeant de gauche à droite horizontalement accompagné de flèches d'indications colorées en vert) ou impossibles (par une flèche rouge bidimensionnelle verticale, accompagnée d'un emoji *sens interdit* pour un mouvement vertical).
5. *Retour d'information (feedback)* : il est montré la ligne horizontale s'épaississant accompagnée d'un bip sonore. L'explication indiquait qu'ils entendront un son et verront la ligne horizontale s'épaissir chaque fois que leur position chevauchera quelque chose (mais il ne leur est pas précisé quoi).

Le texte de chaque étape déjà consultée était disponible à droite de l'écran à tout moment de cette phase d'explication. Une fois les explications terminées, le dernier écran présentait la structure générale de l'expérience : Familiarisation, Phase d'exploration (avec le nombre de blocs et d'essais selon leur étude), Questionnaire, puis Fin.

La seconde séquence de la phase de familiarisation comprenait quatre essais de 30 s chacun, durant lesquels les participants s'exerçaient à contrôler leur avatar dans l'environnement (un agent était présent mais invisible à l'écran). Avant chaque essai, un écran de préparation informait de la visibilité ou non de leur propre avatar et rappelait la durée de l'essai. Un compteur temporel décroissant apparaissait pendant l'essai dans le coin supérieur droit de l'environnement. En dessous de l'environnement un message rappelait les instructions de prise de contrôle et de déplacement. Une mention discrète indiquait qu'en cas de réel problème technique, il était possible de presser sur la touche *ESC* (échap) du clavier pour arrêter de contrôler son avatar et retrouver son curseur de souris.

Durant les essais 1 et 4, l'avatar du participant est visible (carré bleu) tandis qu'il restait invisible durant les essais 2 et 3. Durant tous les essais, un agent était présent, invisible à l'écran, sans que le participant le sache (essai 1 et 2 : Agent Périodique, essai 3 et 4 : Agent Indépendant). L'alternance entre voir ou ne pas voir son propre avatar permet d'habituer le participant à se représenter ce que ses actions de mouvement de souris produisent dans l'environnement et sur leur avatar. En somme, elle permet donc au participant de se familiariser avec les contrôles et les modalités sensorielles de l'environnement tout en le préparant à la phase principale où son avatar sera systématiquement invisible.

Phase principale

Un écran de transition marquait le passage à la phase principale et rappelait que l'avatar serait désormais invisible pour le reste de l'expérience. Cet écran présentait l'instruction centrale pour cette phase :

Pour l'Étude 1 : (traduction) « *Explorez l'environnement* » (c'est-à-dire « *Explore the environment* » dans la langue de l'expérience). Aucune mention de la présence d'un agent ou d'objectif d'interaction n'était formulée.

Pour l'Étude 2 : (traduction) « *Explorez l'environnement et essayez d'interagir avec l'autre* » (c'est-à-dire « *Explore the environment and try to interact with the*

other » dans la langue de l'expérience). Ici, la présence d'un autre est explicitement mentionnée.

Dans l'Étude 1, la phase principale comprend un seul bloc de deux essais consécutifs de 60 s, réalisée avec un seul type d'agent, assigné au début de l'expérience. Avant le premier essai, un écran de préparation indiquait qu'il s'agissait de l'essai 1 sur 2. Un bouton permet de déclencher le lancement de l'essai, initialisant le chronomètre, la position initiale du participant (à 150 px) et celle de l'agent (à 300 px). Contrairement à la phase de familiarisation où du texte était présent sous l'environnement, l'affichage se composait de l'environnement avec la ligne médiane grise et du chronomètre. Pour rappel, le feedback audiovisuel s'active automatiquement lors d'un chevauchement d'au moins $\geq$1 pixel, provoquant un épaississement de la ligne et l'émission d'un bip sonore, qui persistent pendant toute la durée du chevauchement (en tout-ou-rien).

Dans l'Étude 2, les participants rencontrent successivement les quatre types d'agents, avec un agent par bloc. L'ordre des blocs était aléatoirement déterminé pour chaque participant en début d'expérience (mélange de Fisher-Yates). Comme dans l'Étude 1, chaque bloc comprenait deux essais de 60 s.

À l'issue des 60 s, un écran intermédiaire confirmait la complétion de l'essai et invitait le participant à se préparer pour le second essai. À l'issue du second essai, un message l'informait que la phase d'exploration était terminée puis il passait à la phase de questionnaire.

Phase de questionnaire

Durant cette phase, les participants complétaient un questionnaire évaluant leur ressenti de présence sociale. Dans l'Étude 1, c'est seulement à ce moment qu'ils découvrent la nature sociale de l'expérience. Le *Networked Minds Social Presence Inventory* (Harms & Biocca, 2004) a été utilisé en se concentrant sur deux dimensions : la *coprésence* et l'*interdépendance comportementale perçue*.

— La *coprésence* correspond au degré auquel un individu sait qu'il n'est pas seul dans l'environnement, son niveau de conscience (focalisée ou périphérique) de l'autre et son impression que l'autre est réciproquement conscient (de manière périphérique ou focalisée) de l'individu (Harms & Biocca, 2004).

— L'*interdépendance comportementale perçue* renvoie, pour sa part, à la mesure dans laquelle l'individu estime que son propre comportement affecte

et est affecté par le comportement de l'autre avec qui il a interagit (Harms & Biocca, 2004).

Pour chacune de ces deux dimensions, le questionnaire comportait 6 items, évalués sur une échelle de Likert en 7 points allant de « Pas du tout d'accord » à « Tout à fait d'accord ». Plus le score est élevé, plus la présence sociale perçue est forte. Chaque item concernant le participant possède son équivalent miroir concernant le point de vue de l'autre agent.

Dans la dimension de *coprésence*, par exemple : « *J'ai remarqué l'autre* » avec « *L'autre m'a remarqué* » (« *I noticed the other* » et « *The other noticed me* ») ou encore « *Ma présence était évidente pour l'autre* » et « *La présence de l'autre était évidente pour moi*. » (« *My presence was obvious to the other* » et « *The other's presence was obvious to me* »).

Les douze items étaient répartis en trois pages de quatre items. Une barre de progression affichait l'avancement. Sur la dernière page, le bouton « Soumettre » vérifiait que toutes les réponses avaient été renseignées avant de permettre de poursuivre. Pour l'Étude 1, le questionnaire n'était rempli qu'une seule fois, à la fin de l'unique bloc de deux essais avec l'agent. En revanche, dans l'Étude 2, le questionnaire était complété 4 fois, après chaque bloc.

Pour indiquer la fin de l'expérience, un dernier écran présentait un message de remerciement accompagné d'un résumé concis du contexte scientifique. Les participants étaient alors explicitement informés de l'existence d'un autre partenaire (mais pas de sa nature exacte) et de la nature de la tâche. Ensuite, un compte à rebours de 15 s informait de la redirection automatique vers *Prolific*, tandis qu'un bouton permettait de la déclencher immédiatement.

5.2.5 Analyse des données

Prétraitement

Tout d'abord, les enregistrements de tous les participants ont été inspectés pour détecter d'éventuelles anomalies techniques (données vides, partiellement incomplètes ou aberrantes comme des valeurs de temps négatives pour les positions). Pour l'Étude 1, 10 participants (sur 200) ont été exclus, et 3 (sur 92) pour l'Étude 2 en raison d'enregistrements incomplets inexploitables. Pour les participants retenus, malgré un enregistrement des positions toutes les 16 ms,

un ré-échantillonnage avec une grille uniforme à 16 ms a été appliqué aux données temporelles (sans altérer les données de façon substantielle) en raison de variations de performances des ordinateurs des participants et pour faciliter l'analyse.

Analyse des séries temporelles

Pour chaque essai, les séries temporelles de positions ont été converties en séries de vélocité : vélocité du participant, de l'agent ainsi que vélocité relative (dérivée de la distance entre le participant et l'agent). Pour caractériser la structure des corrélations à travers les échelles temporelles, ces séries temporelles de vélocité ont été soumises à une analyse des fluctuations redressées (*DFA*). La *DFA* permet d'estimer l'exposant d'échelle α, indicateur de la structure des fluctuations et de la manière dont elles s'échelonnent avec la taille de l'échelle d'observation, reflétant le degré et la nature des corrélations à longue portée dans le temps. Par exemple, la *DFA* permet de déterminer si les variations de vitesses sont organisées dans le temps ou aléatoires. Ensuite, à partir de cet exposant α, la pente spectrale β est calculée afin d'inspecter les dynamiques dans le domaine des fréquences et de fournir un indicateur de la manière dont la variance de la série temporelle se répartit selon la fréquence. Une *MFDFA* a été conduite pour mesurer la largeur du spectre multifractal (Δh), indicatrice de la diversité des échelles temporelles impliquées dans la dynamique de la série de temps.

Plus précisément, l'analyse *DFA* appliquée aux séries temporelles suit la procédure décrite par Bedia et al. (2014) mais a été adaptée à notre résolution temporelle de 16 ms. Les séries sont d'abord centrées et intégrées (somme cumulée) pour obtenir le profil cumulatif. Ensuite, le profil cumulatif est segmenté en fenêtres de tailles n log-échelonnées comprises entre $n_{\min} = 4$ (soit 64 ms, correspondant à l'échelle minimale utilisée par Bedia et al. (2014) à 1 ms de résolution) jusqu'à $n_{\max} = N/4$, où N représente la longueur de la série. Le pas entre les tailles de fenêtres est défini par un ratio constant de $2^{0.01}$ (soit environ 0.7% d'augmentation entre fenêtres successives). Dans chaque taille de fenêtre n, la tendance linéaire locale est retirée (pour ne pas confondre tendance et réelle fluctuation) par une régression linéaire et l'erreur quadratique moyenne des résidus est calculée. Les fluctuations moyennes quadratiques $F(n)$ sont calculées

sur les erreurs quadratiques moyennes des résidus des segments.

Les paires (n, $F(n)$) sont ensuite transformées en coordonnées log-log ($\log_{10} n$, $\log_{10} F(n)$), et l'exposant d'échelle α est ensuite estimé comme la pente de la régression linéaire sur un sous-ensemble de ces points log-log. Conformément à Bedia et al. (2014), cette régression est restreinte, lorsque cela est possible, à la décade située immédiatement sous un point de coupure (*cutoff*) détectable dans la courbure de cette relation d'échelle. A défaut, une décade centrée autour de la plus petite échelle disponible est utilisée. Ensuite, à partir de cet exposant, la pente spectrale β a été obtenue (selon la relation $\beta = 2\alpha - 1$).

La *MFDFA* reprend les mêmes paramètres que la *DFA* et calcule les fluctuations $F_q(n)$ pour une gamme d'exposants $q \in [-3, 3]$ par pas de 0.25. Cette analyse permet d'estimer la fonction d'échelle pour chaque valeur de q. La fonction d'échelle $h(q)$ est estimée par régression linéaire entre $\log_{10} F_q(n)$ et $\log_{10} n$. La largeur du spectre multifractal est calculée comme $\Delta h = h_{\text{max}} - h_{\text{min}}$, où h_{max} et h_{min} correspondent aux valeurs extrêmes parmi les $h(q)$ valides. Une grande largeur de spectre indique une forte hétérogénéité des dynamiques temporelles et la présence d'intermittences multi-échelles (rafales brèves et intenses séparées par des périodes calmes).

Modélisation

Les exposants α, β et Δh ont été modélisés au moyen de modèles linéaires mixtes (LMM) pour estimer les effets principaux du type d'agent et de l'étude, ainsi que leur interaction, en prenant en compte les intercepts aléatoires par participant. Les effets fixes et leur interaction ont été testés par une ANOVA de type III avec approximation de Satterthwaite pour les degrés de liberté. La variance interindividuelle a été quantifiée par le coefficient de corrélation intraclasse (ICC ajusté). En cas d'interaction significative, des comparaisons post-hoc ont été conduites sur les moyennes marginales estimées (EMM) avec correction des comparaisons multiples selon une procédure de Holm : d'une part, les types d'agents ont été comparés au sein de chaque étude, et d'autre part, les études ont été comparées au sein de chaque type d'agent. La catégorie de référence était l'agent Périodique.

Les nombres de croisements par essai ont été modélisés au moyen de *Generalized Linear Mixed Model* (*GLMM*) à loi binomiale négative avec lien logarith-

mique, incluant comme effets fixes le type d'agent, l'étude, ainsi que leur interaction et des intercepts aléatoires par participant. Les effets fixes ont été évalués par une ANOVA de type III (tests de Wald) sous codage à sommes contraintes. Les moyennes marginales estimées (EMM, nombres attendus) et leurs intervalles de confiance à 95% correspondants ont été obtenus sur l'échelle de réponse. En cas d'interaction significative ($p < .05$), les effets simples ont été examinés en comparaison par paires d'agents au sein de chaque étude et comparaison des paires d'études selon l'agent, avec correction séquentielle de Holm pour les comparaisons multiples. Les résultats sont reportés en rapports de taux (RR) avec intervalles de confiance. En l'absence d'interaction significative, les comparaisons principales ont été menées sur le facteur concerné. L'agent Périodique a servi de référence pour le facteur type d'agent.

Analyse des scores de questionnaire

La consistance interne des deux dimensions étudiées de la présence sociale a été évaluée avec le coefficient α de Cronbach. Ensuite, les analyses ont porté sur le score moyen des six items à chaque dimension. Pour la comparaison entre l'Étude 1 et l'Étude 2, ces scores ont été modélisés au moyen de LMM avec effets fixes pour le type d'agent, l'étude ainsi que leur interaction. De même, il a été inclus les intercepts aléatoires des participants. Les effets fixes ont été évalués par une ANOVA de type III avec approximation des degrés de liberté selon Satterthwaite. En cas d'interaction significative ($p < .05$), des comparaisons post-hoc (avec correction de Holm) sur les EMM ont été réalisées dans les deux directions : (1) entre types d'agents au sein de chaque étude et (2) entre études au sein de chaque type d'agent. En l'absence d'interaction, les comparaisons (avec correction de Holm) ont porté sur l'effet principal pertinent.

Pour la comparaison entre l'Étude 1 et le bloc 1 de l'Étude 2 (un seul agent par participant et aucune mesure répétée), les scores ont été modélisés au moyen d'un modèle linéaire (LM) incluant les effets fixes du type d'agent, de l'étude et de leur interaction. L'inférence reposait sur une ANOVA de type III avec erreurs-types robustes HC3. En cas d'interaction significative, des effets simples étaient testés dans les deux directions : comparaisons entre types d'agents au sein de chaque étude, et comparaisons entre études au sein de chaque type d'agent. À défaut, les comparaisons ont porté sur l'effet principal pertinent. Les analyses

post-hoc auraient été conduites sur les EMM (avec correction séquentielle de Holm). L'agent Périodique servait également de référence.

Chapitre 6

Impact de l'instruction

Après avoir présenté l'environnement expérimental et les méthodes utilisées (Chapitre 5), ce chapitre examine les résultats de l'effet conjoint des propriétés des agents et du type d'instruction donné aux participants (sociale ou neutre) sur la *Présence Sociale* perçue et la dynamique temporelle de l'exploration. Pour rappel, l'Étude 1 est une condition de référence dans laquelle aucune incitation sociale n'est fournie dans l'instruction : les participants reçoivent l'instruction neutre « Explorez l'environnement » et ne sont jamais informés qu'ils interagissent potentiellement avec un « autre » pendant l'exploration. Ce n'est qu'au moment du questionnaire final que la présence d'un « autre » est évoquée, sans préciser sa nature (humaine ou artificielle). L'Étude 2 introduit une incitation explicite avec « Explorez l'environnement et essayez d'interagir avec l'autre ». Cette manipulation permet d'isoler l'effet de l'instruction sur les stratégies d'exploration, la *Présence Sociale* perçue et les signatures temporelles des dynamiques d'interaction. Ce chapitre commence par la présentation des hypothèses, puis détaille successivement les résultats de nos deux études : (1) au questionnaire de *Présence Sociale*, (2) au nombre de croisements observés, et (3) à l'analyse des dynamiques multi-échelles des séries de vélocité (relatives, de celle du participant et de celle de l'agent).

6.1 Hypothèses

Plusieurs hypothèses ont été formulées. Tout d'abord, au sujet de l'instruction sociale, nous nous attendons à des scores de *Présence Sociale* indiquant l'émergence de ce sentiment, même en l'absence d'instruction sociale. Tandis que l'instruction explicite de tenter d'interagir avec l'autre augmentera globalement les scores de *Présence Sociale* par rapport à l'instruction neutre, particu-

lièrement sur la dimension d'Interdépendance comportementale. L'instruction sociale explicite aura également un effet sur les comportements en augmentant le nombre de croisements.

Nous nous attendons à un effet du type d'agent. L'agent Réactif (contingent) suscitera des scores de coprésence et d'interdépendance comportementale plus élevés que les agents non-contingents. Parmi eux, l'agent Indépendant, avec son comportement exploratoire plus riche, induira des scores légèrement supérieurs aux agents oscillatoires. L'agent Réactif induira un nombre significativement plus élevé de croisements que les autres agents tandis que les agents oscillatoires présenteront des niveaux intermédiaires et l'agent Indépendant, étant le plus difficile à trouver, présentera le nombre le plus faible.

Selon l'instruction (consigne neutre ou sociale), le type d'agent ou selon l'interaction entre l'instruction et le type d'agent, nous faisons l'hypothèse que la structure des corrélations des séries temporelles de vélocité du participant ne diffère pas mais que les signatures temporelles issues de la vélocité relative (entre le participant et l'agent) refléteront des dynamiques distinctes selon le type d'agent. En particulier, l'interaction avec l'agent Réactif entraînera des signatures distinctes (α et β) dans la vélocité relative de celles avec des agents non-contingents, et la largeur du spectre multifractal (Δh) sera la plus grande pour les agents n'étant pas cloisonnés à osciller autour du centre.

6.2 Résultats

Les résultats présentés dans cette section proviennent des comparaisons entre l'Étude 1 (instruction neutre, plan inter-sujets) et l'Étude 2 (instruction sociale explicite, plan intra-sujets). Elles examinent l'effet de l'instruction sur le nombre de croisements, sur la *Présence Sociale* perçue et sur la dynamique temporelle de l'exploration. Les analyses incluent la comparaison globale entre l'ensemble de toutes les données (Étude 1 comparée à Étude 2) et la comparaison ciblée entre le bloc unique de l'Étude 1 et le premier bloc de l'Étude 2. Cette dernière comparaison offre une base de comparaison plus propre, car les deux conditions sont structurellement équivalentes, et permet de contrôler les effets potentiels d'apprentissage ou de fatigue liés à l'exposition répétée aux quatre agents dans l'Étude 2.

6.2.1 Présence sociale

La Figure 6.1 présente les diagrammes en boîte comparant les perceptions sur les deux dimensions de coprésence et d'interdépendance comportementale du questionnaire de Harms et Biocca (2004). La consistance interne de chaque dimension était élevée (pour toutes, α de Cronbach > 0.83). La suite de la section précise les résultats pour les deux dimensions évaluées de la *Présence Sociale* .

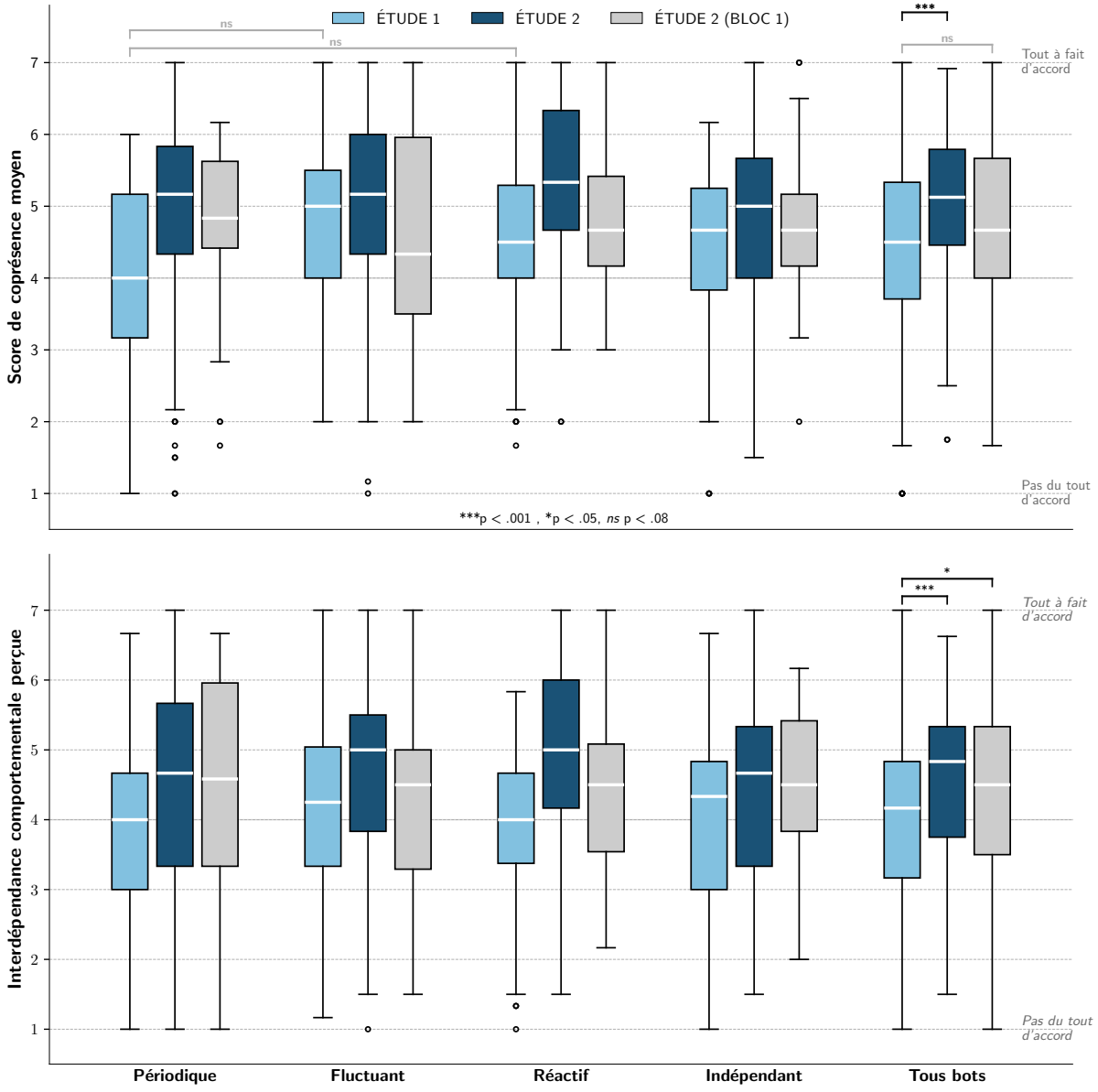
Coprésence

Le coefficient de corrélation intraclasse ajusté était de ICC $= 0.51$, indiquant une variance interindividuelle importante. L'ANOVA de type III (méthode de Satterthwaite) révèle un effet principal significatif du type d'agent, $F(3, 502.92)= 3.39$, $p = .018$, ainsi que de l'étude, $F(1, 240.85)= 18.38$, $p < .001$, mais aucune interaction significative entre les deux facteurs ($p = .115$). La comparaison post-hoc (avec correction de Holm) entre les deux études montre que les participants de l'Étude 2 rapportaient une coprésence plus élevée ($M = 5.00$, $IC_{95\%} = [4.78, 5.21]$) que ceux de l'Étude 1 où l'instruction était neutre ($M = 4.38$, $IC_{95\%} = [4.19, 4.57]$), avec une différence de 0.62 ($t(235.56)= -4.25$, $p < .001$). Pour les comparaisons post-hoc pour l'effet du type d'agent, aucune différence par paire n'était significative ($p > .05$). Le Tableau 6.1 contient les valeurs des moyennes marginales estimées (EMM) de cette dimension selon l'agent.

Concernant l'analyse qui comparait le premier bloc de l'Étude 2 à l'Étude 1 (un seul agent par participant et sans mesures répétées), un LM avec erreurs-types robustes HC3 incluant les effets fixes du type d'agent, de l'étude et de leur interaction a été utilisé. L'ANOVA de type III révèle un effet marginal de l'étude ($F(1, 271)= 3.76$, $p = .053$), sans effet principal du type d'agent, ni interaction significative entre les deux facteurs.

Figure 6.1

Scores aux deux dimensions de présence sociale investiguées



Note. **(en haut)** Coprésence, **(en bas)** Interdépendance comportementale perçue

Table 6.1

Moyennes marginales estimées de la coprésence selon l'agent

Agent	EMM	SE	$IC_{95\%}$
Périodique	4.47	0.12	[4.25, 4.70]
Indépendant	4.58	0.12	[4.35, 4.81]
Réactif	4.84	0.12	[4.61, 5.07]
Fluctuant	4.86	0.12	[4.63, 5.09]

Note. Les EMM sont présentées pour l'effet principal du type d'agent, $F(3, 502.92) = 3.39$, $p = .018$, agrégées sur les deux études (Étude 1 et Étude 2), l'interaction *agent* × *étude* n'étant pas significative, $F(3, 502.92) = 1.99$, $p = .115$.

Interdépendance comportementale perçue

Pour la dimension d'interdépendance comportementale perçue, un LMM a été ajusté suivant la même spécification que pour la coprésence. Le coefficient de corrélation intraclasse ajusté était de ICC $= 0.529$ indiquant également une assez grande variance entre les participants. L'ANOVA de type III n'a révélé aucun effet significatif du type d'agent ($F(3, 498.98) = 1.45$, $p = .227$). En revanche, un effet principal significatif de l'étude a été observé ($F(1, 245.75) = 16.11$, $p < .001$), mais l'interaction entre l'étude et le type d'agent n'était pas significative ($F(3, 498.98) = 1.68$, $p = .171$). Le Tableau 6.2 présente les EMM de cette dimension selon l'étude.

Table 6.2

Moyennes marginales estimées de l'interdépendance comportementale perçue selon l'étude

Étude	EMM	SE	$IC_{95\%}$
Étude 1	3.97	0.10	[3.77, 4.16]
Étude 2	4.58	0.12	[4.35, 4.81]

Note. Effet principal significatif de l'étude, $F(1, 245.75) = 16.11$, $p < .001$. Les participants ont rapporté une plus grande *interdépendance comportementale perçue* dans l'Étude 2 que dans l'Étude 1, avec une différence de 0.61, ($t(235.87) = -3.98$, $p < .001$.). L'étude correspond à l'instruction donnée (une incitation sociale dans l'Étude 2 et une consigne neutre dans l'Étude 1).

La comparaison post-hoc de l'effet de l'étude (ajustement de Holm) a montré que les participants de l'Étude 2 rapportaient une interdépendance significativement plus élevée ($M = 4.58$, $IC_{95\%} = [4.35, 4.81]$) en comparaison à ceux de l'Étude 1 ($M = 3.97$, $IC_{95\%} = [3.77, 4.16]$), avec une différence de 0.61 ($t(235.87) = -3.98$, $p < .001$).

Pour l'analyse entre le bloc 1 de l'Étude 2 et l'Étude 1, le LM avec une ANOVA de type III (Wald avec SE HC3) révèle également un effet significatif de l'étude, $F(1, 271) = 4.69$, $p = .031$, sans effet principal du type d'agent, ni interaction entre les 2 facteurs. La comparaison post-hoc (ajustement de Holm avec erreurs-types robustes) a révélé que les participants du bloc 1 de l'Étude 2 rapportaient une interdépendance comportementale significativement plus élevée ($M = 4.36$, $IC_{95\%} = [4.06, 4.66]$) que ceux de l'Étude 1 ($M = 3.97$, $IC_{95\%} = [3.78, 4.16]$), avec une différence de 0.39 ($t(271) = -2.17$, $p = .031$).

6.2.2 Nombre de croisements

La Figure 6.2 présente les diagrammes en boîte comparant le nombre de croisements selon l'agent et l'étude (l'instruction).

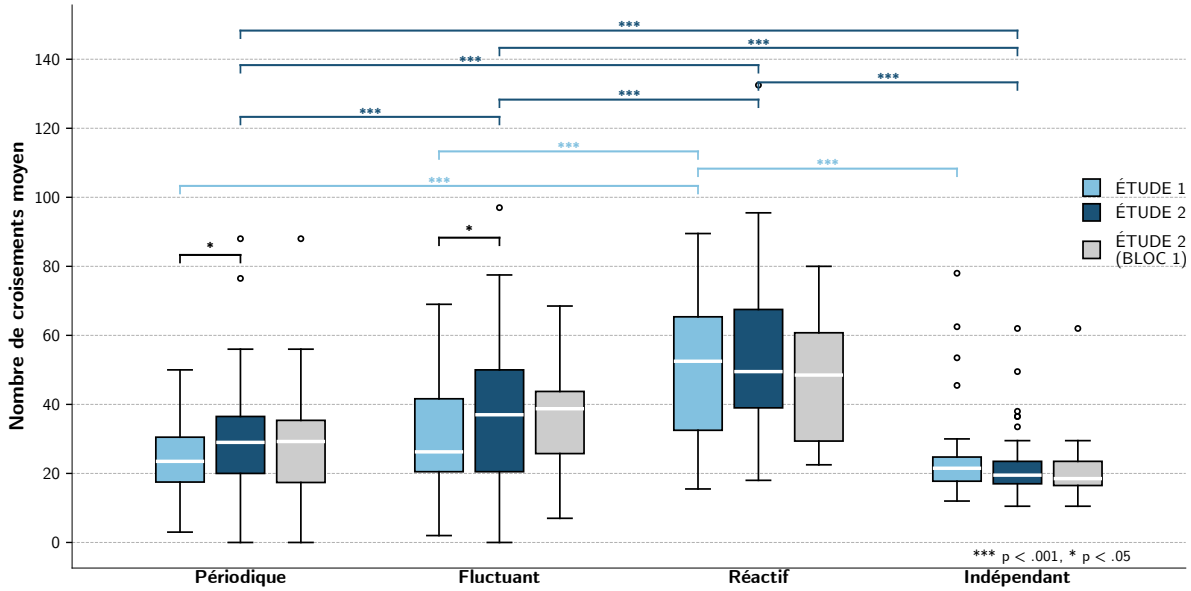
Le nombre de croisements (chevauchements) a été modélisé au moyen de *GLMM* à loi binomiale négative avec lien logarithmique, incluant comme effets fixes le type d'agent, l'étude (instruction neutre ou instruction sociale), ainsi que leur interaction et des intercepts aléatoires par participant.

L'ANOVA de type III (tests de Wald sous codage à sommes contraintes) révèle un effet principal du type d'agent hautement significatif, $\chi^2(3) = 293.60$, $p < .001$, un effet marginal de l'étude, $\chi^2(1) = 3.20$, $p = .074$, mais une interaction significative entre ces deux facteurs, $\chi^2(3) = 9.58$, $p = .022$.

Les effets simples ont été examinés par comparaison des paires d'agents au sein de chaque étude et par comparaison des paires d'études selon le type d'agent, avec correction séquentielle de Holm pour les comparaisons multiples. Les EMM selon l'étude et l'agent (nombre attendu de croisements par essai de 60 s) ainsi que leurs intervalles de confiance à 95% correspondants sont présentés dans le Tableau 6.3.

Figure 6.2

Nombre de croisements selon l'agent et l'Étude



Note. La figure montre le nombre de croisements observés pour chaque type d'agent dans l'Étude 1 (instruction neutre) et l'Étude 2 (instruction sociale) et dans le premier bloc de l'Étude 2.

Table 6.3

Nombre moyen estimé de croisements par essai selon l'agent et l'étude

Étude	Agent	EMM	SE	$IC_{95\%}$
Étude 1	Périodique	22.47	1.47	[19.77, 25.54]
	Indépendant	22.87	1.52	[20.08, 26.04]
	Réactif	47.92	3.13	[42.16, 54.46]
	Fluctuant	27.57	1.81	[24.24, 31.35]
Étude 2	Périodique	27.14	1.30	[24.71, 29.80]
	Indépendant	20.82	1.01	[18.93, 22.89]
	Réactif	51.75	2.43	[47.20, 56.73]
	Fluctuant	33.20	1.58	[30.25, 36.44]

Note. L'interaction entre l'étude et le type d'agent était significative, $\chi^2(3) = 9.58$, $p = .022$. Les agents *Périodique* et *Fluctuant* produisaient plus de croisements dans l'Étude 2 que dans l'Étude 1 ($p < .05$).

Dans l'Étude 1, les comparaisons par paires ont révélé trois différences significatives : l'agent Périodique possède 53% de moins de croisements que l'agent Réactif ($RR = 0.47$, $p < .001$), l'agent Indépendant 52% de moins que l'agent Réactif ($RR = 0.48$, $p < .001$), et pour l'agent Réactif il a été observé 74% plus de croisements que pour l'agent Fluctuant ($RR = 1.74$, $p < .001$).

Dans l'Étude 2, l'agent Indépendant présentait 60% de moins de croisements que l'agent Réactif ($RR = 0.40$, $p < .001$), l'agent Périodique avait 48% de moins de croisements que l'agent Réactif ($RR = 0.52$, $p < .001$), l'agent Indépendant avait 37% de moins de croisements que l'agent Fluctuant ($RR = 0.63$, $p < .001$), l'agent Réactif avait 56% de plus de croisements que l'agent Fluctuant ($RR = 1.56$, $p < .001$), l'agent Périodique avait 30% de plus de croisements que l'agent Indépendant ($RR = 1.30$, $p < .001$), et l'agent Périodique avait 18% de moins de croisements que l'agent Fluctuant ($RR = 0.82$, $p < .001$).

Ensuite, les comparaisons entre études pour chaque type d'agent (ajustement de Holm) ont révélé que seuls les agents Périodique et Fluctuant étaient significativement différents entre l'Étude 1 et l'Étude 2 : l'agent Périodique dans l'Étude 1 avait 17% moins de croisements que dans l'Étude 2 ($RR = 0.83$, $p = .02$) et l'agent Fluctuant avait 17% de croisements en moins dans l'Étude 1 que dans l'Étude 2 ($RR = 0.83$, $p = .02$).

Concernant l'analyse entre le premier bloc de l'Étude 2 et l'Étude 1, l'ANOVA de type III a révélé un effet principal significatif du type d'agent ($\chi^2(3) = 90.21$, $p < .001$), mais aucun effet de l'étude ($\chi^2(1) = 0.30$, $p = .583$), ni d'interaction ($\chi^2(3) = 5.82$, $p = .121$). Les EMM globales suivent une hiérarchie similaire à celle retrouvée entre l'Étude 1 et l'Étude 2 : Réactif (45.9), Fluctuant (29.8), Périodique (23.7), Indépendant (20.9). Les comparaisons post-hoc (correction de Holm) confirment que l'agent Réactif avait significativement plus de croisements que les autres ($p < .001$ pour toutes les comparaisons). L'agent Indépendant avait 30% de moins de croisements que Fluctuant ($RR = 0.70$, $p < .001$) et l'agent Périodique avait 20% de moins de croisements que l'agent Fluctuant ($RR = 0.80$, $p = .025$).

6.2.3 Structure temporelle des vélocités

Pour rappel, trois indicateurs principaux ont été extraits de chaque série temporelle de vélocité (c'est-à-dire la vélocité du participant, celle de l'agent et la

vélocité relative) : (1) l'exposant d'échelle α issu de la *DFA*, qui reflète la présence et la force des corrélations temporelles à longue portée dans la série analysée ; (2) la pente spectrale β (déduite de α via la relation $\beta = 2\alpha - 1$), qui décrit comment la variance se distribue selon les fréquences et permet d'identifier des dynamiques de type $1/f$, sans nécessiter d'estimation spectrale séparée ; (3) la largeur du spectre multifractal Δh issue de la *MFDFA*, qui quantifie l'hétérogénéité multi-échelle (diversité des exposants locaux) et l'intermittence des fluctuations, c'est-à-dire l'alternance de rafales d'activité (grandes variations rapprochées) et de périodes calmes (faibles variations) comme par exemple des accélérations ou des ralentissements rapides dans la vélocité relative suivis de phases plus stables.

Chacun des indices α, β et Δh a été modélisé au moyen d'un LMM pour estimer les effets principaux du type d'agent et de l'étude, ainsi que leur interaction, en prenant en compte les intercepts aléatoires par participant. Les effets fixes et leur interaction ont été testés par une ANOVA de type III avec l'approximation de Satterthwaite pour les degrés de liberté, et la variance interindividuelle a été quantifiée par le coefficient de corrélation intraclasse (ICC ajusté).

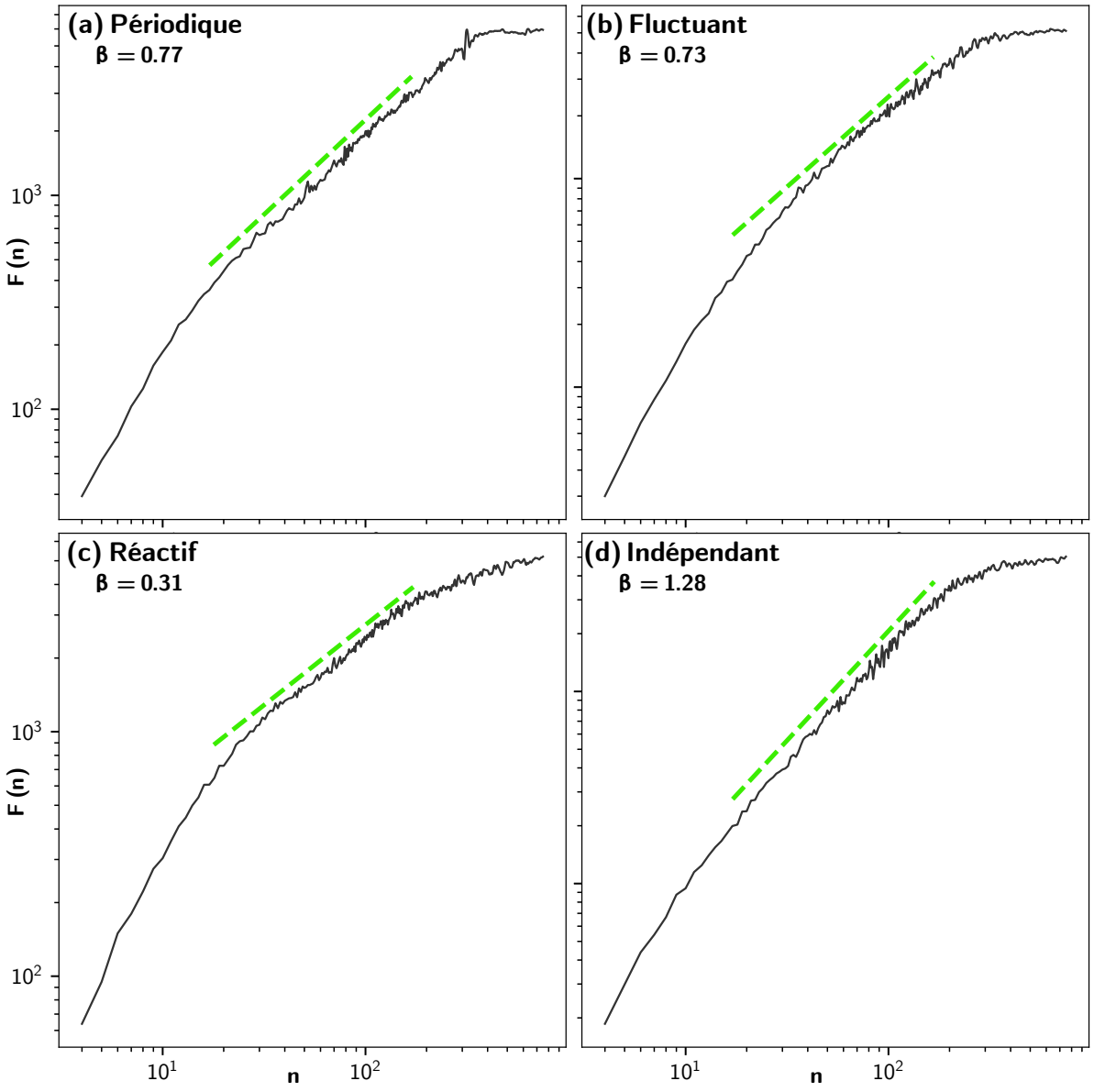
Vélocité relative

Les analyses suivantes examinent les exposants α, β et Δh extraits des séries temporelles de vélocité relative, c'est-à-dire la dérivée temporelle de la distance géodésique sur un cercle de longueur L, définie par $d(t) = \min(|x_p - x_a|, L - |x_p - x_a|)$. La Figure 6.3 contient des exemples de tracés pour la vélocité relative selon le type d'agent rencontré. Contrairement aux vélocités individuelles (qui reflètent les stratégies propres à chaque participant ou à chaque agent), cette mesure capture directement la dynamique collective émergeant de l'interaction et constitue un des indicateurs privilégiés pour caractériser la structure de la coordination sensorimotrice entre les deux agents (le participant et son partenaire agent programmé).

Exposant d'échelle α. Le diagramme en boîte Figure 6.4 présente la distribution des exposants α de la vélocité relative de la dyade (participant et agent) selon l'instruction donnée au participant (Étude 1 vs. Étude 2) et en fonction de l'agent rencontré.

Figure 6.3

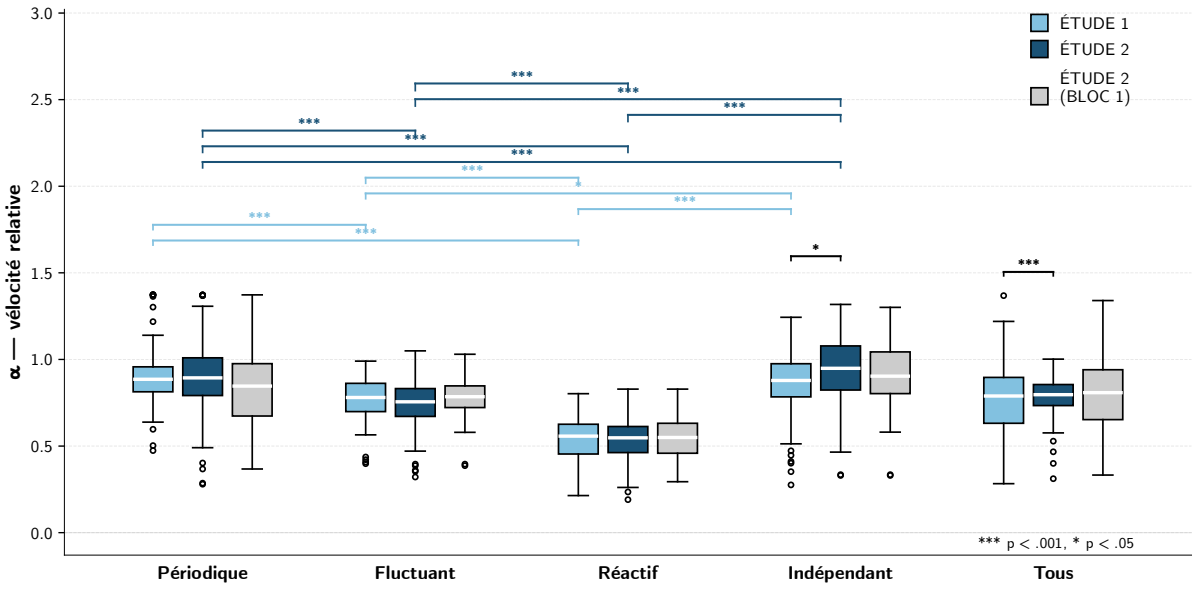
Vélocité relative : Exemples de tracés DFA (log-log) pour la vélocité relative selon le type d'agent rencontré



Note. Chaque panneau présente la relation log–log entre la fluctuation moyenne $F(n)$ et l'échelle temporelle n pour la dynamique entre un participant et quatre types d'agents : **(a)** vs. Périodique, **(b)** vs. Fluctuant, **(c)** vs. Réactif et **(d)** vs. Indépendant.

Figure 6.4

Vélocité relative : Exposant d'échelle α selon l'instruction



Pour ce qui est de l'analyse de l'α de la vélocité relative entre l'Étude 1 et l'Étude 2, l'ANOVA de type III a révélé un effet principal significatif du type d'agent, $F(3, 435.55) = 233.78$, $p < .001$, aucun effet de l'étude, ($F(1, 237.46) = 0.42$, $p = .518$), mais une interaction significative entre ces facteurs, ($F(3, 435.55) = 3.37$, $p = .018$). La variance interindividuelle était substantielle (ICC ajusté $= 0.540$), indiquant une hétérogénéité marquée entre participants. Le Tableau 6.4 présente les détails des EMM pour chaque agent et étude.

Les comparaisons post-hoc par paires au sein de chaque étude (ajustement de Holm) montrent que dans l'Étude 1, la structure avec l'agent Réactif différait de façon significative de toutes les autres (tous $p < .001$) en présentant toujours la valeur de α la plus basse. Avec les agents Périodique ($M_{\text{diff}} = 0.14$, $p < .001$) et Indépendant ($M_{\text{diff}} = 0.10$, $p < .001$) le α présentait des valeurs significativement plus élevées qu'avec l'agent Fluctuant.

Dans l'Étude 2, le même schéma émerge avec l'agent Réactif ($p < .001$), la dynamique avec cet agent est celle ayant le plus faible α. En revanche, une différence significative est présente selon l'agent : des valeurs plus élevées avec les agents Périodique et Indépendant ($M_{\text{diff}} = -0.04$, $p < .001$) qu'avec l'agent Fluctuant ($M_{\text{diff}} = 0.15$ et 0.19, $p < .001$).

Table 6.4

Vélocité relative : Exposant d'échelle α selon l'agent et l'étude

Étude	Agent	EMM	SE	$IC_{95\%}$
Étude 1	Périodique	0.90	0.02	[0.86, 0.94]
	Indépendant	0.87	0.02	[0.82, 0.91]
	Réactif	0.54	0.02	[0.49, 0.58]
	Fluctuant	0.76	0.02	[0.72, 0.80]
Étude 2	Périodique	0.89	0.02	[0.86, 0.92]
	Indépendant	0.94	0.02	[0.91, 0.97]
	Réactif	0.53	0.02	[0.50, 0.56]
	Fluctuant	0.75	0.02	[0.72, 0.78]

Note. L'interaction entre l'étude et l'agent était significative ($F(3, 435.55) = 3.37$, $p = .018$). Dans les deux études, quand l'agent rencontré est l'agent Réactif, la structure de la série de la vélocité relative présentait un exposant α nettement plus faible (proche du bruit blanc) qu'avec tous les autres agents ($p < .001$). Avec l'agent Indépendant, l'exposant α est légèrement plus élevé dans l'Étude 2 que dans l'Étude 1 ($p < .01$), indiquant que les participants ont eu une dynamique légèrement plus corrélée à long terme avec cet agent dans l'Étude 2 (avec incitation sociale).

Les comparaisons post-hoc entre études selon chaque type d'agent rencontré (ajustement de Holm) n'ont révélé qu'une seule différence significative : avec l'agent Indépendant l'exposant α était légèrement inférieur ($M_{\text{diff}} = -0.07$, SE $= 0.03$) dans l'Étude 1 par rapport à l'Étude 2 ($t(342.48) = -2.77$, $p < .01$)

Pente spectrale β. La pente spectrale β suit les mêmes tendances que l'exposant α (voir Figure 6.5). L'ANOVA de type III pour l'exposant β de la vélocité relative entre l'Étude 1 et l'Étude 2 a révélé un effet principal du type d'agent ($F(3, 435.55) = 233.78$, $p < .001$), aucun effet de l'étude ($F(1, 237.46) = 0.42$, $p = .518$), mais une interaction entre ces facteurs ($F(3, 435.55) = 3.37$, $p = .018$). La variance interindividuelle restait assez importante (ICC ajusté$= 0.540$). Le détail des EMM par étude et selon l'agent rencontré est disponible au Tableau 6.5. Les comparaisons post-hoc (ajustement de Holm) entre agents suivent les mêmes types de schémas que pour α. Également, les comparaisons entre études pour chaque agent (ajustement de Holm) confirment que seul l'agent Indépendant diffère entre les deux contextes avec un β légèrement inférieur dans l'Étude 1 par rapport à l'Étude 2 ($M_{\text{diff}} = -0.14$, SE $= 0.05$, $t(342.48) = -2.77$, $p < .01$).

Figure 6.5

Vélocité relative : Pente spectrale β selon l'instruction

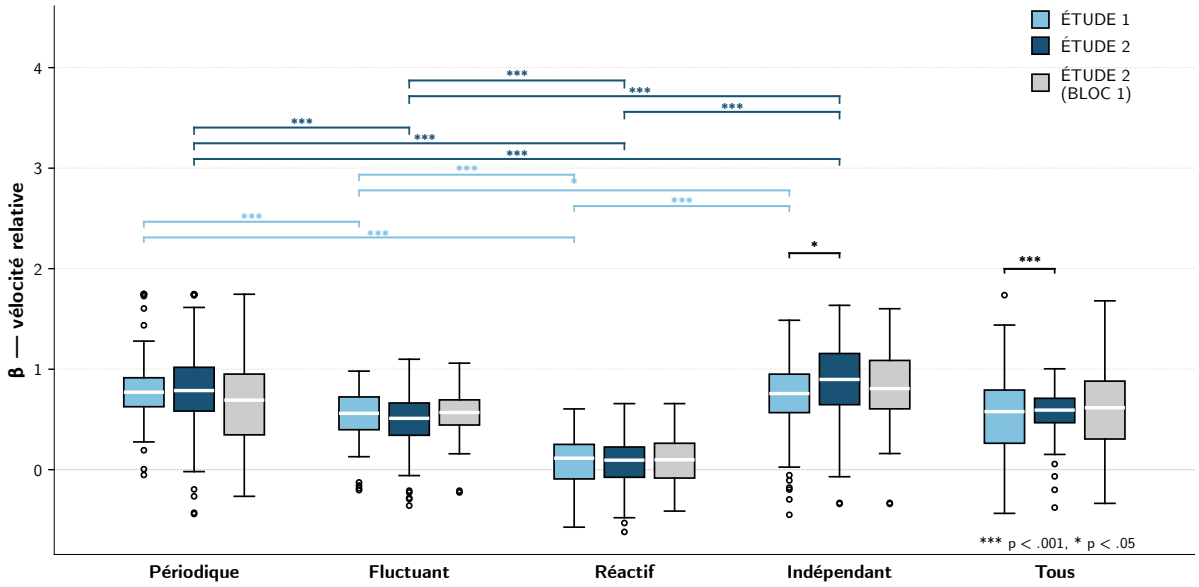


Table 6.5

Vélocité relative : Exposant spectral β selon l'agent et l'étude

Étude	Agent	EMM	SE	$IC_{95\%}$
Étude 1	vs. Périodique	0.81	0.04	[0.73, 0.89]
	vs. Indépendant	0.73	0.04	[0.65, 0.81]
	vs. Réactif	0.07	0.04	[-0.01, 0.16]
	vs. Fluctuant	0.53	0.04	[0.44, 0.61]
Étude 2	vs. Périodique	0.79	0.03	[0.73, 0.85]
	vs. Indépendant	0.88	0.03	[0.82, 0.94]
	vs. Réactif	0.07	0.03	[0.01, 0.13]
	vs. Fluctuant	0.49	0.03	[0.43, 0.56]

Largeur du spectre multifractal Δh. Pour l'indice Δh issu de la *MFDFA* (voir Figure 6.6), l'ANOVA de type III a mis en évidence un effet principal très significatif du type d'agent, $F(3, 585.81) = 70.01$, $p < .001$. À la différence de α et β, il n'y a ni effet principal de l'étude ($F(1, 237.83) = 0.43$, $p = .513$) ni interaction entre agent et étude ($F(3, 585.81) = 1.02$, $p = .384$). La variance interindividuelle reste notable (ICC ajusté$= 0.359$). Les EMM de Δh selon chaque agent rencontré sont présentées dans le Tableau 6.6.

Table 6.6

Vélocité relative : Largeur du spectre multifractal (Δh) selon l'agent

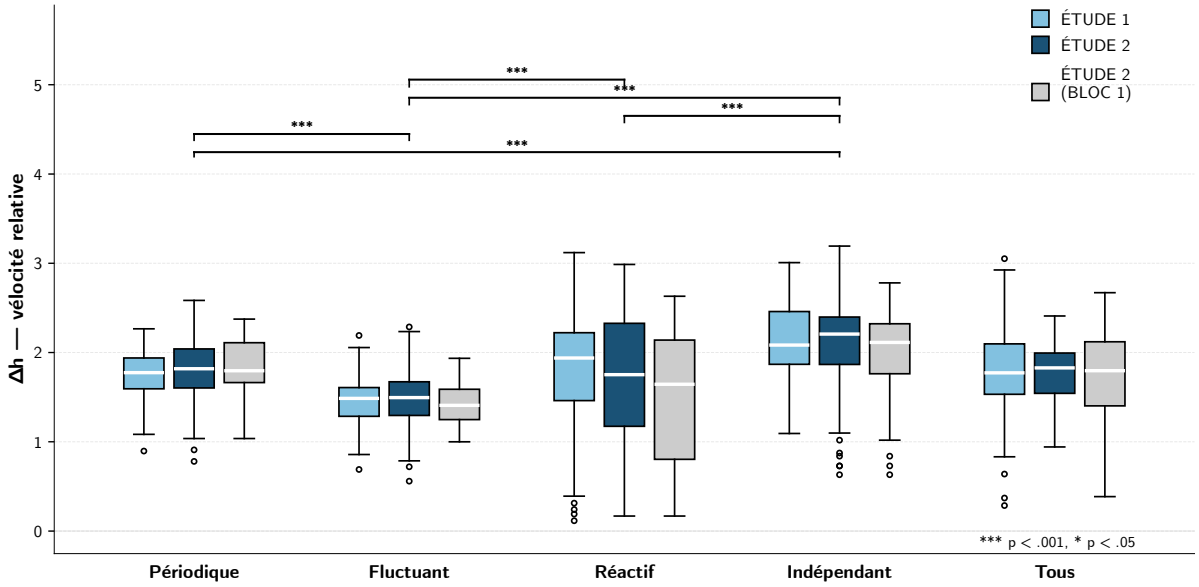
Agent	EMM	SE	$IC_{95\%}$
Périodique	1.77	0.03	[1.71, 1.84]
Indépendant	2.11	0.03	[2.04, 2.18]
Réactif	1.76	0.03	[1.69, 1.83]
Fluctuant	1.47	0.03	[1.41, 1.54]

Note. Seul l'effet principal du type d'agent était significatif, ($F(3, 585.81) = 70.01$, $p < .001$). Cependant, l'interaction avec l'étude n'était pas significative, $F(3, 585.81) = 1.02$, $p = .384$. La largeur du spectre multifractal est plus élevée avec l'agent Indépendant qu'avec tous les autres agents ($p < .001$).

Les comparaisons post-hoc (ajustement de Holm) sur l'effet de l'agent indiquent qu'avec l'agent Indépendant, la structure de la dyade présente un Δh plus élevé qu'avec l'agent Périodique ($M_{\text{diff}} = 0.34$, SE $= 0.04$, $t(583.40) = 7.64$, $p < .001$), Réactif ($M_{\text{diff}} = 0.35$, SE $= 0.04$, $t(578.92) = 7.91$, $p < .001$) et Fluctuant ($M_{\text{diff}} = 0.63$, SE $= 0.04$, $t(581.60) = 14.38$, $p < .001$). En revanche, entre la dyade avec l'agent Périodique et celle avec l'agent Réactif, la largeur du spectre multifractal ne diffère pas ($M_{\text{diff}} = 0.02$, $p = .714$), mais celle avec l'agent Réactif présente un Δh plus élevé qu'avec l'agent Fluctuant ($M_{\text{diff}} = 0.28$, SE $= 0.04$, $t(580.73) = 6.37$, $p < .001$). Ensuite, avec l'agent Périodique il a été retrouvé un Δh plus élevé que dans le cas d'une dyade avec l'agent Fluctuant ($M_{\text{diff}} = 0.30$, SE $= 0.04$, $t(585.32) = 6.82$, $p < .001$). En somme, sans différence entre études, le Δh le plus élevé est celui avec l'agent Indépendant et le plus faible est celui avec l'agent Fluctuant.

Figure 6.6

Vélocité relative : Largeur du spectre multifractal Δh selon l'instruction



Vélocité de l'agent

Les analyses suivantes portent sur les exposants α, β et Δh extraits des séries temporelles de vélocité de l'agent. Ces mesures permettent de caractériser la structure intrinsèque du comportement de chaque agent. La vélocité de l'agent caractérise sa dynamique propre et permet de vérifier que les différents agents présentent effectivement des signatures temporelles distinctes conformément à leur conception algorithmique.

Exposant d'échelle α. Concernant l'exposant α de la vélocité de l'agent pour les données relatives à l'Étude 1 et l'Étude 2 (voir le diagramme en boîte Figure 6.7 pour la distribution), l'analyse a révélé une faible variance interindividuelle (ICC$= 0.090$). Ce résultat est attendu puisque trois agents sur quatre ont des comportements qui ne « dépendent » pas des croisements avec le participant (à l'exception de l'agent Réactif). L'ANOVA de type III a révélé un effet principal significatif de l'agent ($F(3, 894.54) = 17290.19$, $p < .001$), mais aucun effet de l'étude ni d'interaction. Les EMM sont présentées dans le Tableau 6.7 pour chaque agent. Les comparaisons post-hoc (Holm) indiquent que toutes les paires d'agents diffèrent significativement ($p < .001$) avec l'agent Réactif qui

présentait toujours la valeur α la plus basse, suivi de l'agent Fluctuant, puis des agents Indépendant et Périodique.

Table 6.7

Vélocité de l'agent : Exposant α

Agent	EMM	SE	$IC_{95\%}$
Périodique	1.371	0.006	[1.364, 1.378]
Indépendant	1.307	0.006	[1.300, 1.314]
Réactif	0.385	0.006	[0.378, 0.392]
Fluctuant	1.012	0.006	[1.006, 1.019]

Pente Spectrale β. Concernant la pente spectrale β (voir Figure 6.8), celle-ci suit également les mêmes tendances que l'exposant α (ICC $= 0.090$), avec un effet principal très significatif du type d'agent, $F(3, 894.54) = 17290.19$, $p < .001$, sans effet principal de l'étude ni d'interaction significative. Les comparaisons post-hoc indiquent un même schéma que pour α, toutes les paires d'agents étant significativement différentes ($p < .001$) avec l'agent Réactif ayant la valeur de β la plus négative suivi de l'agent Fluctuant, puis des agents Indépendant et Périodique. Le Tableau 6.8 présente les EMM de chaque agent.

Table 6.8

Vélocité de l'agent : Exposant spectral β

Agent	EMM	SE	$IC_{95\%}$
Périodique	1.741	0.011	[1.728, 1.755]
Indépendant	1.614	0.011	[1.600, 1.627]
Réactif	-0.230	0.011	[-0.244, -0.216]
Fluctuant	1.025	0.011	[1.011, 1.039]

Largeur du spectre multifractal Δh. Concernant l'indice Δh (voir le diagramme en boîte Figure 6.9 pour la distribution), tout comme pour α et β, l'ANOVA de type III a mis en évidence un effet principal très significatif du type d'agent, $F(3, 517.72) = 254.98$, $p < .001$, sans effet de l'étude ni d'interaction significative. Les comparaisons post-hoc (ajustement de Holm) sur l'effet de l'agent indiquent

que tous les agents diffèrent entre eux ($p < .001$), avec l'agent Périodique ayant la plus grande largeur de spectre, suivi de l'agent Fluctuant, Indépendant, puis de l'agent Réactif (voir le Tableau 6.9 pour les EMM des Δh de chaque agent).

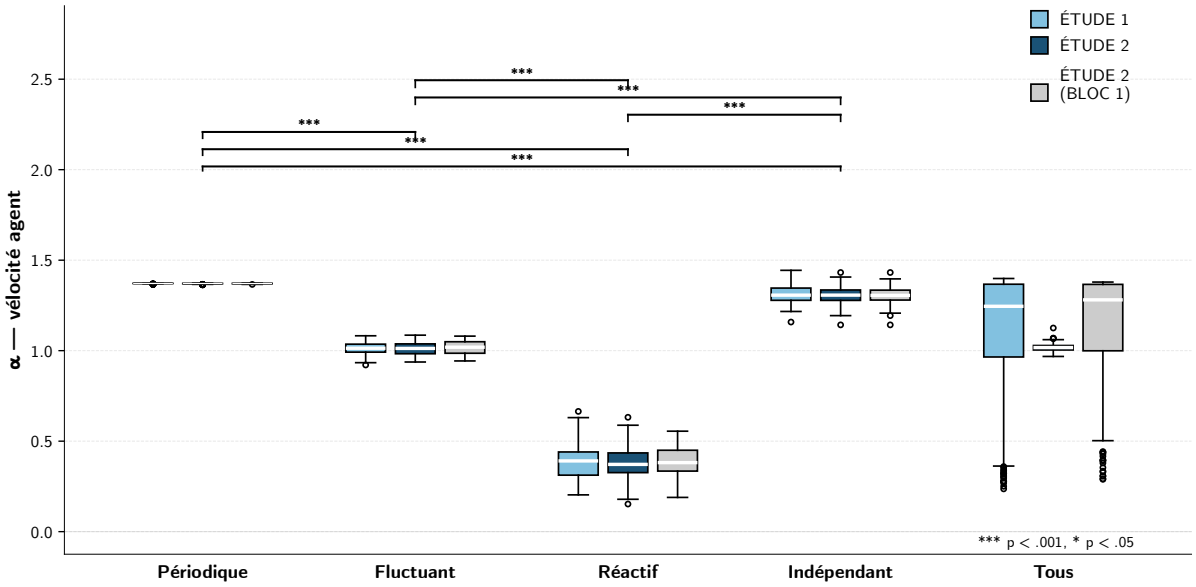
Table 6.9

Vélocité de l'agent : Largeur du spectre multifractal Δh

Agent	EMM	SE	$IC_{95\%}$
Périodique	2.005	0.020	[1.965, 2.045]
Indépendant	1.565	0.021	[1.524, 1.606]
Réactif	1.316	0.021	[1.275, 1.357]
Fluctuant	1.757	0.021	[1.717, 1.797]

Figure 6.7

Vélocité des agents : Exposant d'échelle α selon l'instruction



En somme, en guise de synthèse concise, les trois exposants caractérisant la structure de la vélocité de l'agent convergent vers un profil globalement très structuré et hautement corrélé. L'agent Réactif présente systématiquement les valeurs d'α et de β les plus faibles, caractéristiques d'une dynamique quasi anti-persistante, proche d'un bruit blanc ou même plus aléatoire, avec une faible organisation multi-échelles.

Figure 6.8

Vélocité des agents : Pente spectrale β selon l'instruction

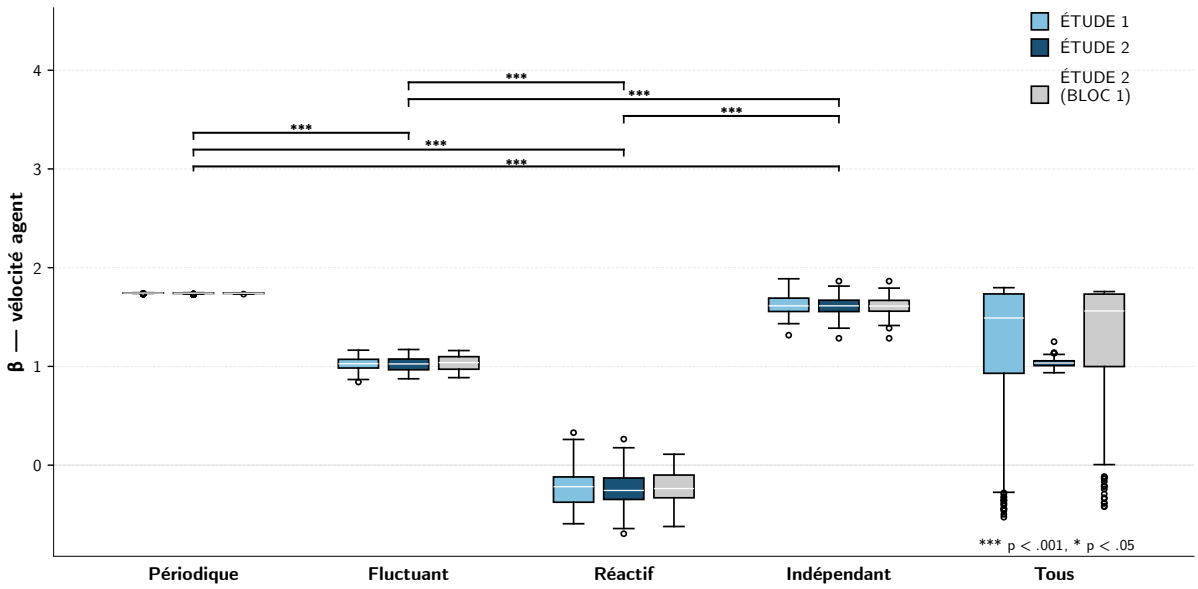
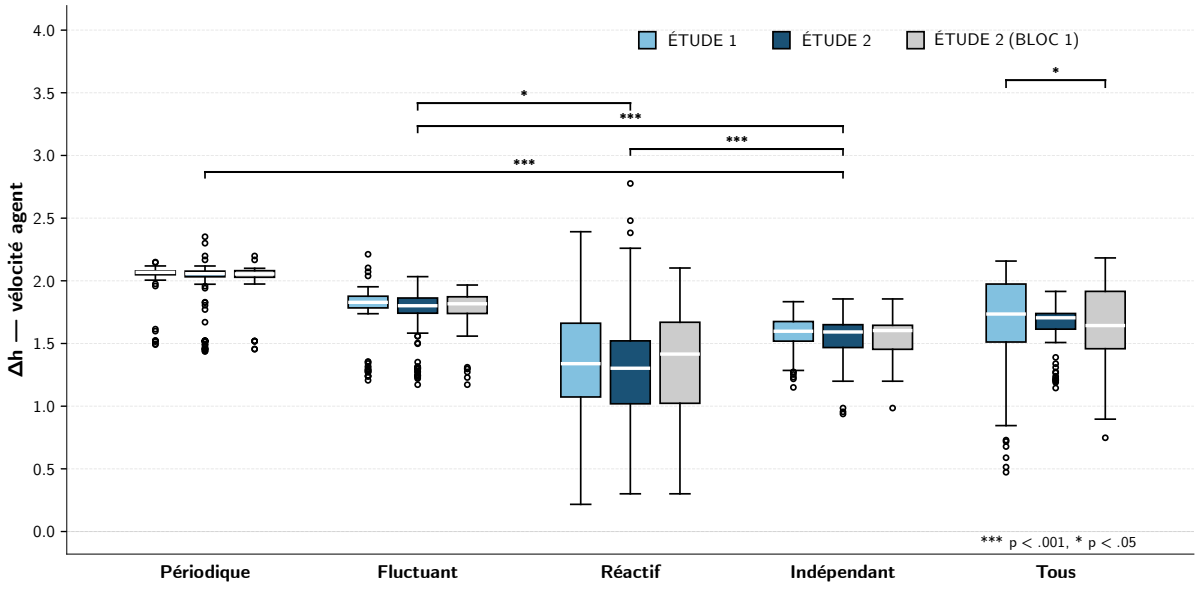


Figure 6.9

Vélocité des agents : Largeur du spectre multifractal Δh selon l'instruction



Quant à l'agent Fluctuant, il montre des valeurs intermédiaires s'apparentant à un bruit rose ($1/f$) dont les fluctuations de vélocité sont corrélées à long terme, traduisant une auto-organisation modérée et une coordination interne stable avec de la variabilité.

Son homologue plus régulier, l'agent Périodique, se caractérise par des valeurs élevées proches du bruit brownien, indiquant une dynamique hautement corrélée et régulière, avec une structure multi-échelles riche imposée par l'oscillation.

Enfin, l'agent Indépendant, « libre » d'explorer dans tout l'environnement conserve un profil également corrélé, au niveau des autocorrélations de ses séries, il s'apparente à un bruit brownien, avec une largeur de spectre plus restreinte que celle de l'agent Périodique, suggérant une complexité temporelle moindre et une variabilité interne stable.

Vélocité du participant

La vélocité du participant reflète les dynamiques d'exploration individuelle, indépendamment du comportement de l'agent. L'analyse de cette variable permet de déterminer si les différences observées dans la vélocité relative résultent principalement des variations dans le comportement de l'agent ou si elles impliquent également des ajustements dans l'activité du participant.

L'ANOVA de type III pour l'exposant α de la vélocité du participant entre l'Étude 1 et l'Étude 2 (voir le diagramme en boîte Figure 6.10 pour la distribution) n'a révélé aucun effet du type d'agent ($F(3, 342.80) = 0.93$, $p = .425$), aucun effet de l'étude ($F(1, 259.59) = 0.20$, $p = .653$), ni interaction significative entre les deux facteurs ($F(3, 342.80) = 0.48$, $p = .698$). La variance interindividuelle était élevée (ICC ajusté$= 0.766$), indiquant une forte hétérogénéité entre participants.

Pour la pente spectrale β (voir le diagramme en boîte Figure 6.11 pour la distribution), l'analyse n'a révélé aucun effet du type d'agent, ($F(3, 342.80) = 0.93$, $p = .425$), de l'étude, ($F(1, 259.59) = 0.20$, $p = .653$), ni d'interaction entre ces deux facteurs ($F(3, 342.80) = 0.48$, $p = .698$).

Également, pour le Δh ($ICC = 0.504$) (voir le diagramme en boîte Figure 6.12 pour la distribution), aucun effet significatif du type d'agent, $F(3, 449.46) = 1.36$, $p = .254$, ni effet principal de l'étude ($F(1, 229.50) = 0.44$, $p = .507$) ni de l'interaction entre ces deux facteurs ($F(3, 449.46) = 1.33$, $p = .264$) n'a été trouvé.

Figure 6.10

Vélocité des participants : Exposant d'échelle α selon l'instruction et le partenaire

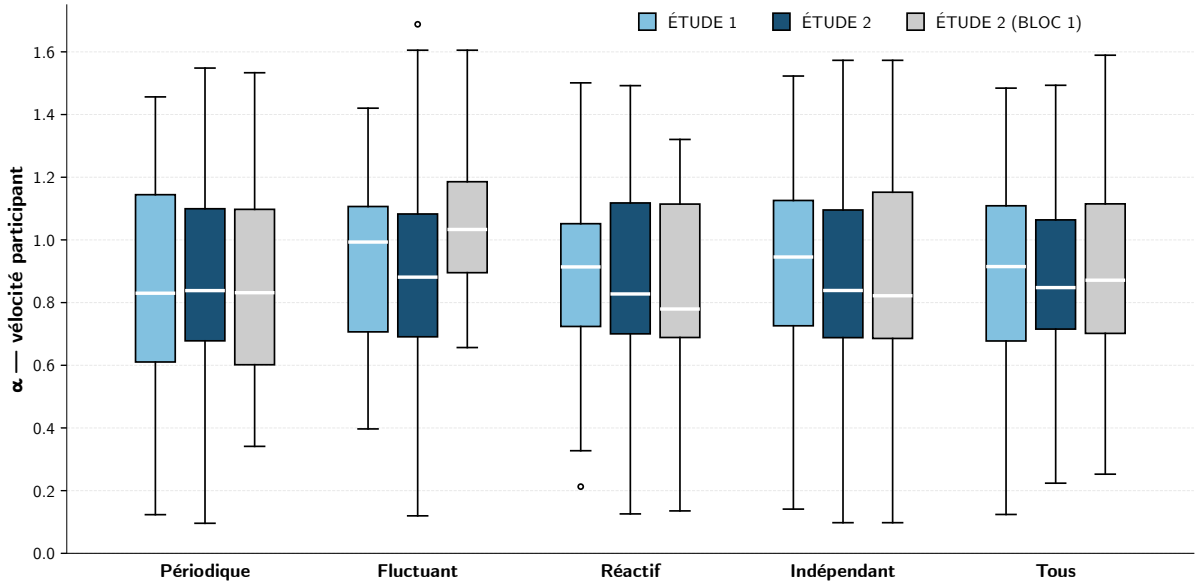


Figure 6.11

Vélocité des participants : Pente spectrale β selon l'instruction et le partenaire

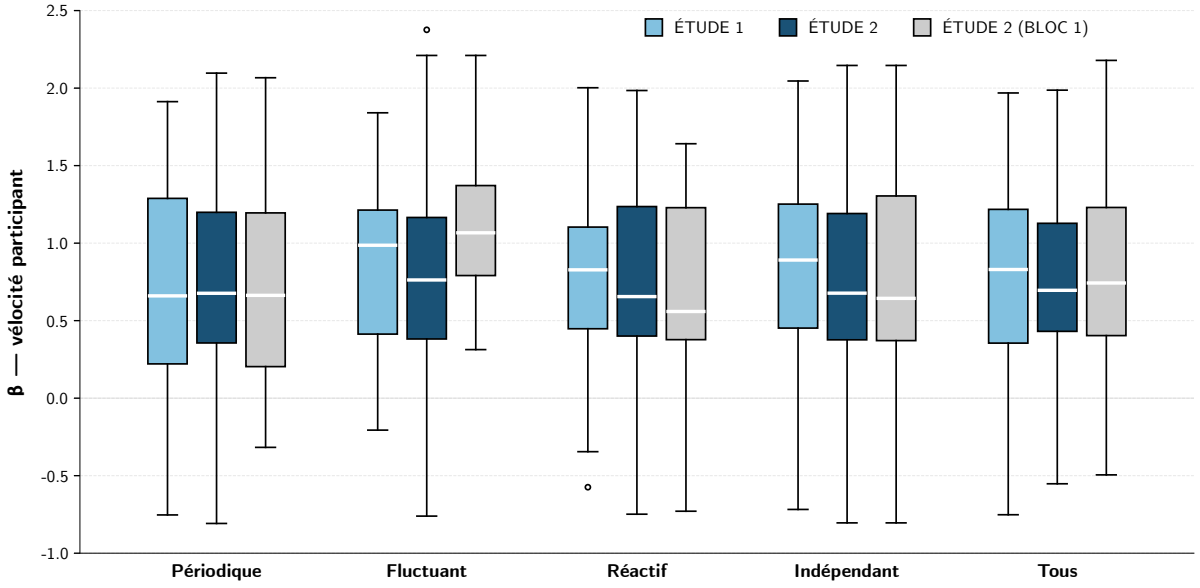
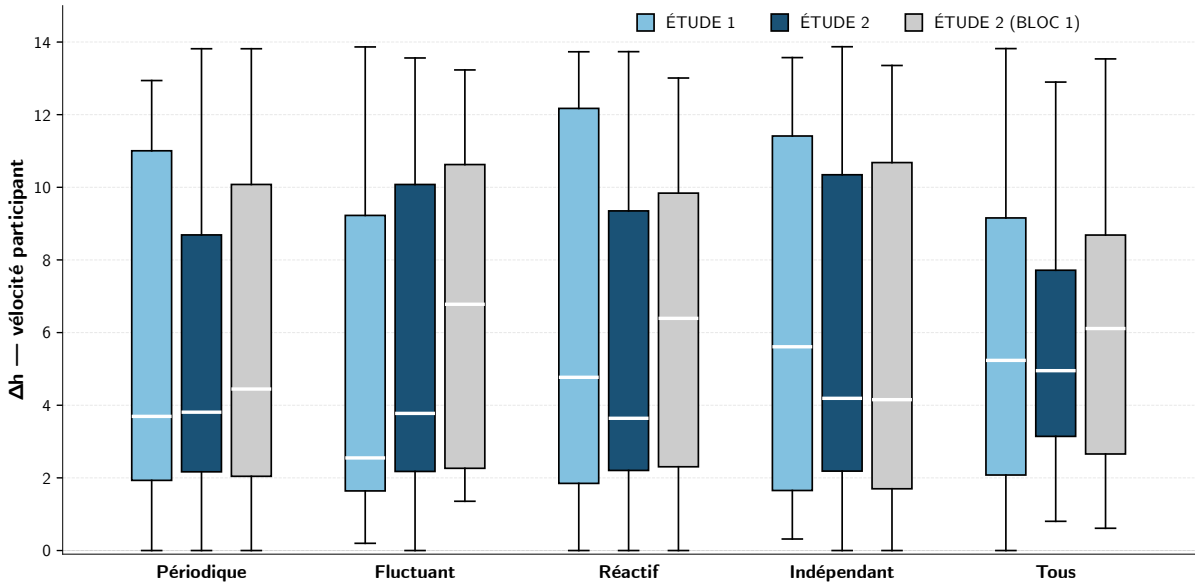


Figure 6.12

Vélocité des participants : Largeur du spectre multifractal Δh selon l'instruction et le partenaire



En synthèse sur la vélocité du participant lui-même, les analyses ne révèlent aucune modulation significative par le type d'agent rencontré ou par l'instruction expérimentale (avec ou sans incitation sociale). La forte variance interindividuelle suggère que les participants adoptent des stratégies d'exploration idiosyncratiques (personnelles) et stables, peu influencées par les propriétés dynamiques de l'agent ou par l'incitation sociale explicite. Cette absence d'effet contraste fortement avec les motifs observés dans la vélocité relative, suggérant que les signatures fractales distinctives émergent probablement au niveau dyadique (c'est-à-dire dans la coordination conjointe) plutôt qu'au niveau individuel.

Comparaison des structures temporelles des vélocités entre l'Étude 1 et le premier bloc de l'Étude 2

Cette comparaison entre l'Étude 1 et le premier bloc de l'Étude 2 constitue une analyse complémentaire structurellement équivalente, les deux conditions comportant un seul essai par participant et un unique type d'agent. Elle offre ainsi une base de comparaison plus propre que l'analyse complète, permettant

de vérifier que les structures temporelles observées traduisent des dynamiques d'interaction initiales plutôt qu'une adaptation progressive aux agents artificiels (ou à la tâche elle-même). Pour une représentation graphique des distributions des différents exposants (Étude 1 vs. Étude 2 bloc 1), se référer aux figures dans chaque sous-sous-section de vélocité comparant Étude 1 et Étude 2 (vélocité relative en 6.2.3, vélocité des agents en 6.2.3, vélocité des participants 6.2.3).

Exposant α de la vélocité relative. Pour l'exposant α de la vélocité relative, entre l'Étude 1 et le bloc 1 de l'Étude 2, l'analyse a révélé une variance interindividuelle plus élevée que pour l'analyse complète précédente (ICC ajusté$= 0.758$). L'ANOVA de type III a mis en évidence un effet principal très significatif du type d'agent, ($F(3, 278.99) = 59.78$, $p < .001$), mais aucun effet de l'étude ($F(1, 278.99) = 0.10$, $p = .752$), ni d'interaction significative ($F(3, 278.99) = 1.31$, $p = .273$), contrairement à l'analyse complète où l'interaction était significative.

Les comparaisons post-hoc (avec ajustement de Holm) indiquent que lors de l'interaction avec l'agent Réactif, la structure fractale diffère significativement de tous les autres ($p < .001$) avec les valeurs les plus faibles. Ensuite pour les autres comparaisons significatives, avec l'agent Fluctuant, α présentait des valeurs significativement plus élevées qu'avec l'agent Réactif ($M_{\text{diff}} = 0.22$, SE$= 0.03$, $t(287.24) = 7.18$, $p < .001$), mais inférieures avec les agents Périodique ($M_{\text{diff}} = -0.11$, SE $= 0.03$, $t(287.24) = -3.67$, $p < .001$) et Indépendant ($M_{\text{diff}} = -0.12$, SE $= 0.03$, $t(287.24) = -4.34$, $p < .001$). Les EMM de α pour la vélocité relative selon l'agent, sont disponibles dans le Tableau 6.10.

Table 6.10

Vélocité relative au bloc 1 : Exposant α

Agent	EMM	SE	$IC_{95\%}$
vs. Périodique	0.877	0.020	[0.84, 0.92]
vs. Indépendant	0.891	0.018	[0.86, 0.93]
vs. Réactif	0.546	0.022	[0.50, 0.59]
vs. Fluctuant	0.768	0.022	[0.72, 0.81]

Exposant β de la vélocité relative. Pour la pente spectrale β, conformément à α, l'ANOVA a révélé un effet principal du type d'agent ($F(3, 279) = 59.78, p < .001$),

sans effet de l'étude ($F(1, 279) = 0.10$, $p = .752$), ni interaction ($F(3, 279) = 1.31$, $p = .273$), (ICC ajusté$= 0.758$). Les comparaisons post-hoc suivent les mêmes schémas que pour α. Les EMM de β pour la vélocité relative selon l'agent rencontré sont disponibles dans le Tableau 6.11.

Table 6.11

Vélocité relative au bloc 1 : Exposant β

Agent	EMM	SE	$IC_{95\%}$
vs. Périodique	0.754	0.040	[0.67, 0.83]
vs. Indépendant	0.782	0.037	[0.71, 0.85]
vs. Réactif	0.091	0.044	[0.01, 0.18]
vs. Fluctuant	0.535	0.044	[0.45, 0.62]

Largeur du spectre multifractal Δh de la vélocité relative. Comparé à l'analyse complète (où seul l'effet de l'agent était significatif), concernant le Δh, l'ANOVA a révélé un effet principal du type d'agent ($F(3, 279) = 28.63$, $p < .001$), un effet principal de l'étude ($F(1, 279) = 4.90$, $p < .05$) et un effet marginal de l'interaction entre ces deux facteurs ($F(3, 279) = 2.34$, $p = .073$), (ICC ajusté$= 0.520$). Les comparaisons post-hoc montrent que toutes les paires selon l'agent diffèrent significativement ($p < 0.01$) à l'exception de la paire Périodique et Réactif qui est marginalement significative ($M_{\text{diff}} = 0.14$, SE $= 0.07$, $t(287.24) = 1.86$, $p = .063$). De même, le Δh est plus grand dans l'Étude 2 que dans l'Étude 1 ($M_{\text{diff}} = 0.11$, SE $= 0.05$, $t(287.24) = 2.81$, $p < .05$).

En somme, l'analyse entre le premier bloc de l'Étude 2 et l'Étude 1 révèle des schémas similaires à l'analyse complète pour la vélocité relative. La dyade avec l'agent Réactif montre les plus faibles corrélations et celle avec l'agent Indépendant la plus grande multifractalité. L'effet de l'étude sur Δh suggère une légère différence globale entre l'Étude 1 et le premier bloc de l'Étude 2, potentiellement due à des facteurs contextuels, contrastant avec l'absence d'effet de l'étude dans l'analyse complète.

Vélocité de l'agent : α, β, Δh. Lorsque l'on s'intéresse à la vélocité de l'agent (bloc 1 de l'Étude 2 vs. Étude 1), l'ANOVA montre un effet principal du type d'agent ($F(3, 279) = 6014.06$, $p < .001$), mais aucun effet de l'étude ($F(1, 279) = 0.02$,

$p = .882$) ni interaction entre ces deux facteurs ($F(3, 279) = 0.26$, $p = .853$). Comparée à l'analyse complète (mêmes effets), les EMM sont quasiment identiques et les analyses post-hoc révèlent également les mêmes tendances (toutes les paires d'agents diffèrent significativement $p < .001$), confirmant la potentielle (ICC$= 0.306$) stabilité des dynamiques algorithmiques des agents.

Pour la pente spectrale β, conformément à l'exposant α et aux résultats de l'analyse complète, l'ANOVA révèle un effet du type d'agent, ($F(3, 279) = 6014.06$, $p < .001$), mais aucun effet principal de l'étude ni interaction. Les tests post-hoc indiquent également le même schéma : Toutes les paires d'agents sont significativement différentes ($p < .001$) avec l'agent Réactif présentant la pente spectrale la plus faible ($\beta = -0.22$, SE $= 0.012$, $IC_{95\%} = [-0.250, -0.201]$)

Concernant la largeur du spectre Δh, l'ANOVA a révélé un effet principal du type d'agent ($F(3, 279) = 81.56$, $p < .001$), aucun effet principal de l'étude ni de l'interaction (ICC$= 0.555$). Les tests post-hoc indiquent également le même schéma, où toutes les paires d'agents sont significativement différentes ($p < .001$) avec l'agent Réactif présentant la valeur la plus faible ($\Delta h = 1.350$, SE = 0.033, $IC_{95\%} = [1.283, 1.413]$) et les autres agents présentant des valeurs fortement similaires à l'analyse complète.

Vélocité du participant : α, β, Δh. Pour la vélocité du participant, l'ANOVA de type III de l'exposant α n'a révélé aucun effet significatif (effet marginal seulement) du type d'agent ($F(3, 279.00) = 2.33$, $p = .075$), de l'étude ($F(1, 279.00) = 0.16$, $p = .685$), ni interaction ($F(3, 279.00) = 1.10$, $p = .348$). La variance interindividuelle était encore plus élevée que dans l'analyse complète (ICC ajusté $= 0.840$), indiquant une hétérogénéité encore plus forte entre les participants.

Les analyses de β et Δh conduisent aux mêmes conclusions : aucun effet significatif du type d'agent, de l'étude ou de leur interaction. Pour la pente spectrale β, les analyses n'indiquent aucun effet significatif du type d'agent ($F(3, 279.00) = 2.33$, $p = .074$), aucun effet principal de l'étude ($F(3, 278.93) = 0.16$, $p = .685$), ni d'interaction ($F(3, 279.00) = 1.10$, $p = .350$). Quant à Δh (ICC $= 0.459$), il n'est observé aucun effet du type d'agent ($F(3, 278.99) = 0.20$, $p = .9$), de l'étude ($F(1, 278.99) = 0.28$, $p = .59$) ni d'interaction entre ces deux facteurs ($F(3, 278.99) = 1.05$, $p = .37$).

En somme, comme dans l'analyse principale, la vélocité du participant reste invariante aux types d'agents et aux types d'instructions (avec ou sans incitation sociale), tandis que les signatures temporelles distinctives apparaissent au niveau de la dynamique issue des agents (participant et son partenaire agent) et non du comportement individuel des participants.

6.3 Discussion

Les travaux décrits dans cette seconde partie avaient pour objectif central d'examiner comment l'instruction donnée au participant et les propriétés dynamiques d'un agent artificiel influencent conjointement (1) le sentiment de *Présence Sociale* ressenti (plus précisément, les dimensions de coprésence et d'interdépendance comportementale perçue) et (2) les dynamiques comportementales et temporelles de l'interaction dans un environnement minimaliste inspiré du paradigme du *Perceptual Crossing*.

À travers deux études expérimentales complémentaires manipulant l'instruction (consigne neutre ou incitation sociale explicite) et le type d'agent (Périodique, Fluctuant, Réactif ou Indépendant), nous avons articulé trois niveaux d'analyse : (1) l'expérience subjective de *Présence Sociale* mesurée par questionnaire, (2) le nombre de croisements et (3) les structures fractales et multifractales des séries temporelles de la vélocité propre à chaque agent et de la vélocité relative (entre l'agent et le participant).

6.3.1 Présence sociale

Les analyses du questionnaire de *Présence Sociale* révèlent un effet de l'instruction expérimentale sur les deux dimensions mesurées. Les participants ayant reçu l'instruction explicite de tenter d'interagir avec l'autre (dans l'Étude 2) rapportent des scores significativement plus élevés de coprésence et d'interdépendance comportementale perçue (réponse « un peu d'accord ») comparativement à ceux ayant reçu l'instruction neutre d'exploration (dans l'Étude 1). Ces effets demeurent significatifs même dans l'analyse comparant uniquement le premier bloc de l'Étude 2 à l'Étude 1 : Il est à noter que pour la dimension de coprésence la significativité était devenue marginale. Néanmoins, cela suggère

que l'amplification de la *Présence Sociale* par l'instruction explicite n'est pas simplement due à un effet cumulatif d'exposition répétée aux quatre agents ou à des effets d'apprentissage, mais reflète un cadrage initial de l'expérience.

Ces résultats étendent les travaux antérieurs sur le *Perceptual Crossing* en démontrant que l'incitation sociale constitue un facteur contextuel modulant l'expérience subjective de *Présence Sociale* dans un tel environnement minimaliste. Les travaux d'Auvray et al. (2009) et de Bedia et al. (2014) fournissaient systématiquement une instruction à caractère sociale explicite dès le début de l'expérience, créant d'emblée une attente sociale. Notre manipulation permet d'isoler l'effet propre de cette instruction et révèle qu'elle joue un rôle dans l'orientation de l'attention des participants.

Cependant, et c'est un résultat important, les scores de *Présence Sociale* dans l'Étude 1 (avec instruction neutre) demeurent globalement au-dessus du point milieu de l'échelle (c'est-à-dire entre la neutralité et la réponse « un peu d'accord »), confirmant une de nos hypothèse : même en l'absence d'instruction sociale explicite, les participants rapportent un sentiment de coprésence et tendre à percevoir une interdépendance comportementale. Ce résultat suggère que la coordination sensorimotrice dans un environnement partagé semble suffire à engendrer une forme basique de *Présence Sociale*, sans nécessiter d'attente sociale préalable explicite.

Pour la coprésence, bien que l'ANOVA révèle un effet principal significatif du type d'agent, aucune comparaison par paire ne demeure significative après correction pour comparaisons multiples. Seule une différence marginalement significative en comparant l'agent Réactif et l'agent Périodique est retrouvée. Pour l'interdépendance comportementale, aucun effet du type d'agent n'est détecté. De plus, aucune interaction significative entre l'instruction et le type d'agent n'émerge pour ces deux dimensions.

Ces résultats contrastent avec l'hypothèse selon laquelle l'agent Réactif, qui est l'agent contingent aux croisements avec le participant, devrait susciter des scores de *Présence Sociale* plus élevés que les agents non-contingents. Pourtant, les comportements au niveau du nombre de croisements révèlent une sensibilité implicite à ces différences. Il est possible que la simple présence d'un mouvement dans l'environnement partagé suffise à engendrer une attribution minimale d'agentivité et un sentiment de *Présence Sociale*, et ce, indépendamment

de la contingence du comportement de l'agent à celui du participant ou bien des propriétés de l'agent. Cela serait cohérent avec les travaux sur l'anthropomorphisme et la *Posture Intentionnelle* (Dennett, 1987 ; Epley et al., 2007) d'après lesquels les humains ont une propension à attribuer des états mentaux et une agentivité même à des entités non-humaines.

De même, il est possible que, dans un environnement aussi minimaliste et dénué d'indices sociaux riches (pas de représentation visuelle anthropomorphe, pas de communication symbolique), les participants ne disposent pas de critères suffisamment saillants pour différencier les agents sur la base de leurs propriétés dynamiques seules lors de l'évaluation subjective post-essai. Contrairement aux travaux de Froese et al. (2014, 2020) dans lesquels les participants devaient juger la clarté de leur perception de l'autre immédiatement après avoir cliqué, notre questionnaire était administré après l'ensemble du bloc d'interaction, introduisant potentiellement un délai dans la mémoire qui aurait pu atténuer le sentiment.

Deuxièmement, bien que l'agent Réactif soit contingent aux actions du participant, cette contingence se manifeste principalement par une propension accrue à générer des croisements (comme en témoignent nos résultats comportementaux), mais peut-être pas par une « qualité » de coordination radicalement différente de celle induite par les autres agents. Autrement dit, la fréquence de contact diffère, mais la dynamique temporelle ressentie des échanges pourrait être perçue comme suffisamment similaire pour ne pas générer de différences marquées dans les jugements de *Présence Sociale*.

Troisièmement, il est possible que les propriétés intrinsèques de l'environnement unidimensionnel (notamment le feedback audiovisuel) limitent l'expression de motifs de coordination suffisamment distinctifs pour être subjectivement discriminables : les travaux d'Auvray et al. (2009) et d'autres comme ceux de Froese et al. (2014) utilisaient également des environnements unidimensionnels mais incluaient des objets leurres ou fixes qui pouvaient servir de points de contraste, facilitant potentiellement la reconnaissance du partenaire humain par opposition aux autres éléments. Dans notre protocole, chaque essai ne présentait qu'un seul agent à la fois, ce qui pourrait avoir réduit la capacité des participants à développer des critères de comparaison stables (surtout dans l'Étude 1 où il n'y a qu'un seul agent rencontré)

6.3.2 Nombre de croisement

Contrairement aux mesures subjectives de *Présence Sociale*, les analyses du nombre de croisements (chevauchements spatiaux entre participant et agent) révèlent un effet massif et consistant du type d'agent.

L'ANOVA révèle un effet principal hautement significatif du type d'agent et une interaction significative entre le type d'agent et l'étude, bien que l'effet principal de l'étude soit seulement marginal. Les moyennes marginales estimées révèlent un classement net et cohérent à travers les deux études : l'agent Réactif génère le nombre de croisements le plus élevé, suivi de l'agent Fluctuant, puis des agents Périodique et Indépendant. Les agents non-contingents (Périodique, Indépendant) suivent des trajectoires indépendantes du participant, rendant les rencontres plus sporadiques et moins durables mais l'agent Fluctuant, bien que non-contingent lui aussi, présente un nombre de croisements intermédiaire, possiblement parce que son comportement oscillatoire d'amplitude variable augmente la probabilité de rencontres fortuites comparativement aux trajectoires plus prévisibles et cloisonnées de l'agent Périodique ou totalement « libres » de l'agent Indépendant.

L'interaction significative entre le type d'agent et l'étude révèle que l'instruction sociale amplifie sélectivement le nombre de croisements pour certains agents. Les comparaisons entre études par type d'agent montrent que seuls les agents Périodique et Fluctuant présentent significativement plus de croisements dans l'Étude 2 que dans l'Étude 1, tandis que les agents Réactif et Indépendant ne diffèrent pas significativement entre études. Cette interaction suggère que l'instruction sociale explicite modifie les stratégies d'exploration des participants d'une manière qui va tendre à augmenter la rencontre avec les agents oscillatoires (dont les trajectoires sont plus prévisibles et localisées autour du centre), mais n'augmente pas substantiellement les rencontres avec l'agent Réactif (qui génère déjà un nombre très élevé de croisements du fait de sa propension naturelle à « chercher » à maintenir le contact) ni avec l'agent Indépendant (dont les trajectoires imprévisibles et distribuées sur tout l'environnement rendent les rencontres difficiles, indépendamment de la stratégie du participant).

Une constatation intrigante est celle du contraste entre les mesures subjectives (questionnaires) et comportementales (croisements) : le type d'agent mo-

dule fortement le nombre de croisements mais module peu les scores de *Présence Sociale*, tandis que l'instruction module fortement les scores de *Présence Sociale* mais marginalement le nombre de croisements. Elle suggère que la *Présence Sociale*, telle que mesurée par questionnaire pourrait ne pas se réduire, dans un tel environnement, au simple effectif par minute des contacts, mais dépendrait plutôt du cadrage cognitif et interprétatif de l'expérience. Les croisements reflètent en partie la coordination sensorimotrice et sont sensibles aux propriétés dynamiques fines des agents, tandis que la *Présence Sociale* subjective est davantage influencée par les attentes sociales préalables et l'interprétation rétrospective de l'expérience.

6.3.3 Analyses fractales et multifractales

Les analyses fractales et multifractales des séries temporelles de vélocité apportent un éclairage complémentaire en révélant des signatures temporelles multi-échelles distinctes selon le type d'agent, indépendantes de l'instruction expérimentale. Ces résultats étendent et apportent de la nuance aux travaux de Bedia et al. (2014) qui avaient identifié une signature fractale $1/f$ ($\beta \approx 1$) caractéristique des interactions humain-humain dans le *Perceptual Crossing*, absente des interactions avec des agents oscillatoires ou leur agent « *shadow* ».

Ces résultats se manifestent principalement dans la vélocité relative (dérivée de la distance entre participant et agent), qui capture directement les dynamiques de coordination émergente. Avec l'agent Réactif, la structure fractale présentait systématiquement les exposants d'échelle α et les pentes spectrales β les plus faibles parmi tous les agents, tant dans l'Étude 1 que dans l'Étude 2. Ses valeurs proches de 0.5 pour l'exposant α et de 0 pour β indiquent une dynamique quasi anti-persistante, proche d'un bruit blanc, caractérisée par une absence de corrélations temporelles à long terme et une organisation multi-échelles très faible. Notre agent Réactif, bien que contingent aux croisements avec le participant et générant un nombre élevé de croisements, ne parvient donc pas à reproduire ni à susciter la qualité de coordination temporelle observée dans les interactions humain-humain. De plus, la largeur du spectre multifractal Δh de la vélocité relative avec l'agent Réactif était significativement inférieure à celle avec l'agent Indépendant qui était son homologue « libre » mais ne différait pas significativement de celle avec l'agent Périodique, et était supérieure à

celle avec l'agent Fluctuant. Un Δh intermédiaire suggère une diversité modérée des dynamiques locales : les interactions avec l'agent Réactif comportent des phases hétérogènes (rafales d'activité intense lors des croisements suivies de périodes calmes), mais cette hétérogénéité reste limitée comparativement à celle avec l'agent Indépendant dont le comportement aléatoire génère une plus grande variabilité d'échelles.

Les séries temporelles de vélocité relative avec les agents Périodique et Indépendant présentaient des exposants α et β significativement plus élevés, proches de ceux attendus pour des dynamiques de type bruit rose ($1/f$), indicatrices de corrélations à long terme et d'une organisation multi-échelles. Avec l'agent Périodique, la vélocité relative affichait des valeurs α autour de 0.90 dans les deux études correspondant à des pentes spectrales β autour de $0.81 et 0.79$. Pour celle avec l'agent Indépendant, elle présentait des valeurs similaires, bien que légèrement supérieures dans l'Étude 2. Ces valeurs élevées d'exposant α (proches de 1) et de β (proches de 1 mais légèrement inférieures) suggèrent que les interactions avec ces agents, bien que moins fréquentes en termes de croisements, génèrent néanmoins des schémas de vélocité relative présentant des corrélations à long terme : les fluctuations passées influencent les fluctuations futures sur plusieurs échelles temporelles, traduisant une forme de coordination implicite ou de synchronisation partielle.

Cette observation apparemment paradoxale avec des corrélations temporelles fortes avec des agents non-contingents, peut s'expliquer par le fait que ces agents présentent des comportements intrinsèquement structurés. L'agent Périodique, avec son mouvement sinusoïdal régulier, impose une structure oscillatoire stable qui se reflète dans les dynamiques de vélocité relative. Même sans ajustement mutuel, la rencontre répétée entre le participant et cette trajectoire périodique génère des schémas temporels prévisibles et corrélés. De même, l'agent Indépendant, bien qu'aléatoire et « libre », explore l'environnement de façon non contrainte spatialement, ce qui peut induire une variabilité d'échelles riches dans les distances et donc dans la vélocité relative lorsqu'il croise occasionnellement le participant.

La largeur du spectre multifractal Δh de la vélocité relative avec l'agent Indépendant était la plus élevée comparée à celle d'avec les autres agents, confirmant une organisation multi-échelles particulièrement riche et une hétérogé-

néité temporelle marquée avec celui-ci. Cette multifractalité élevée indique une alternance entre des phases de dynamique rapide (lorsque l'agent est proche et les distances varient rapidement) et des phases plus lentes (lorsque l'agent est distant et les variations de distance sont minimales).

Pour ce qui est de la dyade avec l'agent Fluctuant, celle-ci présentait des exposants intermédiaires, suggérant une organisation temporelle à la frontière entre une dynamique faiblement corrélée et un bruit rose ($1/f$). Ces valeurs, bien qu'inférieures à celles avec les agents Périodique et Indépendant, restent nettement supérieures à celles de l'agent Réactif, indiquant que les interactions avec l'agent Fluctuant conservent une certaine structure multi-échelles et des corrélations à moyen terme. La structure dans l'interaction avec l'agent Fluctuant, qui est conçu pour explorer une partie de l'environnement avec une amplitude d'oscillation définie à chaque oscillation complète mais spatialement contrainte autour du centre, présente une variabilité temporelle modérée qui se traduit par une signature fractale intermédiaire. La largeur de spectre multifractal Δh de la vélocité relative était la plus faible avec celui-ci parmi celles d'avec les autres agents, suggérant une dynamique temporelle plus homogène et une moindre diversité des échelles locales. En revanche, cette largeur de spectre multifractal au niveau de la vélocité relative avec cet agent étant élevée (homogénéité), pourrait refléter le fait que l'agent Fluctuant, bien que variable, dans la dynamique conjointe mesurée, maintient une distribution relativement stable de ses vitesses sans les phases extrêmes d'accélération et de décélération observées avec d'autres agents.

Aucun effet principal de l'étude (instruction neutre ou instruction sociale explicite) n'a été observé sur les exposants α, β ou Δh de la vélocité relative. En revanche, un effet de l'étude était présent pour l'analyse entre le bloc 1 de l'Étude 2 et l'Étude 1. L'interaction entre l'étude et le type d'agent était significative uniquement pour l'exposant α et β, mais cette interaction ne modifiait pas le motif global des différences selon les agents rencontrés. Ce résultat indique que, bien que l'instruction avec incitation sociale augmente la *Présence Sociale* perçue et module légèrement le nombre de croisements, elle n'affecte pas fondamentalement l'organisation multi-échelles des dynamiques d'interaction. Autrement dit, les structures temporelles fractales émergent principalement des propriétés intrinsèques de l'agent et de la coordination sensorimotrice impli-

cite, indépendamment du cadrage cognitif ou des attentes sociales explicites. Ce contraste que l'on observe entre les effets subjectifs/comportementaux (présence sociale et nombre de croisements) et les structures temporelles suggère ici que les participants peuvent présenter différentes stratégies d'exploration et évaluations subjectives en fonction de l'instruction, mais que l'organisation temporelle de la coordination émerge de processus sensorimoteurs plus implicites. Autrement dit, l'instruction module les stratégies des participants, pas la dynamique d'interaction.

L'analyse des vélocités individuelles (agent et participant) fournit des éclairages complémentaires sur les mécanismes sous-jacents aux schémas observés dans la vélocité relative. Les analyses de la vélocité de l'agent ont confirmé que chaque type d'agent présente une signature distincte, directement déterminée par sa programmation comportementale.

L'agent Réactif, programmé pour explorer et ajuster continuellement son amplitude d'oscillation en fonction des croisements avec le participant présentait des exposants α extrêmement faibles et négatifs, compatibles avec une dynamique quasi anti-persistante ou de type bruit blanc. Cette signature reflète les changements de direction et de vitesse fréquents et brusques induits par le suivi réactif du participant. L'agent Périodique, avec son mouvement sinusoïdal préprogrammé, affichait les valeurs α et β les plus élevées (correspondant à un processus proche d'un mouvement brownien), cohérentes avec une dynamique hautement corrélée et régulière imposée par l'oscillation stricte. L'agent Fluctuant présentait des valeurs intermédiaires, s'apparentant à un bruit rose ($1/f$), indiquant des fluctuations de vélocité corrélées à long terme mais avec une variabilité modérée. Enfin, l'agent Indépendant conservait également un profil corrélé, proche du bruit brownien, mais avec une largeur de spectre plus restreinte que celle de l'agent Périodique, suggérant une variabilité interne plus stable malgré son mouvement aléatoire. Ces résultats confirment que les dynamiques intrinsèques des agents sont fidèlement capturées par les analyses fractales, et que ces propriétés se propagent ensuite dans la vélocité relative en fonction de la façon dont le participant interagit avec chaque agent.

En contraste marqué avec les motifs observés dans la vélocité relative et celle de l'agent, l'analyse de la vélocité du participant seul n'a révélé aucun effet significatif du type d'agent, de l'étude, ni d'interaction entre ces facteurs pour les

exposants α, β ou Δh. La variance interindividuelle était élevée, indiquant une forte hétérogénéité entre participants dans leurs stratégies d'exploration, mais ces stratégies individuelles restaient stables et indépendantes des propriétés de l'agent rencontré ou de l'instruction reçue. Ce résultat démontre que les différences observées dans la vélocité relative ne proviennent pas d'ajustements actifs et différenciés du participant en fonction du type d'agent, mais émergent principalement de la façon dont les dynamiques intrinsèques de l'agent et celles du participant se couplent au niveau dyadique.

Autrement dit, les participants maintiennent une stratégie d'exploration personnelle relativement constante, et c'est la rencontre de cette stratégie avec les différentes dynamiques des agents qui génère les motifs fractals distinctifs observés dans la vélocité relative. Cette absence d'adaptation comportementale individuelle pourrait également expliquer en partie l'absence de différenciation entre agents : si les participants n'ajustent pas consciemment leur comportement en fonction de l'agent, ils peuvent ne pas percevoir les propriétés distinctives de celui-ci comme suffisamment saillantes pour moduler leur sentiment de *Présence Sociale*.

6.3.4 Limites

Plusieurs limitations de ces études méritent d'être soulignées. Premièrement, bien que notre protocole en ligne ait permis de recruter un large échantillon de participants, la nature à distance de l'expérimentation introduit une variabilité technique (environnement de la passation, performances des ordinateurs personnels ou différences matérielles comme la souris ou la taille de l'écran) difficile à contrôler rigoureusement. Nous avons tenté de contrôler certains écarts de performances avec un rééchantillonnage uniforme des séries temporelles à 16 ms et par l'exclusion des enregistrements aberrants, mais une certaine variabilité résiduelle persistait probablement.

Deuxièmement, le fait d'utiliser des agents programmés dans un environnement virtuel unidimensionnel reposant sur un feedback audiovisuel, bien que conforme à l'esprit minimaliste du paradigme du *Perceptual Crossing*, limite considérablement la richesse des interactions possibles et les ressentis comparativement à un feedback tactile ou à une interaction avec un autre humain.

Troisièmement, le questionnaire de *Présence Sociale* était administré après

l'ensemble du bloc d'interaction, lui-même composé de 2 essais, introduisant potentiellement un biais et un effet de moyennage qui pourrait avoir atténué la perception. Des mesures de *Présence Sociale* plus granulaires, recueillies immédiatement après chaque essai ou en temps réel pourraient révéler des variations plus fines.

Quatrièmement, bien que les agents présentaient des signatures dynamiques contrastées, aucun d'entre eux ne parvenait à susciter et reproduire fidèlement la « qualité » de coordination multi-échelles de type bruit *1/f* observée dans les interactions humain-humain par Bedia et al. (2014). Le développement d'agents plus sophistiqués, potentiellement avec de l'apprentissage automatique ou intégrant des mécanismes d'anticipation encore plus avancés, de mémoire temporelle réelle ou de synchronisation plus adaptative, pourrait permettre de se rapprocher de cette signature fractale distinctive et de potentiellement générer des expériences de *Présence Sociale* plus robustes et différenciées.

Cinquièmement, nos analyses se sont concentrées sur trois mesures principales : *Présence Sociale* subjective, nombre de croisements et signatures fractales. Le contraste observé entre mesures subjectives, comportementales et temporelles soulève des questions méthodologiques et théoriques importantes. Il est possible que les questionnaires de *Présence Sociale* auto-rapportés, bien que largement utilisés dans la littérature (Harms & Biocca, 2004), ne capturent qu'une facette limitée et réflexive de l'expérience sociale, manquant des dimensions plus implicites ou pré-réflexives de la perception d'autrui.

Enfin, nos résultats soulèvent une question théorique fondamentale concernant les mécanismes de la *Présence Sociale* dans les interactions minimales. Si le nombre de contacts (croisements) et les motifs de coordination temporelle (signatures fractales) ne se traduisent pas directement en différences subjectives de *Présence Sociale*, quels sont les facteurs critiques qui déterminent ce sentiment ? Nos données suggèrent que, dans un environnement aussi dépouillé d'indices sociaux, l'instruction explicite, qui oriente l'attention et l'interprétation, joue un rôle prépondérant, potentiellement plus important que les propriétés dynamiques fines de l'agent. Cela soulève l'hypothèse que la *Présence Sociale* pourrait résulter moins d'une détection sensorimotrice directe de la contingence (comme le suggèrent certaines approches énactives radicales) que d'un processus inférentiel de haut niveau, modulé par les attentes, le contexte et le

cadrage cognitif.

Alternativement, il est possible que la *Présence Sociale* nécessite un seuil minimal de contingence (que tous nos agents, y compris l'agent Réactif, ne parviennent pas à atteindre de façon équivalente à un humain), au-delà duquel les variations quantitatives de contingence ne modulent plus substantiellement l'expérience subjective. Des études futures reproduisant notre protocole entre humains ou manipulant plus systématiquement le type et le degré de contingence de l'agent programmé (par exemple, via des retards dans la réactivité, ou des probabilités graduées de réponse aux actions du participant) pourraient permettre de tester et mettre à l'épreuve directement cette hypothèse.

En conclusion, les travaux décrits dans cette Partie II ont permis d'examiner expérimentalement comment les propriétés des agents et les instructions données aux individus influencent la *Présence Sociale* et l'interaction. Le Chapitre 5 rapportait en détail notre adaptation méthodologique de l'expérience de *Perceptual Crossing* d'Auvray et al. (2009), ainsi que les protocoles de nos études expérimentales visant à étudier l'impact d'une instruction avec incitation sociale et des comportements des agents sur la *Présence Sociale* et les dynamiques d'interaction. Ensuite, le Chapitre 6 nous a permis de rapporter les résultats de ces deux études et de les discuter.

Dans la partie suivante, il sera question de la manière dont les délais de réponse et le style de communication d'un robot affectent la façon dont on le perçoit.

Troisième partie

Attentes temporelles dans les discussions humain-robot

Attentes temporelles dans les discussions humain-robot

La Partie II a exploré l'interaction sociale dans un environnement minimaliste, notamment les dynamiques temporelles et le sentiment d'être avec un autre. Nous nous tournons maintenant vers l'exploration d'un autre type d'échange dans l'interaction : les échanges verbaux avec les robots et plus précisément, leur temporalité. Comment les humains perçoivent-ils et réagissent-ils aux délais de réponse d'un robot dans une conversation ? Dans quelle mesure le style de communication d'un robot influence-t-il ces attentes temporelles ? Et surtout, opèrent-ils de la même manière que dans les interactions entre humains ?

Cette partie présente deux études complémentaires qui examinent ces questions. La première étude adopte une approche psychophysique pour identifier le délai de réponse optimal auquel un robot social doit nous répondre. Elle explore aussi comment différents styles de communication modulent ce délai et la perception du robot. La seconde étude approfondit ces questions en examinant comment des délais fortement déviants de cet optimal, combinés aux styles de communication, influencent l'évaluation globale du robot. Ces deux études ont donné lieu à un article accepté sous réserve de modifications et est actuellement en cours de révision pour publication.

Chapitre 7

Style de communication et perception des délais

7.1 Introduction

Comme nous l'avons vu au Chapitre 4, la coordination temporelle dans la conversation humaine repose sur une synchronisation remarquablement précise. Les humains alternent leurs tours de parole avec des transitions moyennes d'environ 200 ms (Stivers et al., 2009). Cette prouesse est rendue possible par des mécanismes prédictifs sophistiqués qui anticipent la fin du tour de l'interlocuteur (Ekstedt & Skantze, 2020 ; Garrod & Pickering, 2015 ; Levinson & Torreira, 2015). Il est important de noter que dans les conversations entre humains, le délai porte une signification et affecte la perception de l'échange mais aussi celle de l'interlocuteur : des délais de réponse plus rapides sont associés à une connexion sociale accrue (Templeton et al., 2022), tandis que des délais prolongés peuvent affecter négativement les jugements sur la compétence perçue du locuteur (Matzinger et al., 2023).

La question suivante se pose quant à ces normes temporelles finement calibrées de la conversation humaine : les humains appliquent-ils les mêmes attentes temporelles aux robots sociaux ? En effet, les systèmes de dialogue actuels peinent à reproduire la fluidité naturelle des échanges humains. Ils sont limités par des contraintes techniques (Skantze & Irfan, 2025) et s'appuient souvent sur des seuils de silence fixes (par exemple, 600 ms ; Cuijpers et Van Den Goor, 2017) qui créent des pauses artificielles dans l'interaction (Skantze, 2021). D'autre part, la coordination du tour de parole devient d'autant plus complexe lorsque le robot doit synchroniser paroles et actions (Hough & Schlangen, 2016).

Or, plusieurs résultats suggèrent que, face aux robots, les préférences tem-

porelles ne semblent pas recouvrir exactement celles qui sont appliquées avec un autre humain : Shiwa et al. (2009) montrent par exemple, au sujet de la vitesse à laquelle un robot doit nous répondre (sans faire varier de cadrage social comme par exemple, le statut du robot, son registre de politesse ou son style de communication), que les utilisateurs semblent préférer un léger délai de réponse d'environ 1 s, alors qu'ils préfèrent une réponse instantanée (0 s) avec une interface graphique.

Cette divergence suggère que les utilisateurs pourraient avoir développé des modèles temporels spécifiques pour l'interaction humain-robot, distincts de ceux appliqués avec les humains. Cette hypothèse d'attentes temporelles différenciées trouve un écho dans les recherches sur l'adaptation temporelle dans d'autres contextes d'interaction asymétrique. Casillas et al. (2016) ont montré que dans les interactions adulte-enfant, les adultes affichent des latences de réponse médianes d'environ 371 ms lorsque les enfants posent des questions. Ces latences sont nettement plus longues que dans les conversations entre adultes. Tandis que les enfants affichent des latences médianes d'environ 625 ms lorsque les adultes posent des questions. Cette adaptation suggère que les humains peuvent ajuster leurs attentes temporelles en fonction de la catégorie d'interlocuteur et des caractéristiques qu'ils lui attribuent. Dans le contexte de l'interaction humain-robot, cette capacité d'adaptation pourrait être mobilisée de façon comparable, les utilisateurs développant des attentes temporelles propres aux systèmes robotiques basées sur la perception de leurs capacités et limites.

Au-delà des considérations techniques, un aspect étant encore peu exploré concerne l'influence du style de communication du robot sur les attentes temporelles. En effet, les humains projettent spontanément des traits de personnalité sur les agents robotiques (Fong et al., 2003), mais l'impact de ces attributions sur la perception temporelle reste mal compris. Appliquée à l'interaction humain-robot, cette modulation des attentes pourrait opérer selon deux mécanismes distincts. Premièrement, le style de communication d'un robot pourrait modifier le délai optimal lui-même : un robot adoptant un style autoritaire pourrait être perçu comme devant répondre plus rapidement qu'un robot au style enfantin, reflétant des normes sociales associées à ces positions relationnelles. Deuxièmement, le style pourrait affecter la tolérance aux écarts par rapport au délai optimal : indépendamment de celui-ci, les utilisateurs pourraient se mon-

trer plus indulgents envers les variations temporelles d'un robot présentant certaines caractéristiques (par exemple, un style soumis ou enfantin) plutôt que d'autres. Ce chapitre présente une première étude adoptant une approche psychophysique pour identifier le délai de réponse optimal d'un robot social dans une tâche de questions-réponses fermées et cherche à examiner comment différents styles de communication modulent cette perception temporelle.

7.2 Méthode

Cette première étude visait donc un double objectif : d'une part (1) identifier le délai de réponse perçu comme optimal pour un robot social à des questions-réponses fermées (oui/non) et d'autre part (2) déterminer si, et de quelle manière, le style de communication du robot module cette perception temporelle.

7.2.1 Conception

Ainsi, l'Étude 1 a examiné, au moyen d'un plan mixte, l'influence du style de communication d'un robot sur la perception de son délai de réponse. L'étude comportait un facteur inter-sujets comprenant cinq conditions expérimentales relatives au style de communication (Autoritaire, Soumis, Neutre, Enfantin et Rideau) et un facteur intra-sujets (délai de réponse du robot à travers quinze délais différents : 100 à 1500 ms). Chaque participant était assigné à l'une des cinq conditions correspondant au style de communication du robot et exposé aux quinze délais de réponse. L'expérience comprenait trois phases : une phase d'introduction ou de familiarisation, une phase expérimentale principale, puis une phase de questionnaire. Au cours de la phase principale, les participants visionnaient cent-cinquante vidéos d'interactions verbales entre un robot et un humain et devaient juger si les différents délais de réponse du robot étaient « trop rapides » ou « trop lents ». Les variables dépendantes comprenaient ces jugements temporels (trop rapide ou trop lent) ainsi que les scores sur les dimensions du questionnaire de Ho et MacDorman (2017), qui évalue la perception d'*Humanité*, d'*Attrait* et d'*Étrangeté* relative au robot.

7.2.2 Participants

Deux cent dix participants ($N = 210$) ont été recrutés via la plateforme *Prolific* selon des critères stricts : être âgé entre 18 et 65 ans, avoir l'anglais comme première langue et langue maternelle, disposer d'un ordinateur avec une résolution minimale de 1024 px ainsi que d'un système audio fonctionnel et ne présenter aucune déficience auditive.

Les participants ont donné leur consentement éclairé avant de prendre part à l'étude, laquelle a été approuvée par le Comité Éthique de l'Université de Lille (réf. : 2021-467-S90). Concernant le genre, l'échantillon comprenait 109 hommes (51.9%), 100 femmes (47.6%) et une personne (0.5%) qui a préféré ne pas le préciser. L'âge des participants variait de 19 à 65 ans (Mdn $= 36.5$, $M = 37.9$, SD $= 11.5$). La majorité des participants résidaient au Royaume-Uni (51.9%), suivis des États-Unis (16.2%), du Canada (8.1%), de l'Afrique du Sud (6.7%) et du Nigeria (6.7%).

7.2.3 Matériel et Apparatus

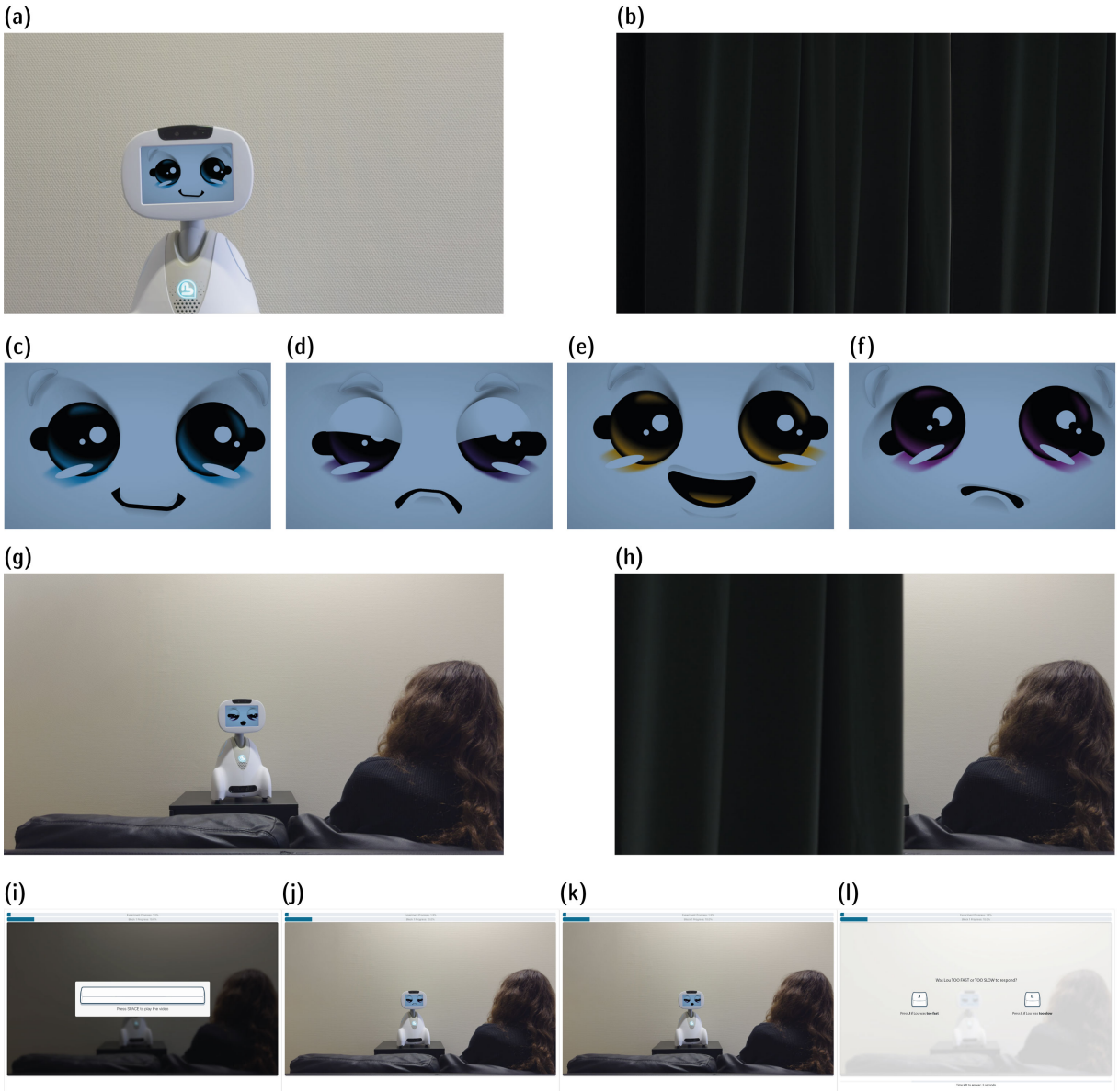
L'expérience et son interface ont été développées en *Javascript*, notamment via la bibliothèque *React*, et étaient accessibles aux participants via le navigateur web de leur ordinateur (voir Figure 7.1 pour un aperçu de l'interface). Un contrôle automatisé empêchait l'accès depuis un smartphone ou une tablette afin de garantir des conditions de passation homogènes.

Stimuli et amorçage du style de communication

Le robot utilisé était *Buddy* (*Blue Frog Robotics*). Cinq conditions expérimentales ont été créées : une condition de contrôle, dans laquelle le robot était caché derrière un rideau, et quatre conditions où le robot affichait sur l'écran placé au centre de son visage une expression correspondant à un style de communication spécifique (voir Figure 7.1). Les stimuli auditifs ont tous été générés à l'aide de l'outil de synthèse vocale d'*ElevenLabs* (https://elevenlabs.io), en s'appuyant sur de brèves indications insérées dans les prompts (par exemple, des points de suspension « … », point d'exclamation « ! », des phrases telles que *« — dit d'une voix autoritaire, froide, directe et sobre »* (*« — said in an authoritative, cold, direct, and sober voice »*) pour le style Autoritaire, ou *« — dit d'une voix hésitante et timide »* (*« — said in a hesitant and shy voice »*) pour le style Soumis).

Figure 7.1

Aperçu des interfaces, vidéos et visages du robot



Note. Captures d'écran de l'expérience montrant différentes phases et conditions. (**a-b**) Phase d'introduction pour les conditions Neutre et Rideau. (**c-f**) Expressions faciales du robot (de gauche à droite) : Neutre affichant son expression par défaut, Autoritaire affichant une expression sévère, Enfantin affichant une expression joyeuse, et Soumis affichant une expression triste. (**g-h**) Aperçus de la phase principale pour les conditions Autoritaire et Rideau, où l'humain (vu de dos) pose des questions au robot. (**i-l**) Interface vidéo pendant la phase principale (de gauche à droite) : écran invitant à appuyer sur la touche ESPACE pour démarrer la vidéo, aperçu vidéo avant et après la réponse oui/non du robot, écran invitant à juger le délai (appuyer sur J ou L).

Toutes les conditions utilisaient le même timbre de voix. Les cinq conditions étaient les suivantes :

— **Neutre** : dans cette condition, la voix ne comportait aucun élément supplémentaire au-delà du rendu par défaut généré par *ElevenLabs* tandis que *Buddy* affichait son visage neutre de base.

— **Autoritaire** : dans cette condition, l'utilisation d'un ton directif suggère des ordres venant d'un supérieur, le discours va droit au but et le visage de *Buddy* est grincheux.

— **Enfantin** : dans cette condition, la voix est modulée pour avoir un ton enthousiaste tandis que *Buddy* utilise des tournures de phrases enfantines, accompagnés d'un visage joyeux.

— **Soumis** : dans cette condition, la voix est modulée pour avoir un discours hésitant, avec des micro-pauses et des formulations très prudentes. Le robot affichait un visage pouvant être perçu comme triste ou timide.

— **Rideau** : cette condition fait office de contrôle. Elle est identique à la condition Neutre mais ici le robot caché est derrière un rideau.

Ces styles de communication ont été sélectionnés afin de représenter des dynamiques relationnelles distinctes, susceptibles d'influencer les attentes temporelles dans l'interaction humain-robot. Le style Autoritaire représente une position hiérarchique dominante, pouvant susciter des attentes en matière d'efficacité et de précision temporelle. À l'inverse, le style Soumis évoque une position subordonnée, avec laquelle les humains adultes pourraient être davantage enclins à tolérer des délais de réponse plus longs. Le style Enfantin correspond à un autre registre d'attentes sociales : les adultes font généralement preuve d'une plus grande tolérance face aux délais et à la temporalité lorsqu'ils interagissent avec des enfants. Le style Neutre servait de condition de référence pour les comparaisons, permettant d'isoler l'effet des variations liées au style. Enfin, la condition rideau constituait une condition contrôle supplémentaire, venant compléter la référence Neutre, permettant de distinguer les effets du style verbal (la voix seule) de ceux des indices visuels (apparence des expressions du visage) dans la perception temporelle. Le style de communication assigné au robot était maintenu identique pour chaque participant tout au long de l'expérience. Cette cohérence a été assurée à travers les différentes phases clés

(voir Annexe A.1.1 pour les scripts complets) Les stimuli de la phase principale comprenaient cent-cinquante vidéos (en anglais) montrant une scène d'interaction entre un humain et le robot. Dans toutes les conditions, la structure des vidéos étaient strictement identiques : un humain vu de dos posait une question en anglais au robot (voir Figure 7.1), lequel répondait, en anglais, par un « Oui » ou un « Non » identique dans chaque condition. Un exemple de question est, par exemple, (voir Annexe A.2 pour la liste complète) : (traduction) « *Est-ce que tu peux t'ennuyer ?* » (« *Can you get bored ?* ») ou « *Est-ce que t'es déjà allé à un concert ?* » (« *Have you ever been to a concert ?* »).

Chaque vidéo était de courte durée, avec une moyenne d'environ 6 à 8 s, incluant la question posée par l'humain, le délai de réponse variable du robot, puis la réponse du robot. Les cent-cinquante vidéos étaient réparties en cinq blocs, séparés par de courtes pauses.

Le style de communication était transmis (amorcé) soit par la combinaison de la voix et du visage affiché sur l'écran, soit par le visage seul, selon la phase :

— Lors des phases d'introduction et de familiarisation, le robot présentait les consignes dans le style de communication qui lui avait été attribué (voix et visage).

— Tout au long de la phase principale, durant laquelle le robot répondait aux questions uniquement avec les mêmes enregistrements « Oui » ou « Non », seule l'expression faciale servait d'amorçage du style (visage uniquement).

— Lors des quatre pauses séparant les cinq blocs expérimentaux, le robot s'exprimait de nouveau dans son style caractéristique (voix et visage).

Cette conception garantissait que les différences dans la perception des délais de réponse ne pouvaient être attribuées à des variations dans les audios des réponses oui/non, telles que des différences de hauteur ou du délai d'établissement du voisement (*« voice onset time »*, qui correspond au temps qui sépare la libération d'une consonne du début des vibrations des cordes vocales). Par ailleurs, dans la condition Rideau, le robot étant caché, les expressions faciales affichées à l'écran n'étaient pas visibles pour le participant.

Variations du délai

Dans les vidéos de la phase principale, le robot répondait après un délai variable parmi quinze niveaux : *100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 720, 865, 1040, 1250* et *1500 ms*. Une granularité plus fine autour de la plage 100-450 ms permettait d'analyser précisément les seuils de perception proches de ceux rapportés dans la littérature sur la prise de parole entre humains, tandis que les intervalles plus larges au-delà de 500 ms capturent les effets des délais extrêmes et ceux de certains systèmes robotiques (tels que 700-1000 ms mentionnés par Skantze, 2021).

Pré-tests : sélection de la voix et validation des styles

Dans le but de minimiser les biais potentiels liés au genre dans la perception de différentes conditions vocales, le timbre de voix du robot était celui d'une voix androgyne. Afin de sélectionner cette voix, une étude préliminaire a été menée auprès d'un panel de onze participants (sept hommes et quatre femmes) qui ont évalué plusieurs échantillons de voix sur un continuum allant de « Très masculin » à « Très féminin ». La voix sélectionnée, *River - gender neutral* d'*ElevenLabs*, a obtenu les évaluations les plus divergentes ($50/50$) parmi les évaluateurs, traduisant une perception ambiguë et variable de son genre. Cette ambiguïté renforce la neutralité requise pour notre protocole expérimental, garantissant que l'attention des participants se concentre sur les caractéristiques propres à chaque condition plutôt que sur le genre perçu de la voix.

De même, afin de valider la manipulation du style de communication, un pré-test complémentaire a été mené auprès de vingt-sept personnes (12 hommes, 13 femmes et deux non-binaires) selon un plan intra-sujet. Dans ce pré-test, les participants ont été exposés, dans un ordre aléatoire, aux quatre styles de communication. Pour chaque style, ils devaient répondre à une question ouverte sur ce que leur évoquait le robot, puis répondre à d'autres questions et questionnaires, et enfin, choisir le nom du style qui le représentait le mieux. Les résultats ont confirmé la validité de la manipulation de style, avec des taux de catégorisation corrects nettement supérieurs au hasard ($\chi^2 = 9.24$, $p = .026$). (voir l'Annexe A.4 pour plus de détails).

7.2.4 Procédure

L'expérience se déroulait en trois phases, avec des contrôles d'attention intégrés tout au long. Ces contrôles comprenaient : un test audio initial (où les participants devaient retranscrire correctement un mot entendu), un item de vérification de l'attention où le robot indique au participant d'appuyer sur une touche précise du clavier et un autre supplémentaire avant le bloc 3, demandant aux participants de saisir un mot.

Phases d'introduction

Lors de la phase d'introduction, les participants visionnaient une vidéo dans laquelle le robot (dénommé *Lou* dans l'expérience) présentait les instructions selon le style de communication qui lui avait été attribué. Cette vidéo était suivie de trois essais de familiarisation, comprenant un item de vérification de l'attention (appuyer sur la touche demandée par le robot).

Phase principale

Lors de la phase principale, les participants visionnaient cent-cinquante vidéos, réparties en cinq blocs de trente vidéos, avec des pauses entre chacun des blocs. Chaque vidéo était déclenchée par le participant en appuyant sur la touche *ESPACE* et présentait une scène d'interaction dans laquelle un humain posait une question, suivie de la réponse du robot (Oui/Non).

Après chaque vidéo, une fenêtre de réponse apparaissait avec l'instruction (traduction) : « *Est-ce que Lou était TROP RAPIDE ou TROP LENT pour répondre ?* » (« *Was Lou TOO FAST or TOO SLOW to respond ?* »). Les participants devaient alors évaluer le délai de réponse du robot par rapport à la question de l'humain en appuyant sur **J** s'ils percevaient le délai de réponse du robot comme *TROP RAPIDE* ou sur **L** s'il était *TROP LENT* (voir Figure 7.1). Cette méthode psychophysique classique permet d'identifier le *Point d'Égalité Subjective* (*PSE*), c'est-à-dire le délai pour lequel le participant était le plus indécis, autrement dit celui qui était jugé la moitié du temps comme trop long et l'autre moitié comme trop court.

L'ordre des cinq blocs et des vidéos au sein de chaque bloc a été randomisé pour chaque participant à l'aide d'un fichier *JSON* unique. Cette procédure ga-

rantissait que chaque bloc contenait exactement deux instances de chaque niveau de latence et qu'aucun participant n'était exposé au même ordre de présentation. Lorsque la fenêtre d'évaluation du délai apparaissait, les participants ne pouvaient pas répondre pendant les premières 500 ms afin d'éviter les réponses précipitées et mécaniques ; les participants disposaient ensuite de 10 s pour soumettre leur évaluation (par pression de la touche). Pendant les pauses entre les blocs, une vidéo (d'une durée moyenne de 20 s) était présentée, dans laquelle le robot s'exprimait selon le style de communication qui lui avait été attribué afin de maintenir l'effet d'amorçage (voir l'Annexe A.1.1 pour le détail du contenu du discours). Enfin, les participants pouvaient se reposer pendant 45 s avant le bloc d'essais suivant.

Phase de questionnaire

Après les cinq blocs, les participants complétaient le questionnaire de Ho et MacDorman (2017), destiné à évaluer leur perception du robot. Ce questionnaire se compose de plusieurs échelles sémantiques différentielles (paires d'adjectifs opposés) en 7 points allant de -3 à $+3$, sur lesquelles les participants notaient le robot. Chaque échelle était ancrée par des adjectifs opposés. Par exemple, « Mouvement mécanique–Mouvement biologique » ou « Inanimé–Vivant » (*Inanimate–Living*). Les valeurs numériques étaient masquées afin de minimiser les effets d'ancrage. Les dimensions évaluées avec ce questionnaire sont l'*Humanité* (*« Humanness »*), l'*Attrait* (« *Attractiveness* ») et l'*Étrangeté* (« *Eeriness* »).

7.2.5 Analyse des données

Les jugements temporels ont été analysés afin de déterminer le délai de réponse optimal ainsi que la sensibilité aux écarts par rapport à ce dernier. Pour obtenir le délai de réponse optimal à chaque condition, le *PSE* a été calculé pour chaque participant en ajustant une fonction sigmoïde incluant un paramètre de décalage vertical (« *vertical offset* ») aux proportions de réponses « trop lent » de chaque participant. La fonction sigmoïde utilisée est définie comme suit :

$$f(x) = \frac{1}{1 + e^{-\beta_1(x-\beta_2)}} + b \tag{7.1}$$

où le paramètre β_1 contrôle la pente de la courbe, β_2 correspond au *PSE* et b est un paramètre de décalage vertical modélisant d'éventuels biais de réponse dans les jugements. Les *PSE* individuels ont ensuite été comparés entre conditions à l'aide d'un test de Kruskal-Wallis.

Afin d'examiner plus finement les jugements temporels, la tolérance temporelle autour du *PSE* a été analysée. Pour ce faire, la probabilité de réponse « trop lent » a été modélisée à l'aide de *GLMM* avec une fonction de lien logit. Le modèle incluait des intercepts aléatoires pour les participants, ainsi que des effets fixes de la condition, du délai de réponse normalisé (min-max = [0,1]) et leur interaction :

$$logit(p_{ij}) = \underbrace{\beta_0 + u_i}_{\text{Intercepts}} + \underbrace{\beta_1 x_{ij} + \beta_2 c_i + \beta_3 (x_{ij} \times c_i)}_{\text{Effets Fixes}} \tag{7.2}$$

Ici, p_{ij} est la probabilité qu'un essai j du participant i soit jugé « trop lent », x_{ij} la latence normalisée (min-max [0, 1]), c_i la condition du participant, et u_i l'intercept aléatoire du participant i.

Concernant la perception du robot, les réponses aux trois dimensions (*Humanité, Étrangeté, Attrait*) du questionnaire de Ho et MacDorman (2017) ont été comparées entre les différentes conditions au moyen d'un test de Kruskal-Wallis. Dans le cas d'un résultat global significatif, des comparaisons post-hoc avec le test de comparaisons multiples de Dunn (avec correction de Bonferroni) étaient effectuées.

7.3 Résultats

7.3.1 Point d'Égalité Subjective

Un *PSE* global d'environ ≈703.7 ms (SD $= 204.3$ ms, $IC_{95\%}= [672.8, 734.6]$) a été identifié pour les quatre conditions où le robot était visible. La comparaison des *PSE* entre les quatre styles de communication au moyen d'un test de Kruskal-Wallis n'a révélé aucune différence significative ($H(3)= 0.10$, $p = .99$). Cela indique que le *PSE* était remarquablement stable à travers ces quatre styles de communication, suggérant que le délai optimal perçu est indépendant du style de communication du robot. Les moyennes des *PSE* sont présentées au Ta-

bleau 7.1, et les statistiques descriptives des paramètres de sigmoïde (β_1, β_2 et b) peuvent être trouvées dans le Tableau en Annexe A.3.

Table 7.1

Point d'Égalité Subjective moyen selon la condition

Condition	PSE moyen (ms)	SD	$IC_{95\%}$
Global[a]	703.7	204.3	[672.8, 734.6]
Global[b]	700.1	203.5	[672.6, 727.6]
Neutre	704.1	180.1	[649.7, 758.6]
Autoritaire	703.8	225.8	[635.5, 772.1]
Enfantin	704.7	203.9	[643.1, 766.4]
Soumis	702.1	212.2	[638.0, 766.3]
Rideau	685.6	202.2	[624.5, 746.8]

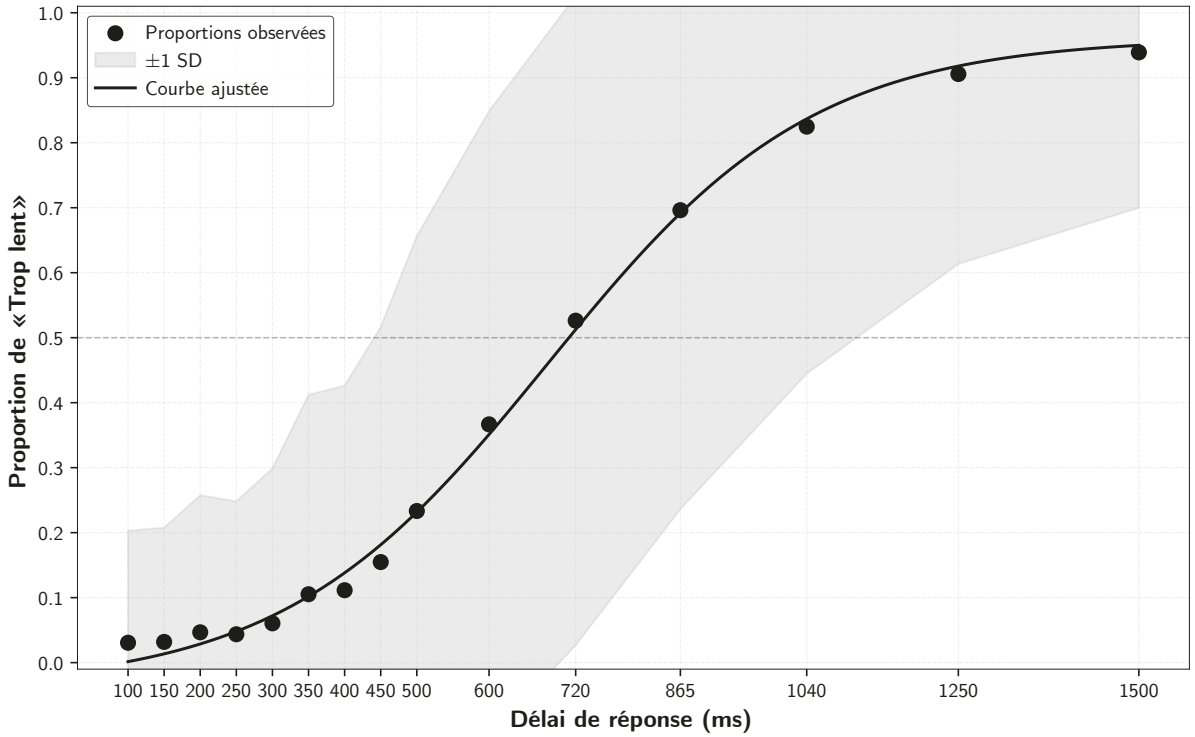
Note. Ces valeurs de PSE ont été estimées à partir de l'ajustement de fonctions sigmoïdes aux données individuelles, puis moyennées par condition. [a] Moyenne sur l'ensemble des participants, à l'exception de la condition *Rideau*. [b] Moyenne sur l'ensemble des participants, toutes conditions confondues.

L'inclusion de la condition Rideau ne modifiait pas substantiellement ce résultat, avec un *PSE* global de ≈700.1 ms et aucune différence significative entre les cinq conditions ($H(4) = 0.12$, $p = .99$) ou entre cette condition comparée aux autres ($H(1) = 0.04$, $p = .85$), suggérant que voir le robot, ou plus précisément ce robot (*Buddy*), n'impacte pas ce délai optimal perçu.

Pour l'ensemble des participants, les valeurs de *PSE* présentaient la dispersion suivante : Me $= 674.8$ ms, IQR $= 242.4$ ms, min-max$= [209.6, 1413.9]$ ms, SD $= 203.5$ ms. La Figure 7.2 présente l'évolution de la proportion de réponses « trop long » en fonction du délai de réponse avec la fonction psychométrique ajustée sur l'ensemble des données (tous participants et toutes conditions confondues).

Figure 7.2

Fonction psychométrique globale pour les jugements des délais de réponse

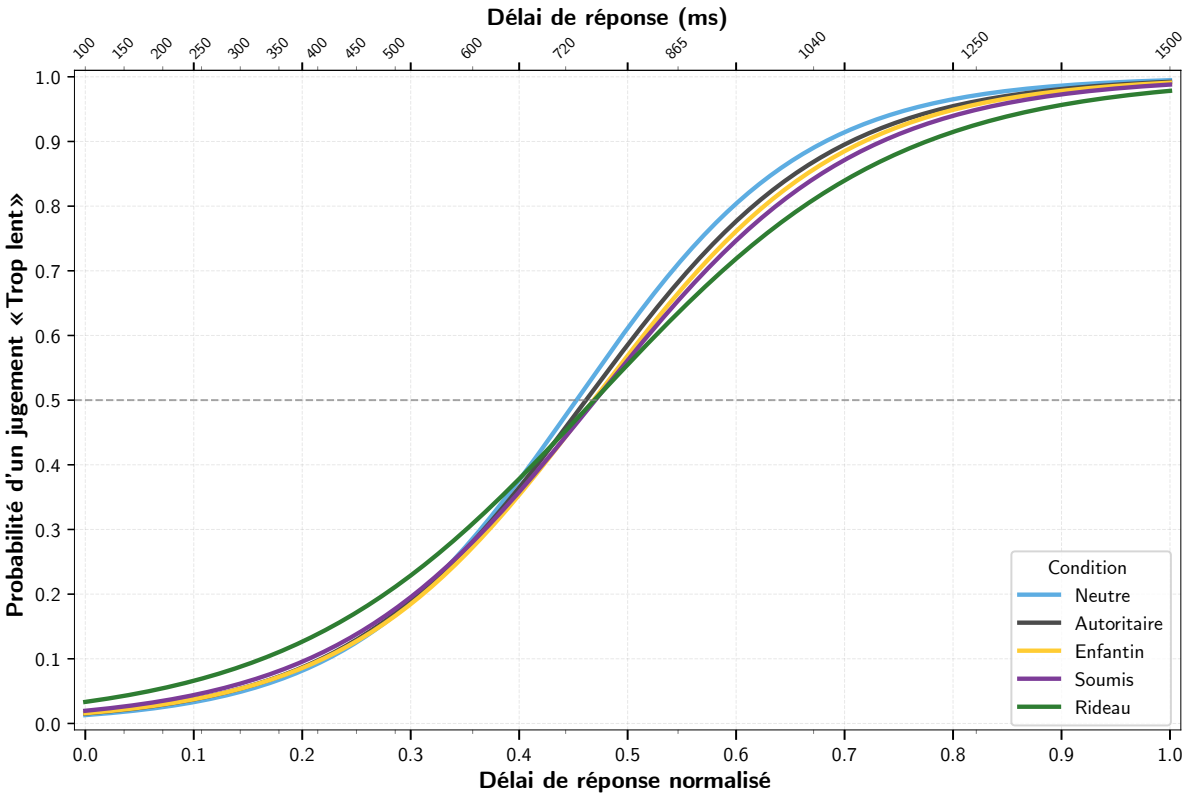


Note. La figure montre l'évolution de la proportion de réponse *trop lent* en fonction du délai de réponse (ms). Les points noirs représentent les proportions observées, moyennées sur l'ensemble des participants et des conditions (y compris *Rideau*). La ligne continue montre la fonction sigmoïde ajustée. La zone grisée représente ± 1 écart-type. La ligne horizontale en pointillés à 0.5 indique le seuil du *PSE*

7.3.2 Sensibilité aux écarts temporels

Figure 7.3

Courbes psychométriques issues du GLMM des cinq conditions



Note. Les lignes continues représentent les prédictions du modèle issues de l'analyse GLMM de la probabilité de jugements *trop lent* pour chacune des cinq conditions. Avec *Enfantin* (jaune), *Autoritaire* (gris), *Soumis* (violet), *Neutre* (bleu, référence), *Rideau* (vert). L'axe des abscisses présente les valeurs normalisées (en bas) et en millisecondes (en haut). La ligne horizontale en pointillés à 0.5 sur l'axe des ordonnées indique le seuil du Point d'Égalité Subjective (voir le Tableau 7.2 pour les valeurs de *PSE*).

Bien que le *PSE* ne diffère pas selon les conditions, les analyses complémentaires ont mis en évidence des différences dans la manière dont les participants évaluent les écarts autour de ce délai optimal, comme l'indiquent les variations de pente des fonctions psychométriques entre les conditions.

Tout d'abord, seules les quatre conditions principales où le robot était visible ont été examinées. Dans la condition de référence *Neutre*, un effet très significatif de la latence a été observé : la probabilité de répondre « trop lent »

Table 7.2

Points d'Égalité Subjective estimés à partir du GLMM selon la condition

Condition	PSE (ms)
Neutre	733.87
Autoritaire	746.29
Enfantin	755.95
Soumis	758.15
Rideau	757.04

Note. Les valeurs de PSE ont été calculées à partir des coefficients du GLMM au seuil de probabilité de 0.5 des jugements *trop lent*. Ces estimations reposent uniquement sur les effets fixes du modèle et ne permettent pas de calculer d'écart-type, contrairement aux PSE dérivés des ajustements sigmoïdes individuels.

augmente significativement avec la latence ($\beta_{\text{latence}} = 9.59$, $p < .001$). Comparativement à cette condition *Neutre*, l'interaction entre latence et style de communication révèle des fonctions psychométriques plus plates dans la condition *Soumis* ($\beta = -1.26$, $p < .001$) et *Enfantin* ($\beta = -0.76$, $p = .019$), ainsi qu'une tendance similaire mais non significative pour la condition *Autoritaire* ($\beta = -0.57$, $p = .085$).

Dans une seconde analyse, incluant la condition *Rideau*, la courbe s'est révélée encore plus plate pour cette dernière ($\beta = -2.40$, $p < .001$), tandis que les coefficients des autres conditions présentant des styles de communication restent pratiquement inchangés (voir en Annexe le Tableau A.2), ce qui atteste de la robustesse des effets. Ces pentes plus faibles (voir Figure 7.3) indiquent une zone d'incertitude plus large autour du *PSE*, où les participants étaient moins catégoriques dans leurs jugements temporels. Concrètement, ces résultats suggèrent une plus grande tolérance aux écarts par rapport au *PSE* lorsque le robot adopte un style *Soumis* ou *Enfantin*, ou lorsqu'il est caché derrière un *Rideau*. Dans ces conditions, les participants acceptaient une plus grande variété de délais comme étant ni trop rapides ni trop lents par rapport à la condition *Neutre*.

À partir des coefficients du GLMM, un PSE a également été estimé pour chaque condition (voir Tableau 7.2) en résolvant l'égalité *logit(p)=0* (c'est-à-dire $p = .5$) dans le prédicteur linéaire. Cela donnait une latence normalisée qui était ensuite reconvertie en millisecondes. Ces valeurs de *PSE* issues du modèle ($\approx$733-758 ms) étaient légèrement supérieures à celles obtenues par l'ajustement sig-

moïde (≈700 ms). Cette différence provient des contraintes mathématiques propres aux deux approches : la sigmoïde modifiée s'adapte aux proportions de réponses qui s'approchent asymptotiquement de 1.0 sans toutefois l'atteindre en cas de délais extrêmes (comme le montrent nos données empiriques).

Malgré cette légère différence, les deux méthodes convergent vers la même idée : le moment optimal n'est pas différent dans toutes les conditions, tandis que la sensibilité aux écarts par rapport à cet optimal (représentée par la pente de la fonction psychométrique) varie considérablement en fonction du style de communication du robot.

7.3.3 Perception du robot

La Figure 7.4 présente les diagrammes en boîte comparant les évaluations des participants sur les trois dimensions (*Humanité, Attrait, Étrangeté*) du questionnaire de Ho et MacDorman (2017) dans les cinq conditions. La consistance interne de chaque dimension était élevée ($\alpha = .818$ pour *Humanité*, $\alpha = .862$ pour *Attrait*, $\alpha = .883$ pour *Étrangeté*). Globalement, les évaluations du robot, tous styles de communication confondus, indiquent une faible humanité perçue, une étrangeté légèrement négative et un attrait plus variable selon les conditions. Les moyennes par condition figurent au Tableau 7.3.

Humanité

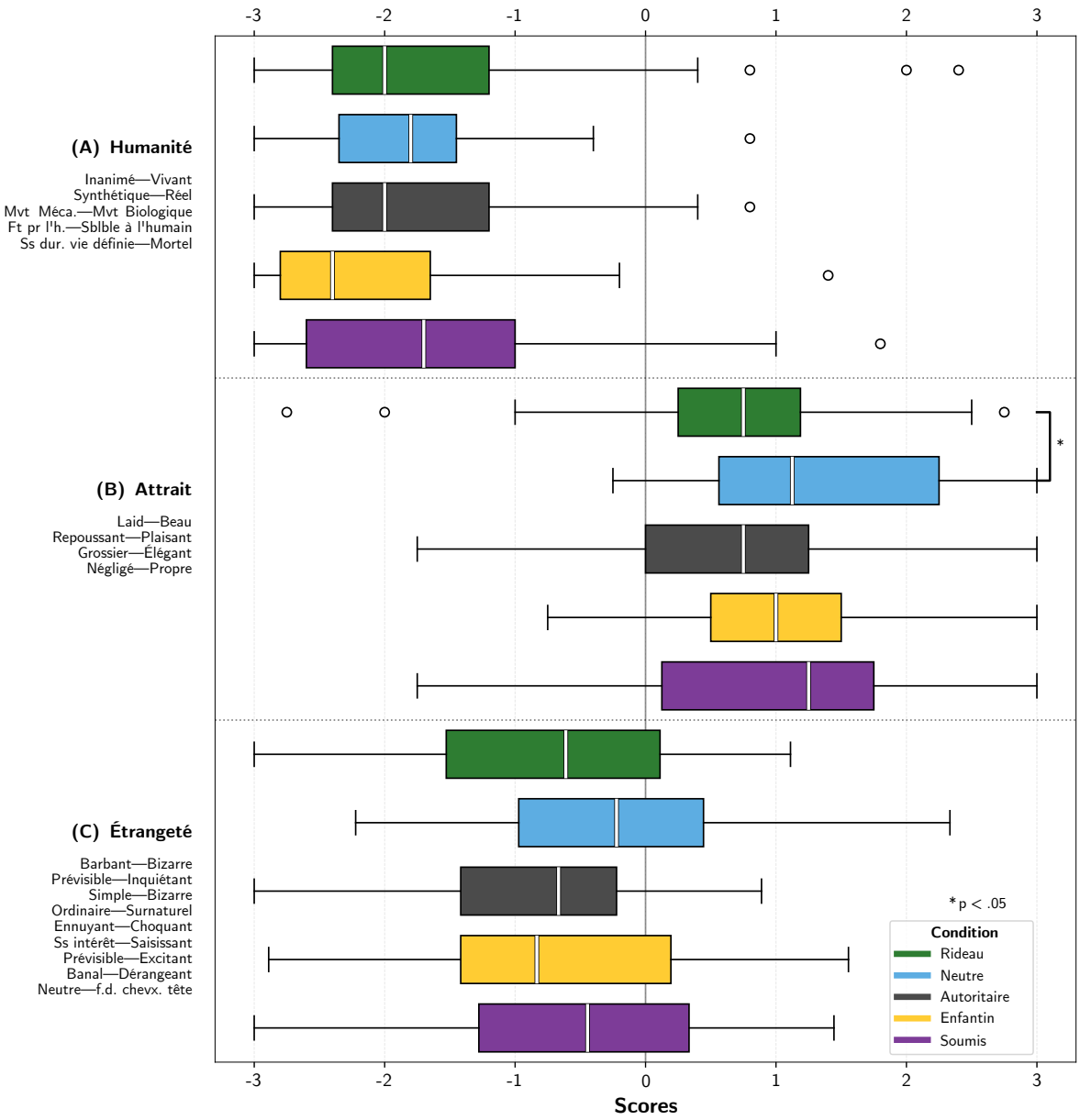
Aucune différence significative entre conditions n'a été observée pour la dimension *Humanité*, que l'on considère uniquement les quatre conditions principales ($H(3) = 4.49$, $p = .214$), ou l'ensemble incluant la condition Rideau ($H(4) = 5.66$, $p = .226$). Sans surprise, les moyennes sont toutes négatives, indiquant que le robot était perçu comme artificiel et mécanique plutôt que comme une entité vivante et biologique, et ce indépendamment du style de communication.

Attrait

Pour cette dimension, aucune différence significative n'a été observée entre les quatre conditions principales $H(3) = 5.92$, $p = .116$. En revanche, en incluant la condition Rideau, les différences devenaient significatives $H(4) = 11.99$, $p = .017$. Les scores étaient globalement positifs. La condition Neutre présente la

Figure 7.4

Scores selon la condition aux dimensions de Ho et MacDorman (2017)



Note. Scores moyens (-3 à $+3$) par condition aux dimensions (Humanité, Attrait, Étrangeté) du questionnaire de Ho et MacDorman (2017) dans l'Étude 1. La seule différence significative a été observée entre les conditions Neutre et Rideau (*$p < .05$, test post-hoc de Dunn avec correction de Bonferroni). Les boîtes représentent l'intervalle interquartile avec la ligne médiane. Les moustaches s'étendent jusqu'aux valeurs extrêmes non considérées comme des valeurs aberrantes. Les points individuels (cercles) représentent les valeurs aberrantes.

Table 7.3

Moyennes des scores aux dimensions de Ho et MacDorman (2017) par condition

Condition	Humanité	Attrait	Étrangeté
Neutre	-1.83	1.36^{a}	-0.34
	(0.84)	(1.01)	(1.01)
Autoritaire	-1.78	0.73	-0.79
	(1.00)	(1.02)	(0.90)
Enfantin	-2.07	0.98	-0.67
	(0.95)	(0.82)	(1.07)
Soumis	-1.62	0.97	-0.57
	(1.19)	(1.08)	(1.21)
Rideau	-1.52	0.57^{b}	-0.67
	(1.31)	(0.99)	(1.09)

Note. Les scores varient de -3 à $+3$ (l'écart-type est indiqué entre parenthèses). Pour chaque condition de l'Étude 1, $n = 42$. Les différentes lettres en exposant (^{a}, ^{b}) indiquent des différences significatives entre conditions selon le test post-hoc de Dunn avec correction de Bonferroni ($p < .05$). Par exemple, sur la dimension *Attrait*, la condition *Neutre*^{a} diffère significativement de la condition *Rideau*^{b}.

moyenne la plus élevée tandis qu'il s'agissait de la condition Rideau qui était la plus proche de la réponse neutre. Les comparaisons post-hoc avec le test de Dunn (correction de Bonferroni) n'ont révélé qu'une différence significative entre la condition Neutre et la condition Rideau ($p = .014$). Dans l'ensemble, ces scores positifs suggèrent que le robot est jugé esthétiquement plaisant, beau ainsi qu'agréable. Toutefois, dans la condition Rideau, les participants avaient plus de difficultés à évaluer l'attrait du robot en raison de l'absence d'indices visuels concernant son apparence, donnant lieu à des jugements plus neutres.

Étrangeté

Pour la dimension *Étrangeté*, les scores entre les conditions n'étaient pas significativement différents, que ce soit pour les quatre conditions principales $H(3) = 4.27$, $p = .234$ ou pour l'analyse incluant la condition Rideau $H(4) = 4.35$, $p = .361$. Les moyennes étaient modérément négatives dans toutes les conditions, indiquant que le robot était perçu comme plutôt prévisible et ordinaire que inquiétant ou dérangeant.

7.4 Discussion partielle

Cette première expérience apporte plusieurs éclairages sur la relation entre le style de communication d'un robot et la perception de ses délais de réponse. Premièrement, un délai de réponse optimal a été identifié autour de 700 ms, indépendant du style de communication du robot (Autoritaire, Enfantin, Neutre et Soumis) comme de son apparence ou visibilité (Rideau). Cette valeur, nettement plus longue que les ≈ 200 ms maximum typiques des discussions humain-humain (Levinson & Torreira, 2015 ; Stivers et al., 2009), suggère que les humains pourraient appliquer des modèles temporels distincts lorsqu'ils discutent avec les robots sociaux.

Une première explication de ce délai optimal plus long est qu'il pourrait refléter une adaptation intuitive aux limites techniques propres aux systèmes robotiques (par exemple, des seuils de réponse fixes, comme le souligne Skantze (2021). Cela rejoint des observations faites dans les interactions entre adultes et enfants, où des latences plus longues sont naturellement intégrées (Casillas et al., 2016), ce qui laisse penser que les humains ajustent leurs attentes lorsque de tels délais paraissent inévitables.

Une autre possibilité est que cet écart provient de la méthodologie. Étant donné que les temps de réponse dans les conversations humaines sont mesurés quantitativement via des chronométrages précis d'échanges *naturels* (par exemple des délais entre changements de locuteur en millisecondes ; Stivers et al., 2009), notre étude s'appuie sur une approche psychophysique demandant aux participants de juger explicitement si un délai est trop rapide ou trop lent. Les interactions pourraient reposer sur des normes implicites et des signaux multimodaux (par exemple, le regard ou la prosodie) afin de maintenir une coordination fluide sans longs blancs ni chevauchements (Stivers et al., 2009), alors que des jugements subjectifs explicites, comme c'est le cas dans cette étude, pourraient amener les participants à tolérer, voire même à préférer, des délais plus longs dans leur processus d'évaluation.

Même si le délai optimal n'était pas différent selon les conditions, l'analyse de la sensibilité montre que la tolérance aux écarts temporels varie significativement selon le style de communication du robot. Ces différences observées à travers des courbes psychométriques plus plates pour les styles Soumis et Enfan-

tin, reflètent une forme de tolérance à l'égard de délais de réponse s'écartant de l'optimal : lorsque le robot adopte ces styles, les participants manifestent une plus grande « zone d'incertitude » autour du délai optimal, une fenêtre temporelle étendue où le délai n'est jugé ni clairement « trop rapide » ni « trop lent ».

Cette plus grande tolérance peut être interprétée comme une potentielle adaptation des attentes sociales : des hésitations, des délais de réponse plus longs pourraient être plus attendus de la part d'un agent Soumis, de la même manière, un comportement temporel imprécis peut sembler plus naturel chez un agent adoptant un style enfantin.

De plus, dans la condition Rideau, où le robot est caché, les participants ont également montré une acceptation plus large aux écarts temporels, probablement parce qu'en l'absence d'indices visuels de la part du robot, ils ont formé d'autres attentes sur les capacités du système ou sur ses potentielles limites.

A l'inverse, les courbes plus raides pour les styles Neutre et Autoritaire indiquent des attentes plus strictes, plus définies, suggérant qu'un robot standard ou adoptant une position dominante est perçu comme devant adhérer à des normes temporelles plus précises. Ces résultats indiquent que, malgré un *PSE* ne différant pas entre les styles de communication du robot, la tolérance aux écarts par rapport à ce délai optimal varie selon le style de communication du robot.

Toutefois, l'influence du style de communication n'apparaît pas non plus dans les évaluations subjectives du questionnaire de Ho et MacDorman (2017). Quel que soit le style, dans toutes les conditions, le robot était perçu comme non-humain (scores négatifs à la dimension d'*Humanité*), et comme ordinaire et prévisible plutôt que dérangeant (scores modérément négatifs d'*Étrangeté*), sans différence significative entre les conditions pour ces deux dimensions.

En revanche, un effet significatif est apparu pour la dimension *Attrait*, où la condition Neutre a reçu un score plus élevé que la condition Rideau. Cette différence pourrait être expliquée par l'absence d'information visuelle sur le robot dans la condition Rideau, les participants ne disposant alors pas des indices nécessaires pour former un jugement esthétique à propos du robot.

Cette expérience a révélé que le style de communication du robot semblait moduler la tolérance des utilisateurs face aux variations de délai autour du temps de réponse jugé optimal.

Mais qu'en est-il lorsqu'un robot adopte systématiquement un délai non optimal ? Par exemple, est-ce qu'un robot avec un style Autoritaire serait perçu différemment s'il répond constamment trop rapidement ou trop lentement comparé au délai optimal identifié ? Ces interrogations ont motivé la conception d'une seconde étude (Chapitre 8), visant à explorer spécifiquement comment un délai de réponse non optimal influence la perception du robot en fonction de son style de communication.

Chapitre 8

Perception sociale face aux délais non-optimaux

8.1 Introduction

L'étude précédente (Chapitre 7) a démontré que les humains peuvent détecter les variations de délai de réponse d'un robot et que leur tolérance à ces écarts dépend du style de communication adopté par celui-ci. Cependant, qu'advient-il de la façon dont le robot est perçu socialement s'il répond avec un délai extrême, soit trop rapide ou soit trop lent ?

Il convient de rappeler que dans les interactions humain-humain, le délai auquel nous répondons porte une signification sociale : des réponses rapides sont fortement associées à un sentiment de connexion (Templeton et al., 2022), tandis que des délais prolongés diminuent la compétence perçue du locuteur (Matzinger et al., 2023) ou conduisent l'auditeur à inférer une issue négative, comme un refus, une critique ou un désaccord et à choisir l'interprétation la plus négative d'un énoncé poli (Bonnefon et al., 2015). Reste à savoir si les robots sociaux sont soumis aux mêmes interprétations temporelles.

Cette question est d'autant plus pertinente que les recherches en robotique sociale, comme évoqué au Chapitre 2, ont mis en évidence des schémas d'évaluation parfois contre-intuitifs et divergents de ceux retrouvés face aux humains. Cela suggère que les humains pourraient appliquer aux agents artificiels des logiques d'interprétation différentes de celles qu'ils utilisent entre eux. Deux possibilités se dessinent : soit les schémas d'interprétation temporelle propres aux interactions humaines sont transposés aux robots, soit, au contraire, le délai est perçu comme une simple caractéristique fonctionnelle, sans signification sociale. Un délai trop long traduirait alors uniquement une limitation technique

plutôt qu'un indice relationnel, et n'aurait pas d'implications pour son évaluation sociale.

Ce chapitre présente une seconde étude qui cherche à approfondir l'analyse menée dans le chapitre précédent, en examinant comment des délais fortement déviants de l'optimal identifié, combinés aux styles de communication, influencent l'évaluation globale du robot.

8.2 Méthode

8.2.1 Conception

Dans cette seconde étude, les participants visionnaient trente vidéos issues de l'Étude 1 et répondaient ensuite à deux questionnaires. L'étude reposait sur un plan inter-sujets avec douze conditions expérimentales, croisant trois niveaux de délai de réponse (trop rapide : 200 ms, optimal : 700 ms d'après le *PSE* identifié dans l'Étude 1, trop lent : 1500 ms) avec quatre styles de communication du robot (Autoritaire, Soumis, Neutre, Enfantin). L'objectif était d'examiner comment la combinaison de ces délais fixes et des styles de communication influence la perception du robot.

L'expérience comprenait trois phases : une phase d'introduction, une phase expérimentale principale, puis une phase de questionnaire. Au cours de la phase principale, les participants visionnaient des vidéos du robot. Dans la phase suivante, ils répondaient au questionnaire de Ho et MacDorman (2017) et aux dimensions choisies du modèle *Almere* (Heerink et al., 2010).

8.2.2 Participants

Un total de quatre-cent-vingt ($N = 420$) nouveaux participants (trente-cinq participants par condition, soit trois délais de réponse par quatre styles de communication du robot) ont été recrutés via *Prolific*, selon les mêmes critères que pour l'Étude 1, mais avec des restrictions géographiques plus strictes : seules les personnes résidant dans les principaux pays anglophones pouvaient participer (c'est-à-dire le Royaume-Uni, les États-Unis, l'Irlande, l'Australie, le Canada et la Nouvelle-Zélande), permettant ainsi de mieux contrôler le niveau de maîtrise de l'anglais. Les participants de l'Étude 1 étaient exclus. L'échantillon comprenait

235 hommes (56.0%), 183 femmes (43.5%), ainsi que 2 personnes n'ayant pas souhaité préciser leur genre (0.5%), âgés de 18 à 65 ans (Mdn $= 38.0$, $M = 39.7$, SD $= 11.5$).

8.2.3 Matériel et Apparatus

Plateforme expérimentale

L'expérience était accessible sur un navigateur d'ordinateur. Comme pour l'Étude 1, elle a été implémentée en *JavaScript* à l'aide de la bibliothèque *React*.

Stimuli

Les stimuli (trente vidéos par participant) provenaient de ceux de l'Étude 1 et ont été sélectionnés en privilégiant les séquences où les questions pouvaient être adressées aussi bien à un interlocuteur humain qu'à un robot (les questions et réponses sont disponibles en Annexe A.6). Chaque participant était assigné à un seul style de communication (Autoritaire, Soumis, Neutre ou Enfantin) et à un seul délai de réponse (200, 700 ou 1500 ms). La condition Rideau n'a pas été incluse dans cette seconde étude, l'objectif étant centré sur l'interaction entre le style de communication et le délai de réponse, plutôt que sur des effets liés à la visibilité, qui ne sont pas centraux aux questions de recherche ici.

L'amorçage du style de communication a suivi les mêmes principes que dans l'Étude 1, principalement transmis durant la phase d'introduction et les pauses où le robot prononçait son discours selon sa prosodie et son style de communication.

8.2.4 Procédure

L'expérience a été divisée en trois grandes phases. Contrairement à l'Étude 1, il n'y avait pas de familiarisation. Deux contrôles de l'attention étaient intégrés : l'un à mi-parcours des trente vidéos (après quinze vidéos, il fallait saisir un mot affiché à l'écran) et l'autre à mi-parcours des deux questionnaires (il fallait alors sélectionner l'item demandé)

Phase d'introduction

Durant la phase d'introduction, les participants visionnaient une vidéo où le robot se présentait et délivrait les consignes. Le contenu de son discours variait selon son style de communication et servait d'amorçage.

Phase principale

Durant la phase principale, trente vidéos étaient présentées, selon le même modèle que dans l'Étude 1. Cependant, le délai de réponse du robot était ici fixe et identique pour toutes les vidéos, correspondant à l'une des trois conditions temporelles assignées (rapide : 200 ms, optimal : 700 ms ou lent : 1500 ms). Avant chaque vidéo, les participants devaient appuyer sur la touche ESPACE pour démarrer le visionnage.

Phase de questionnaire

Afin d'approfondir l'analyse de la perception du robot menée dans l'Étude 1 et de mieux examiner comment le style de communication et le délai de réponse influencent conjointement la façon dont on évalue le robot, le questionnaire de Ho et MacDorman (2017) a été complété par un autre questionnaire issu du modèle *Almere* Heerink et al. (2010), qui appréhende des dimensions socio-émotionnelles plus larges.

Il convient de préciser que seules six dimensions du modèle *Almere* ont été utilisées : *Anxiété*, *Attitude*, *Plaisir perçu*, *Sociabilité perçue*, *Présence Sociale* et *Confiance*. Les autres dimensions (allant de l'intention d'usage du robot à l'usage de celui-ci) n'étaient pas jugées pertinentes au vu du contexte de l'expérience.

Pour certaines des dimensions sélectionnées, les items ont été adaptés afin de refléter la nature observationnelle de l'étude (par opposition à une interaction directe). Par exemple, « *J'apprécie que le robot me parle* » (« *I enjoy the robot talking to me* ») est devenu « *J'ai bien aimé que le robot parle à l'humain* » (« *I enjoyed the robot talking to the human* ») et « *J'ai l'impression que le robot me comprend* » (« *I feel the robot understands me* ») a été modifié en « *J'ai l'impression que le robot comprend les humains* » (« *I feel the robot understands humans* »).

Enfin, les items de présence sociale ont également été ajustés : « *Lorsque j'interagissais avec le robot, j'avais l'impression de parler à une vraie personne* »

(« *When interacting with the robot I felt like I'm talking to a real person* ») est devenu « *En observant l'humain interagir avec le robot, j'avais l'impression qu'il parlait à une vraie personne* » (« *When observing the human interacting with the robot, I felt as if he was talking to a real person* »)

Les deux questionnaires ont été présentés successivement, en commençant par le questionnaire du modèle *Almere* (Heerink et al., 2010), suivi de celui de Ho et MacDorman (2017) avec un ordre aléatoire des items à l'intérieur de chaque dimension.

8.2.5 Analyse des données

L'analyse des données s'est déroulée en deux temps. Premièrement, la consistance interne des dimensions a été évaluée avec l'α de Cronbach. Deuxièmement, pour chaque dimension, les hypothèses de normalité ont été vérifiées à l'aide du test de Shapiro-Wilk et l'homogénéité des variances à l'aide du test de Levene. Une ANOVA factorielle à deux facteurs sur échantillons indépendants avait été prévue afin d'examiner les effets principaux et les interactions. En cas de violation des conditions d'application, le test de Scheirer-Ray-Hare (son extension non paramétrique) a été utilisé. Les tailles d'effet ont été calculées (ε^2, pour le test de Scheirer-Ray-Hare), et les comparaisons post-hoc ont été effectuées à l'aide du test de Dunn avec correction de Bonferroni lorsque des effets principaux significatifs étaient observés.

8.3 Résultats

8.3.1 Dimensions de Ho et MacDorman (2017)

L'analyse des réponses des participants au questionnaire de Ho et MacDorman (2017) a révélé une consistance interne acceptable pour les dimensions d'*Humanité* ($\alpha = .776$), *Attrait* ($\alpha = .753$), ainsi qu'une excellente consistance interne pour *Étrangeté* ($\alpha = .881$).

L'examen des scores pour chacune des trois dimensions à l'aide des tests de Scheirer-Ray-Hare n'a montré aucun effet principal du délai de réponse (tous $p > .05$, voir dans la partie Annexe au Tableau A.3 et Figure A.4 pour des résultats détaillés par délai et style de communication). Autrement dit, le délai (200,

700 ou 1500 ms) auquel le robot répondait aux questions de l'humain n'a pas influencé la façon dont il a été évalué, que ce soit sur l'*Humanité*, l'*Étrangeté* ou l'*Attrait*.

Tout comme dans l'Étude 1, les scores selon le style de communication reflétaient une perception non-humaine du robot (scores négatifs d'*Humanité*). Les évaluations de l'*Attrait* présentaient davantage de variabilité. Pour la dimension d'*Étrangeté*, les participants ont également perçu le robot comme prévisible et ordinaire plutôt que comme dérangeant. La Figure 8.1 et le Tableau 8.1 illustrent les scores moyens aux dimensions du questionnaire de Ho et MacDorman (2017) (*Humanité, Attrait, Étrangeté*) à travers les styles de communication.

Table 8.1

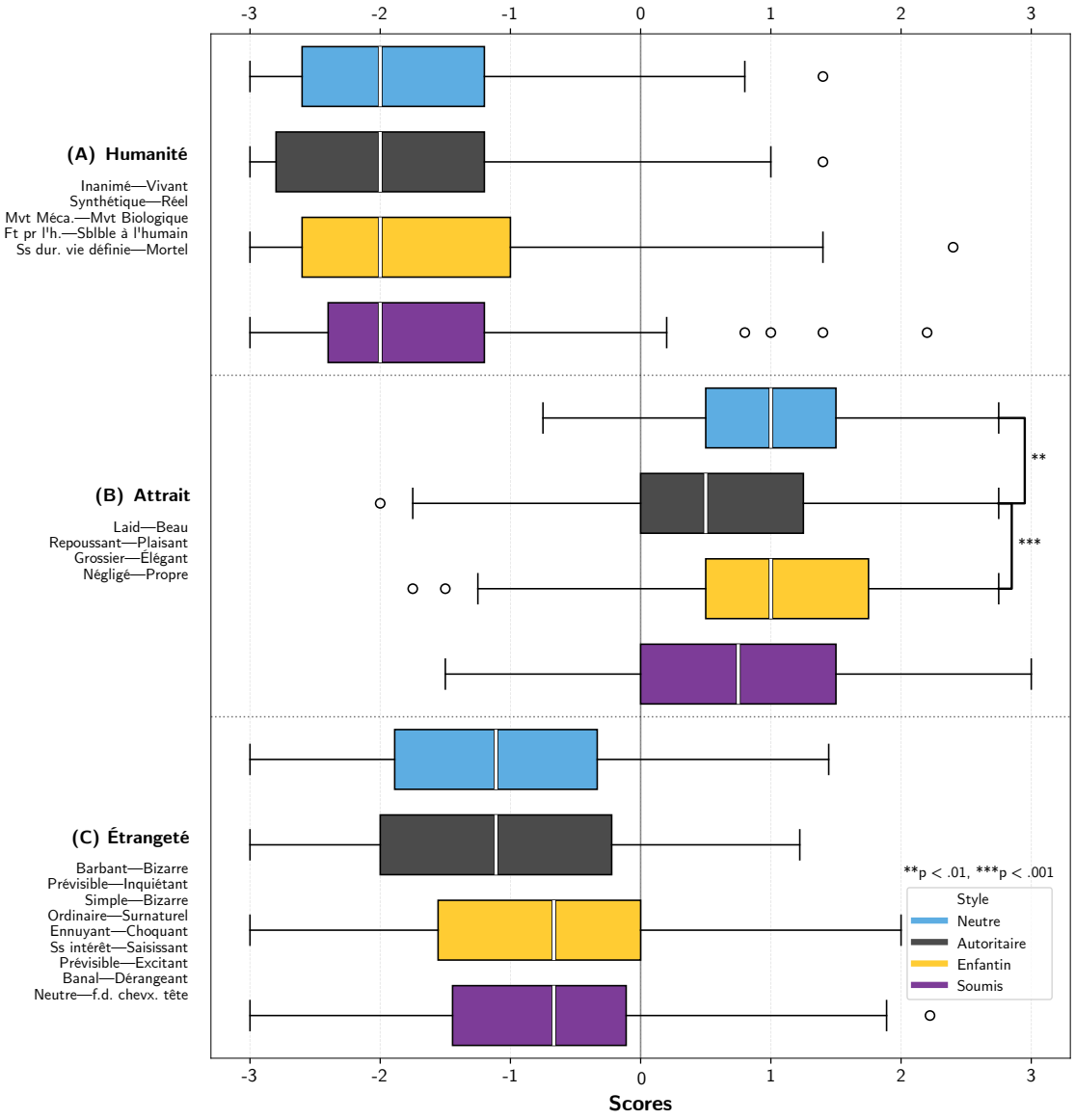
Moyennes des scores selon le style de communication et le délai de réponse aux dimensions de Ho et MacDorman (2017)

Style	Délai	Humanité	Attrait	Étrangeté
Autoritaire	Tous (n=105)	-1.85 (1.09)	0.51^{a} (1.00)	-1.04 (1.10)
	200 ms (n=35)	-1.86 (1.15)	0.20 (1.01)	-1.08 (1.08)
	700 ms (n=35)	-1.70 (1.11)	0.69 (0.99)	-0.87 (1.00)
	1500 ms (n=35)	-1.98 (1.02)	0.65 (0.94)	-1.17 (1.21)
Enfantin	Tous (n=105)	-1.71 (1.11)	1.05^{b} (0.90)	-0.74 (1.02)
	200 ms (n=35)	-1.71 (0.97)	0.99 (0.87)	-0.70 (0.98)
	700 ms (n=35)	-1.65 (1.32)	1.05 (1.01)	-0.70 (1.13)
	1500 ms (n=35)	-1.77 (1.03)	1.11 (0.85)	-0.83 (0.97)
Neutre	Tous (n=105)	-1.87 (0.95)	0.99^{b} (0.79)	-1.04 (1.07)
	200 ms (n=35)	-1.81 (1.10)	1.06 (0.85)	-0.78 (1.10)
	700 ms (n=35)	-1.83 (0.89)	0.91 (0.73)	-1.01 (0.99)
	1500 ms (n=35)	-1.96 (0.85)	1.01 (0.81)	-1.31 (1.07)
Soumis	Tous (n=105)	-1.69 (1.02)	0.80 (1.05)	-0.74 (1.08)
	200 ms (n=35)	-1.66 (1.13)	0.81 (0.88)	-0.87 (1.00)
	700 ms (n=35)	-1.68 (1.11)	0.89 (1.03)	-0.83 (1.09)
	1500 ms (n=35)	-1.74 (0.83)	0.71 (1.22)	-0.51 (1.13)

Note. Les scores varient de -3 à $+3$ (écart-type entre parenthèses). *Tous* indique la moyenne pour l'ensemble des délais de réponse. Des lettres différentes (^{a}, ^{b}) indiquent des différences significatives entre styles de communication selon les tests post-hoc de Dunn avec correction de Bonferroni ($p < .05$). Par exemple, pour la dimension *Attrait*, le style Autoritaire^{a} diffère significativement des styles Enfantin^{b} et Neutre^{b}.

Figure 8.1

Scores selon le style de communication aux dimensions de Ho et MacDorman (2017)



Note. Scores moyens (-3 à $+3$) aux dimensions du questionnaire de Ho et MacDorman, 2017 (Humanité, Attrait, Étrangeté) à travers tous les délais. Les différences significatives (test post-hoc de Dunn avec correction de Bonferroni) entre les styles de communication sont indiqués comme suit : **$p < .01$, ***$p < .001$. Les boîtes représentent l'intervalle interquartile avec la ligne médiane. Les moustaches s'étendent jusqu'aux valeurs extrêmes non considérées comme des valeurs aberrantes. Les points individuels (cercles) représentent les valeurs aberrantes.

Humanité

Pour la dimension *Humanité*, aucun effet significatif du style de communication du robot ($H(3)= 2.79$, $p = .425$, $\varepsilon^2= .007$), du délai de réponse ($H(2)= 0.69$, $p = .709$, $\varepsilon^2= .002$) ou de leur interaction ($H(6)= 1.46$, $p = .962$, $\varepsilon^2= .003$) n'a été observé. La moyenne des scores était négative pour chaque condition.

Attrait

Pour la dimension *Attrait*, aucun effet du délai ($H(2)= 0.95$, $p = .620$, $\varepsilon^2 = .0023$) ni d'interaction entre le style de communication et le délai ($H(6)= 4.38$, $p = .625$, $\varepsilon^2= .010$) n'a été observé. En revanche, un effet principal du style de communication a été trouvé ($H(3)= 20.32$, $p < .001$, $\varepsilon^2= .048$). Les comparaisons post-hoc effectuées à l'aide du test de Dunn avec correction de Bonferroni ont révélé que le style Autoritaire obtenait des scores significativement plus bas que les styles Enfantin ($p < .001$) et Neutre ($p = .003$). Le style Autoritaire était donc perçu comme le moins attrayant.

Étrangeté

Concernant la dimension *Étrangeté*, aucun effet significatif du délai ($H(2)= 1.02$, $p = .599$, $\varepsilon^2 = .002$) ni d'interaction entre délai et style de communication ($H(6)= 5.98$, $p = .426$, $\varepsilon^2 = .014$) n'a été observé. Cependant, un effet significatif du style de communication a été trouvé ($H(3)= 8.69$, $p = .034$, $\varepsilon^2= .021$), mais les comparaisons post-hoc (test de Dunn avec correction de Bonferroni) n'ont révélé aucune différence significative par paire. En regardant de plus près au niveau de chaque style de communication, les scores moyens suggèrent toutefois que le robot était perçu comme moins « inquiétant » (plus prévisible et ordinaire) lorsqu'il adoptait un style Autoritaire ($M = -1.0$) ou Neutre ($M = -1.0$), comparé aux styles Enfantin ($M = -0.7$) et Soumis ($M = -0.7$).

8.3.2 Dimensions du modèle Almere (Heerink et al. 2010)

Les dimensions du modèle *Almere* Heerink et al., 2010 utilisées dans cette étude ont montré des valeurs α de Cronbach acceptables ($\alpha > .78$) pour toutes les dimensions, à l'exception de l'échelle *Anxiété* ($\alpha = .670$). Ainsi, étant donné

ce problème de consistance interne et en se basant sur des distinctions conceptuelles, la dimension *Anxiété* a été séparée en deux sous-dimensions : (1) *Anxiété liée à l'Utilisation* (ANX-U, $\alpha = .677$), reflétant les craintes des participants à l'idée d'utiliser le robot, et (2) *Anxiété Sociale* (ANX-S, $\alpha = .781$) reflétant les préoccupations sociales et émotionnelles.

Lorsqu'elle a été analysée séparément, la sous-dimension « *Anxiété Sociale* » a pu atteindre une consistance interne acceptable, tandis que la sous-dimension « *Anxiété liée à l'Utilisation* », dont la consistance interne restait insuffisante, a été interprétée avec prudence.

Comme pour l'autre questionnaire, pour toutes les dimensions investiguées, aucun effet principal significatif du délai de réponse n'a été observé (tous les $p > .05$), ce qui indique que le délai de réponse du robot (200, 700 ou 1500 ms) n'a pas influencé les évaluations des participants (voir les résultats des tests Scheirer-Ray-Hare dans l'Annexe A.3).

Comme l'illustre la Figure 8.2, les scores donnés par les participants pour les dimensions *Almere* ont montré des tendances variées. Dans l'ensemble, les scores pour les dimensions « *Attitude* » et « *Confiance* » étaient modérément positifs pour tous les styles de communication. Le style Autoritaire, en revanche, a systématiquement reçu des évaluations plus négatives sur plusieurs dimensions par rapport aux autres styles de communication, en particulier pour « Anxiété sociale », « *Plaisir perçu* » et « *Sociabilité perçue* ». Enfin, les scores relatifs à la *Présence Sociale* sont restés généralement faibles dans toutes les conditions, ce qui suggère que les participants ont perçu une entité sociale limitée lorsqu'ils observaient les interactions du robot, quel que soit le style de communication ou le délai de réponse. Une analyse détaillée de chaque dimension est présentée ci-dessous. Les notes moyennes attribuées par les participants aux différentes dimensions en fonction du style de communication et du délai sont présentées dans le Tableau 8.2.

Anxiété

L'analyse de l'*Anxiété Sociale* a révélé un effet significatif du style de communication du robot ($H(3)= 16.86$, $p < .001$, $\varepsilon^2 = 0.040$). Les comparaisons post-hoc à l'aide du test de Dunn avec correction de Bonferroni ont montré que le style Autoritaire ($M = 2.0$) suscitait des niveaux d'*Anxiété Sociale* significativement

Table 8.2

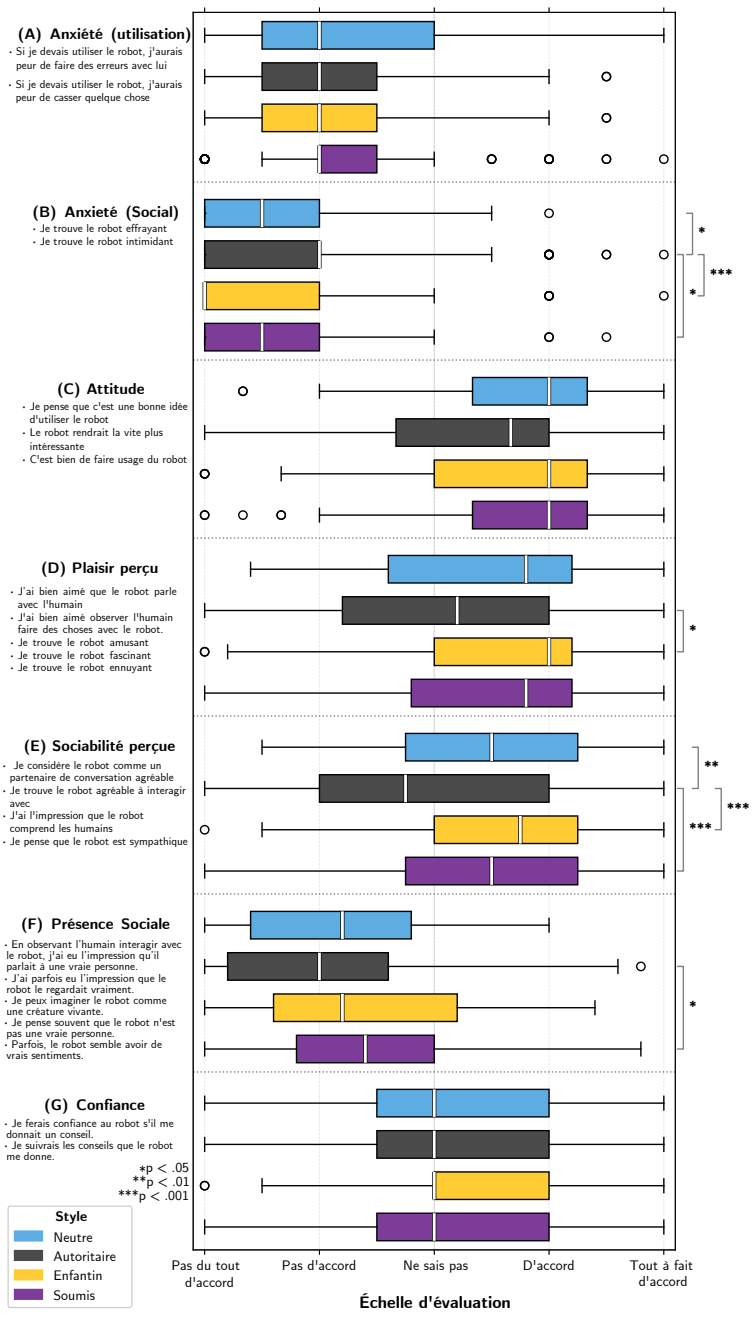
Moyennes des scores selon le style de communication et le délai au questionnaire de Almere (Heerink et al. 2010)

Style	Délai	ANX-U	ANX-S	ATT	PENJ	PS	SP	TRUST
Autoritaire	Tous	2.18	2.00^{a}	3.37	3.15^{a}	2.85^{a}	2.10^{a}	3.18
		(0.96)	(1.00)	(1.07)	(1.13)	(1.13)	(0.91)	(0.94)
	200 ms	2.19	2.19	3.25	3.18	2.76	2.46	3.04
		(0.98)	(1.12)	(1.13)	(0.69)	(1.18)	(0.80)	(0.97)
	700 ms	2.17	2.00	3.37	3.14	2.79	2.65	3.10
		(1.03)	(0.97)	(1.06)	(0.70)	(1.09)	(0.66)	(0.92)
	1500 ms	2.19	1.80	3.48	3.21	2.99	2.63	3.40
		(0.90)	(0.88)	(1.04)	(0.79)	(1.15)	(0.84)	(0.91)
Enfantin	Tous	2.02	1.55^{b}	3.68	3.58^{b}	3.53^{b}	2.30	3.26
		(0.92)	(0.79)	(0.98)	(1.09)	(0.92)	(0.91)	(0.94)
	200 ms	1.94	1.57	3.66	3.45	3.59	2.85	3.29
		(0.89)	(0.77)	(0.98)	(0.58)	(0.87)	(0.67)	(0.77)
	700 ms	2.13	1.67	3.63	3.43	3.56	2.74	3.34
		(0.91)	(0.80)	(0.99)	(0.71)	(0.96)	(0.81)	(1.03)
	1500 ms	2.00	1.40	3.75	3.38	3.45	2.71	3.14
		(0.96)	(0.79)	(0.99)	(0.62)	(0.94)	(0.76)	(1.00)
Neutre	Tous	2.17	1.61^{b}	3.66	3.51	3.41^{b}	2.20	3.11
		(0.98)	(0.68)	(0.90)	(0.98)	(0.91)	(0.84)	(0.87)
	200 ms	2.33	1.54	3.70	3.38	3.42	2.74	3.11
		(1.15)	(0.65)	(0.95)	(0.59)	(0.97)	(0.77)	(0.97)
	700 ms	2.27	1.67	3.73	3.49	3.54	2.71	3.01
		(0.92)	(0.65)	(0.84)	(0.60)	(0.82)	(0.65)	(0.82)
	1500 ms	1.90	1.61	3.55	3.25	3.28	2.60	3.20
		(0.81)	(0.74)	(0.94)	(0.58)	(0.94)	(0.64)	(0.82)
Soumis	Tous	2.23	1.60^{b}	3.63	3.48	3.44^{b}	2.43^{b}	3.18
		(0.89)	(0.69)	(0.90)	(1.05)	(0.99)	(0.89)	(0.90)
	200 ms	2.31	1.67	3.62	3.34	3.41	2.92	3.06
		(0.92)	(0.65)	(0.86)	(0.69)	(0.83)	(0.80)	(0.86)
	700 ms	2.07	1.47	3.73	3.43	3.59	3.01	3.26
		(0.94)	(0.55)	(0.99)	(0.70)	(1.08)	(0.75)	(1.00)
	1500 ms	2.31	1.66	3.54	3.22	3.31	2.79	3.23
		(0.81)	(0.83)	(0.84)	(0.69)	(1.04)	(0.76)	(0.86)

Note. Scores de 1 à 5 (avec leur SD). Par style sur l'ensemble des délais : $n = 105$; par délai et par style : $n = 35$. ANX (-U, -S) = anxiété (utilisation, sociale), ATT = attitude, PENJ = plaisir perçu, PS = sociabilité perçue, SP = présence sociale, TRUST = confiance. Les scores varient de 1 à 5 (SD entre parenthèses). Des lettres différentes (^{a}, ^{b}) indiquent des différences significatives entre styles de communication selon les tests post-hoc ($p < .05$). Par exemple, pour la *Présence sociale*, le style Autoritaire^{a} diffère significativement du style Soumis^{b}.

Figure 8.2

Scores selon le style de communication au questionnaire du modèle Almere (Heerink et al. 2010)



Note. Scores moyens (de *Pas du tout d'accord* à *Tout à fait d'accord*) aux dimensions du questionnaire de Heerink et al. (2010) selon le style de communication pour l'ensemble des délais de réponse (200, 700 et 1500 ms). *$p < .05$, **$p < .01$, ***$p < .001$.

plus élevés que les trois autres styles (Enfantin : $M = 1.5$, Neutre : $M = 1.6$, Soumis : $M = 1.6$, tous $p < .05$). Cependant, aucun effet significatif du délai de réponse ($H(2) = 3.38$, $p = .184$, $\varepsilon^2 = 0.008$) ou de l'interaction entre le style de communication et le délai ($H(6) = 4.01$, $p = .675$, $\varepsilon^2 = 0.010$) n'a été constaté.

Pour l'*Anxiété liée à l'Utilisation*, aucun effet significatif n'a été observé concernant le style de communication du robot ($H(3) = 4.36$, $p = .225$, $\varepsilon^2 = 0.010$), le délai de réponse ($H(2) = 0.50$, $p = .780$, $\varepsilon^2 = 0.001$) ou leur interaction ($H(6) = 6.57$, $p = .362$, $\varepsilon^2 = 0.016$).

Sociabilité perçue et Plaisir perçu

L'analyse de la *Sociabilité perçue* a montré un effet significatif du style de communication du robot ($H(3) = 24.33$, $p < .001$, $\varepsilon^2 = 0.058$). Le style Autoritaire ($M = 2.9$) a obtenu un score significativement inférieur à toutes les autres conditions (Enfantin : $M = 3.5$, Neutre : $M = 3.4$, Soumis : $M = 3.4$, tous $p < .01$). Cela signifie que le style Autoritaire est perçu comme le moins apte à adopter un comportement sociable. Ni l'effet du délai de réponse ($H(2) = 1.47$, $p = .479$, $\varepsilon^2 = 0.004$) ni l'interaction entre le style de communication du robot et le délai ($H(6) = 3.56$, $p = .735$, $\varepsilon^2 = 0.009$) n'étaient significatifs.

De même, l'analyse du *Plaisir perçu* a révélé un effet significatif du style de communication du robot ($H(3) = 8.52$, $p = .036$, $\varepsilon^2 = 0.020$), avec une différence significative entre le style Autoritaire ($M = 3.2$) et le style Enfantin ($M = 3.6$, $p = .036$), suggérant que le robot au style de communication autoritaire était jugé le moins agréable. Le délai de réponse ($H(2) = 1.98$, $p = .371$, $\varepsilon^2 = 0.005$) et son interaction avec le style de communication ($H(6) = 3.74$, $p = .711$, $\varepsilon^2 = 0.009$) n'étaient pas significatifs.

Présence Sociale

Pour la *Présence Sociale*, un effet significatif du style de communication du robot a été trouvé ($H(3) = 8.51$, $p = .037$, $\varepsilon^2 = 0.020$), avec une différence significative entre le style Autoritaire ($M = 2.1$) et le style Soumis ($M = 2.4$, $p = .028$). Autrement dit, le visionnage de vidéos avec le style Soumis évoque le plus efficacement l'expérience de percevoir une entité sociale. Comme pour les dimensions précédentes, ni le délai de réponse ($H(2) = 1.39$, $p = .499$, $\varepsilon^2 = 0.003$) ni son

interaction avec le style de communication ($H(6)= 1.29$, $p = .972$, $\varepsilon^2 = 0.003$) n'étaient significatifs.

Attitude et Confiance

Pour la dimension *Attitude*, aucun effet significatif n'a été observé pour l'effet du style de communication du robot ($H(3)= 5.96$, $p = .113$, $\varepsilon^2 = 0.014$), ni pour le délai de réponse ($H(2)= 0.42$, $p = .811$, $\varepsilon^2 = 0.001$), ni pour leur interaction ($H(6) = 3.53$, $p = .740$, $\varepsilon^2 = 0.001$).

De même, la dimension *Confiance* n'a montré aucun effet significatif du style de communication du robot ($H(3)= 1.50$, $p = .683$, $\varepsilon^2 = 0.004$), du délai de réponse ($H(2)= 1.09$, $p = .581$, $\varepsilon^2 = 0.003$) ni de leur interaction ($H(6)= 5.15$, $p = .524$, $\varepsilon^2 = 0.012$).

8.4 Discussion partielle

L'Étude 2 a examiné comment les délais de réponse d'un robot, s'écartant du délai optimal (identifié dans l'Étude 1), interagissent avec son style de communication pour influencer la perception des participants. Contrairement à ce qui était attendu, le délai de réponse n'a pas eu d'effet significatif sur la perception du robot, telle que mesurée par les dimensions des questionnaires de Ho et MacDorman (2017) et de *Almere* (Heerink et al., 2010). Cette absence d'effet est assez surprenante étant donné que les délais extrêmes choisis (200 ms et 1500 ms) représentent des écarts importants par rapport au délai optimal identifié dans l'Étude 1. Une explication possible est que lorsque le délai reste constant tout au long de l'interaction, les participants s'adaptent rapidement à ce rythme, quelle que soit sa vitesse absolue et cela n'impacte donc pas la façon dont ils évalueront le robot.

En revanche, le style de communication du robot a exercé un effet constant sur la perception des utilisateurs à travers plusieurs dimensions sociales et liées à l'interaction. La condition Autoritaire s'est particulièrement démarquée, générant systématiquement des évaluations moins favorables : une *Anxiété Sociale* plus élevée, une *Sociabilité perçue*, un *Plaisir perçu*, une *Présence Sociale* et un *Attrait* réduits. Ces résultats concordent avec ceux de Saunderson et Nejat (2021), qui ont montré qu'un robot est plus persuasif lorsqu'il est positionné comme un

pair donnant des incitations positives, plutôt que comme une figure autoritaire imposant un contrôle ou faisant des incitations négatives. Dans cette même étude, les auteurs ont également révélé qu'un robot autoritaire est évalué de façon plus négative, avec notamment une acceptation plus faible et une plus grande résistance de la part des participants.

8.5 Discussion générale

Ces deux études ont permis d'examiner de manière complémentaire comment le style de communication d'un robot social interagit avec le délai de réponse pour moduler les perceptions des individus dans le cadre d'une discussion simple de type question-réponse (oui/non). La première étude visait à identifier le délai de réponse jugé « optimal » pour ce type d'échange et à explorer comment différents styles de communication d'un robot influencent cette perception. La deuxième étude a examiné l'impact de délais s'écartant fortement de cet optimal, combinés à des styles de communication, sur la perception globale qu'ont les participants du robot.

L'Étude 1 a permis de déterminer un délai de réponse optimal (PSE) d'environ $\approx$700 ms, qui peut varier d'une personne à une autre (au vu de la dispersion des délais de réponse préférés) mais qui semble néanmoins constituer une valeur de référence, indépendamment du style de communication adopté par le robot. Ces résultats peuvent indiquer que les humains développent des attentes temporelles spécifiques concernant le tour de parole et les réponses des robots, qui se distingueraient de celles mobilisées dans la communication entre humains. Pour rappel, le délai de réponse dans un échange entre humains est d'environ $\approx$200 ms (Stivers et al., 2009).

En revanche, certains styles de communication rendent les écarts temporels par rapport à cet optimal plus acceptables que d'autres. Les individus montrent une tolérance nettement plus grande aux écarts par rapport au délai optimal avec les robots ayant un style de communication soumis ou enfantin comparativement à un style neutre. Le style autoritaire suscitait une tolérance similaire à celle du style neutre. Cela suggère que les robots projetant des personnalités « douces » peuvent bénéficier d'une plus grande souplesse ou une marge de manœvre plus grande dans leur délai de réponse, tandis que les robots perçus

comme plus dominants sont soumis à des attentes plus strictes.

L'Étude 2 a révélé un résultat particulièrement inattendu : le fait que les délais de réponse d'un robot soient beaucoup plus courts ou beaucoup plus longs que le délai optimal de 700 ms, n'a aucun impact mesurable sur la perception faite du robot, sur son *Humanité*, son *Attrait*, son *Étrangeté*, sa *Sociabilité perçue*, ou même l'*Anxiété* qu'il peut provoquer. Ceci est particulièrement intéressant puisque l'Étude 1 montre clairement que les participants perçoivent bien les variations temporelles et qu'il existe un délai préférable : ils jugeaient plusieurs délais comme « trop rapides » ou « trop lents » par rapport à leur point optimal.

Pourtant, plusieurs études ont montré que les temps de réponse peuvent affecter divers aspects des perceptions sociales dans les conversations humaines : des réponses plus rapides (inférieures à 250 ms) sont associées à un lien social plus fort (Templeton et al., 2022), tandis que des temps de réponse plus lents peuvent inciter les interlocuteurs à interpréter les propos de manière plus négative dans des contextes impliquant des commentaires sensibles ou négatifs (Bonnefon et al., 2015). De même, il a été démontré que des pauses plus longues avant de répondre à des questions de connaissances affectent le jugement sur la compétence et la confiance de l'orateur Matzinger et al., 2023.

Étant donné les effets bien documentés des schémas temporels dans la communication humaine, le contraste entre ces résultats et l'absence d'impact mesurable dans notre étude soulève plusieurs possibilités :

— les questionnaires utilisés, bien que sensibles aux variations dans le style de communication des robots, ne permettent peut-être pas de saisir les dimensions sociales spécifiques influencées par le délai de réponse ;

— les délais de réponse n'influencent peut-être tout simplement pas la façon dont les humains évaluent les robots (du moins pour l'instant) ;

— les humains peuvent appliquer aux robots des attentes temporelles fondamentalement différentes de celles qu'ils appliquent aux interlocuteurs humains ;

— enfin, les délais de réponse des robots pourraient être traités comme des caractéristiques fonctionnelles et mécaniques plutôt que comme des signaux sociaux.

Ce constat ouvre de nouvelles questions et perspectives quant à la possibi-

lité que les humains appliquent différemment la cognition sociale aux agents artificiels et aux autres humains : ils attribuent facilement une signification sociale au style de communication d'un robot, mais pas à ses schémas temporels dans l'interaction. Dans cette perspective, nos deux expériences indiquent que le style de communication d'un robot influence fortement son évaluation globale, contrairement à son délai de réponse. En particulier, le style autoritaire se distingue par des évaluations moins favorables, et ce, indépendamment du délai, sur des dimensions telles que l'*Anxiété*, le *Plaisir perçu* ou la *Sociabilité perçue*.

Cette réception négative du style de communication autoritaire rejoint et s'aligne avec la théorie de la réactance psychologique (Brehm, 1966), qui postule que des menaces (perçues) à la liberté d'action déclenchent un mécanisme de défense sous la forme d'un état motivationnel visant à restaurer cette liberté, se manifestant souvent par de la colère et des cognitions de résistance. Roubroeks et al. (2011) ont par exemple montré que, dans une interaction avec un agent artificiel social, l'usage d'un langage hautement contrôlant, c'est-à-dire menaçant l'autonomie de l'utilisateur (par exemple, « Tu dois... »), provoque une réactance bien plus élevée que l'emploi d'un langage faiblement contrôlant ou non menaçant (par exemple, « Tu pourrais... »). Cette réactance se manifestait par des scores de colère plus élevés et un pourcentage plus élevé de pensées négatives. De façon similaire, ces résultats rejoignent ceux de Saunderson et Nejat (2021), qui ont montré qu'un robot adoptant un rôle d'autorité est moins persuasif qu'un robot se positionnant comme un pair, les récompenses (incitations positives) s'avérant plus efficaces que les punitions (incitations négatives) dans les deux cas, et la résistance des participants étant particulièrement marquée envers le robot en position d'autorité.

Ainsi, même si les styles de communication autoritaires pourraient éventuellement se révéler intéressants dans certaines situations fonctionnelles nécessitant des directives claires (par exemple, les robots de sécurité ou d'instruction), il semble que leur utilisation à plus grande échelle soit limitée et puisse même entraîner des réactions contre-productives.

8.5.1 Limites

Plusieurs limites doivent être prises en compte lors de l'interprétation de ces résultats. Premièrement, le protocole repose sur des vidéos d'interactions (point de vue à la troisième personne) plutôt que sur des interactions directes en temps réel, ce qui a pu influencer les jugements temporels des participants et leurs évaluations du robot. Ce choix méthodologique offrait un plus grand contrôle expérimental mais il pourrait potentiellement avoir réduit la validité écologique étant donné que les interactions en temps réel mobilisent des indices supplémentaires et des adaptations dynamiques qui n'étaient pas capturées par le format vidéo. À la différence des études sur l'interaction entre humains, nous n'avons pas mesuré les intervalles effectifs entre tours de parole, le délai optimal obtenu a été extrait des jugements explicites des participants concernant chaque délai de réponse.

Deuxièmement, ce protocole s'est exclusivement concentré sur des réponses *oui/non* à des questions, qui ne représentent qu'une sous-catégorie des échanges conversationnels. Des interactions plus complexes, impliquant des questions ouvertes ou des dialogues avec plusieurs tours, pourraient révéler des dynamiques temporelles différentes ainsi qu'une sensibilité accrue au délai des réponses. En effet, Funakoshi et al. (2010) rapportent que les préférences des utilisateurs en matière de latence varient selon les schémas d'interaction : les participants attendent des réponses verbales plus rapides aux salutations et aux réponses affirmatives, mais tolèrent des réponses verbales plus lentes à des requêtes complexes lorsque le système émet un accusé de réception non verbal immédiat (par exemple, via une LED clignotante).

Troisièmement, notre protocole s'est appuyé sur des questionnaires (Heerink et al., 2010 ; Ho & MacDorman, 2017) pour évaluer la perception du robot, ce qui permet de recueillir des jugements subjectifs, mais non les réactions comportementales implicites. Il se pourrait que des délais non optimaux influencent des réactions physiologiques ou non conscientes qui n'ont pas été mesurées ici. Des travaux recourant à l'oculométrie ou à des mesures physiologiques pourraient fournir une compréhension plus complète de la façon dont le délai de réponse affecte l'interaction.

Quatrièmement, même si la condition Rideau a permis d'aider à isoler l'influence des indices visuels, les résultats demeurent largement spécifiques au ro-

bot *Buddy* avec son apparence physique particulière ; d'autres incarnations de robots pourraient susciter des attentes temporelles différentes.

Cinquièmement, le choix méthodologique d'avoir utilisé des audios *oui/non* identiques pour l'ensemble des essais permettait de favoriser un plus grand contrôle expérimental, mais un tel choix a peut-être, dans le même temps, réduit la probabilité que les participants attribuent une signification sociale aux délais. Cela pourrait en partie expliquer pourquoi les délais de réponse, bien que détectés (dans l'Étude 1) n'ont pas eu d'impact sur l'évaluation sociale du robot (dans l'Étude 2). Dans de tels contextes de discussion entre humains et robots, il se pourrait que l'attribution de sens social au délai dépende de son intégration à d'autres indices temporels (par exemple, des marqueurs d'hésitations, des variations prosodiques, ou de la vitesse d'articulation).

Enfin, même si la manipulation expérimentale a effectivement permis de créer différents styles de communication, les quatre styles étudiés sont des simplifications de style de communication et ne sont pas des profils de personnalité complets. Des futurs travaux devraient prendre en considération ces limites en examinant notamment des interactions en temps réel avec divers robots, en intégrant des structures de conversation plus complexes et en explorant des dimensions de la communication plus subtiles.

En somme, cette Partie III a permis d'examiner la temporalité des échanges verbaux avec un robot ainsi que l'impact du style de communication qu'il adopte lorsqu'il s'adresse à l'humain. Le Chapitre 7 présentait une étude nous ayant permis d'identifier le délai de réponse optimal d'un robot et de montrer que la tolérance aux écarts par rapport à ce délai dépend du style de communication utilisé par un robot. Le Chapitre 8 a ensuite montré que les délais de réponse d'un robot n'influencent pas pour autant la perception que les individus ont de lui. Finalement, ce qui façonne cette perception n'est pas tant le temps qu'il faut au robot pour répondre que la manière dont il s'adresse à nous.

Après avoir analysé la dynamique temporelle et le style de communication du robot, il convient désormais de s'interroger sur un autre aspect fondamental de la conversation : le contenu même du discours. La Partie IV explorera ainsi la façon dont les humains traitent les discours d'un robot lorsque ceux-ci dépassent les limites de ce qui lui est réellement accessible ou possible.

Quatrième partie

Approche EEG des frontières et limites face au discours robotique

Approche EEG des frontières et limites face au discours robotique

Les travaux présentés dans cette dernière partie examinent comment les humains traitent les énoncés produits par des robots sociaux lorsque ceux-ci dépassent les limites de ce qui leur est accessible et possible. L'objectif est de tester expérimentalement la façon dont nous évaluons le robot et son discours et d'identifier les potentielles frontières cognitives que nous établissons face au discours des robots sociaux. Dans cet objectif, la Partie IV explore deux situations où le robot tient un discours potentiellement incongruent avec ses capacités : (1) lorsqu'il parle d'actions physiques impossibles au vu de sa morphologie (Chapitre 9), et (2) lorsqu'il parle de ses propres émotions, au risque de dépasser les limites que l'observateur est prêt à lui accorder en tant qu'entité artificielle (Chapitre 10).

Le Chapitre 9 décrit la première étude visant à valider un paradigme N400, centrée sur la réaction des participants face au discours d'un robot dépourvu de jambes et de bras qui parle d'actions physiques qui lui sont inaccessibles. Le Chapitre 10 présente la seconde étude, qui applique ce paradigme à des énoncés portant sur les émotions, abordant cette fois non plus les limites physiques de ses capacités, mais celles relevant d'un registre expérientiel et subjectif qui ne lui est pas propre. L'ensemble des travaux présentés dans cette Partie IV a donné lieu à plusieurs publications (Gigandet & Nazir, 2025 ; Gigandet et al., 2023, 2024)

Chapitre 9

Un robot parlant d'actions physiques impossibles : validation du paradigme

9.1 Introduction

Lorsqu'un autre humain énonce quelque chose qui contredit manifestement nos connaissances du monde, nous mobilisons plusieurs mécanismes. Ceux-ci portent à la fois sur l'individu source, en évaluant sa compétence et sa bienveillance ainsi que sur le contenu du message en jugeant la crédibilité intrinsèque de l'énoncé et sa cohérence (Sperber et al., 2010).

La composante N400 se présente comme un indicateur particulièrement pertinent pour investiguer la façon dont les humains réagissent au niveau cérébral à de tels énoncés. Pour rappel, la N400 est une composante des ERP, généralement induite par une incongruité sémantique, avec une amplitude légèrement plus importante dans l'hémisphère droit que dans l'hémisphère gauche, qui apparaît environ 400 ms après la présentation d'un mot ou stimulus qui est sémantiquement incongruent avec son contexte (Kutas & Hillyard, 1980; Luck, 2005).

La N400 n'est pas seulement associée à la congruence du langage, mais aussi à la congruence de l'action (Reid & Striano, 2008; van Elk et al., 2008). Des images présentées avec des mots incongruents (Friedrich & Friederici, 2004; Hamm et al., 2002), ainsi qu'un contexte où l'incongruence est créée par le locuteur (van Berkum et al., 2008) la déclenchent également. (Kutas & Federmeier, 2011). van Berkum et al. (2003) ont en effet montré qu'une phrase cohérente telle que « *Chaque soir, je bois un peu de vin avant d'aller me coucher* » suscite une réponse N400 plus forte lorsqu'elle est prononcée par la voix d'un enfant que lorsqu'elle est prononcée par une personne adulte (van Berkum et al., 2008).

Une telle phrase ne passe donc pas inaperçue : le contenu ne correspond pas aux attentes liées à l'âge du locuteur. En somme, cette évaluation rapide de la cohérence entre *qui* parle et *ce qui est dit* se produit de façon automatique, en quelques centaines de millisecondes.

Mais comment traitons-nous le discours d'un robot ? Serions-nous surpris par l'incongruence entre ce qui est dit et ce que l'on perçoit de ses capacités ? Autrement dit, comment l'humain traite-t-il un discours qui, bien que grammaticalement correct et formulé d'une façon compréhensible, semble incompatible avec l'agent qui le prononce ?

En transposant cette idée à un robot, on peut s'attendre à ce que les énoncés décrivant des actions incompatibles avec le corps du robot suscitent une augmentation de la N400 chez la personne écoutant un tel discours. Cependant, l'étendue de cette réaction pourrait dépendre des croyances préalables concernant les capacités du robot.

En effet, dans la culture populaire, les robots sont fréquemment mis en scène comme possédant des membres rétractables ou des fonctionnalités dissimulées qui défient leur apparence initiale, à l'image du personnage d'*Eve* dans le film *Wall-E* ou des personnages de la franchise *Transformers*. La croyance en de telles capacités cachées pourrait réduire, voire éliminer, l'incongruence perçue dans les énoncés du robot entre sa morphologie et les capacités qu'il doit posséder pour affirmer ces énoncés. Autrement dit, une personne convaincue que « tout est possible » avec un robot verra ses attentes moins contredites par un énoncé décrivant une action apparemment impossible, contrairement à un individu plus sceptique, qui se fiera davantage à l'apparence visible du robot pour juger de la plausibilité de l'énoncé.

Ainsi, ce chapitre présente la validation d'un paradigme expérimental permettant d'observer l'effet N400 lié aux énoncés incongruents d'un robot, en se concentrant sur le cas des actions physiquement impossibles. Il examine également dans quelle mesure cet effet varie selon la croyance des participants dans les éventuelles capacités cachées du robot.

9.2 Méthode

9.2.1 Conception

Cette étude a adopté un plan mixte dans lequel le facteur inter-sujets correspondant à l'apparence du corps du robot comporte deux conditions : (1) la condition « *BODY* » où le corps entier du robot est visible ; (2) la condition « *HEAD* » (contrôle) où seule la tête du robot est visible. Chaque participant était assigné à une condition d'apparence. Le facteur inter-sujets relatif à la congruence de l'énoncé du robot opposait deux modalités : (1) il contenait un mot cible incompatible avec les capacités physiques visibles du robot, (2) il contenait un mot cible compatible avec les capacités physiques visibles du robot.

L'expérience comprenait quatre phases : une phase de vérifications techniques et de recueil du consentement, une phase d'introduction, une phase principale et enfin une phase de questionnaire.

Au cours de la phase principale, les participants visionnaient des vidéos du robot s'exprimant, pendant que le signal EEG était enregistré. Dans la phase suivante, ils répondaient au questionnaire de Ho et MacDorman (2010) ainsi qu'à cinq traits supplémentaires relatifs à l'*Imagination*, l'*Intelligence*, l'*Indépendance*, la *Créativité* et le fait d'être *Bavard*. Les variables dépendantes de cette étude comprennent l'amplitude moyenne du pic N400 mesurée sur treize électrodes d'intérêt, ainsi que les réponses aux trois dimensions du questionnaire de Ho et MacDorman (2010), aux cinq traits supplémentaires et à deux items évaluant la possibilité que le robot possède des bras ou des jambes dissimulés.

9.2.2 Participants

Au total, cinquante-six participants droitiers dont la langue maternelle est le français (dont 1 personne non-binaire, 37 femmes et 17 hommes) et âgés de 18 à 58 ans ($M = 24.25$, $Mdn = 23$) ont pris part à l'expérience. Parmi ceux-ci, trente-deux participants ont été assignés à la condition dans laquelle le robot était visible dans son intégralité et vingt-quatre participants à la condition dans laquelle seule la tête du robot était visible. Les participants ont été rémunérés 15 € pour leur participation.

La population de l'échantillon ne présentait pas de troubles neurologiques ou psychiatriques et ne recevait pas de médicaments neuroleptiques. Les personnes présentant des troubles de la vision tels que la myopie étaient autorisées à participer, à condition de porter leurs lunettes correctrices. Enfin, la latéralité manuelle des participants a été vérifiée à l'aide du *Edinburgh Handedness Inventory* (Oldfield, 1971). Les personnes ayant déjà rencontré le robot ont été automatiquement assignées à la condition *BODY* pour empêcher leurs connaissances préalables d'influencer les données : cette assignation explique le nombre plus élevé de participants dans la condition *BODY* que dans la condition *HEAD*.

9.2.3 Matériel et Apparatus

Plateforme expérimentale

L'expérience a été menée dans le laboratoire EEG de la plateforme de la *FR 2052 SCV - Sciences et Cultures du Visuel* à Tourcoing (France). L'affichage des stimuli de l'expérience sur l'écran de l'ordinateur du participant (vidéos, écrans noirs, croix de fixation, etc.) a été créé en utilisant *Psychopy*.

Stimuli

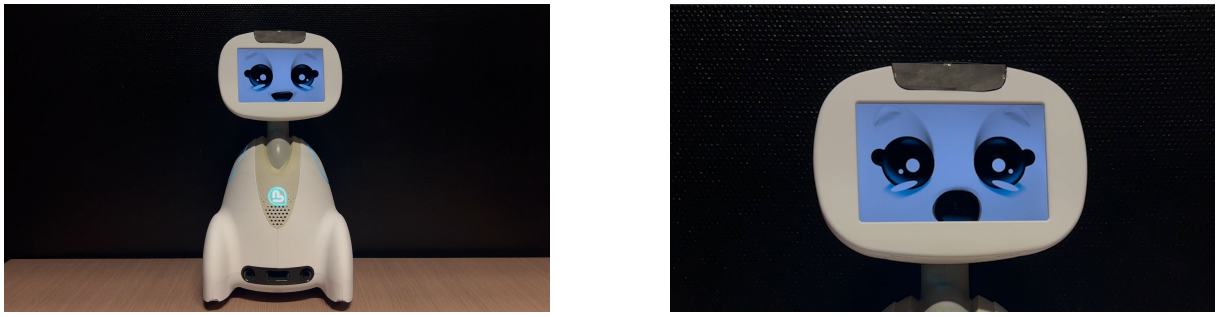
Dans cette expérience, le robot *Buddy* (*Blue Frog Robotics*), dépourvu de bras comme de jambes, a été utilisé. Il est nommé « *Lou* » pour les besoins de l'étude. Le robot a été présenté aux participants à travers des vidéos, selon deux conditions : (1) corps entier (BODY) ou (2) tête seulement (HEAD). Dans chaque condition, le robot s'exprimait à l'oral dans soixante courtes vidéos scénarisées. Dans la condition où les participants ne voyaient que la tête du robot (*HEAD*), le cadrage de la vidéo se limitait à la tête et à son visage. Ce cadrage permettait d'occulter l'absence de membres (bras et jambes) et empêchait le participant d'inférer l'éventuelle existence de ceux-ci. Cette condition servait de contrôle, permettant d'évaluer l'effet de la visibilité du corps sur la perception du discours du robot. Dans la condition où le corps entier du robot était visible (*BODY*), son apparence révélait clairement l'absence de bras et de jambes. Cette condition permettait ainsi de rendre manifeste l'incongruence potentielle entre certains

énoncés du robot et ses capacités physiques observables. La Figure 9.1 présente des captures d'écran des vidéos de chacune des deux conditions.

Figure 9.1

Captures d'écran des vidéos des conditions HEAD et BODY

(a) *(b)*



Note. **(a)** Condition où le corps entier du robot est visible (BODY) ; **(b)** Condition où seule la tête du robot est visible (HEAD)

Concernant le discours du robot, nous avons d'abord construit soixante énoncés comportant typiquement une ou deux phrases et se terminant par le mot cible. Chaque énoncé existe avec deux fins alternatives : une version avec un mot cible congruent (possible) et une version avec un mot cible incongruent (physiquement impossible compte tenu de l'apparence du robot). Les cent-vingt énoncés finaux étaient des phrases simples en français dans lesquelles le robot parlait à la première personne. Les sujets évoqués couvraient une gamme d'actions simples (ramasser un objet, caresser un chat ou des loisirs tels que jardiner, courir) avec des situations diverses comme partir en vacances ou s'habiller d'une certaine manière.

Le premier type de fin comprenait un mot cible qui était congruent avec l'apparence physique du robot dans les conditions *HEAD* et *BODY*. Par exemple, « *J'espère aller un jour tout en haut de la Tour Eiffel. Pour monter, j'emprunterai l'**ascenseur*** ». Le second type de fin avait un mot cible qui était incongruent dans la condition *BODY*, faisant référence à des actions qui impliquent des bras et/ou des mains ou des jambes. Par exemple, pour la même phrase, le robot dira « [...] *j'emprunterai l'**escalier*** ». Une telle incongruence ne pouvait pas être perçue comme telle dans la condition HEAD, étant donné que les participants ne

pouvaient pas voir que le robot ne disposait pas de bras ou de jambes.

Les mots cibles dans les deux conditions (par exemple, ascenseur/escalier) étaient contrôlés pour diverses propriétés linguistiques tirées de la base de données *LEXIQUE* (New et al., 2001). Il est important de souligner que la *cloze probability*, qui est la probabilité qu'une phrase se termine par un mot spécifique en fonction du contexte, exerce une influence sur la composante N400 (Desroches et al., 2009). Une *cloze probability* plus faible signifie un mot moins anticipé, donc nécessitant plus d'efforts pour l'intégration sémantique et pouvant ainsi potentiellement susciter une réponse N400 plus prononcée. Cela souligne la nécessité d'un contrôle précis.

Pour gérer cette variable, chacune des soixante phrases françaises a d'abord été évaluée par vingt-cinq évaluateurs humains qui devaient choisir l'un des deux mots alternatifs pour compléter la phrase, par exemple : « [...] *Pour monter, je prendrai...* » (ascenseur/escalier). Il était donc attendu que ces mots cibles de fin de phrase soient choisis de façon équilibrée. Ceux avec des *cloze probability* déséquilibrées étaient ensuite remplacés et soumis à nouveau à une réévaluation par un autre ensemble de vingt-cinq évaluateurs. Les listes de stimuli définitives ont été déterminées après cinq rounds de tests, chacun impliquant au moins vingt-cinq évaluateurs différents. En raison du défi inhérent à l'obtention d'un équilibre parfait pour la *cloze probability* dans toutes les phrases, certains mots ont été acceptés avec une *cloze probability* minimale de 0.30. Cependant, en moyenne sur l'ensemble de la liste de phrases, la *cloze probability* était de 0.50. En raison des contraintes imposées par l'équilibrage de celle-ci, il n'a pas été possible d'obtenir un équilibre parfait pour certaines autres variables linguistiques, telles que le nombre de lettres, de syllabes et de phonèmes.

Par conséquent, la condition *HEAD* (contrôle) a joué un rôle pivot dans la validation de notre matériel de phrase : il a été considéré que les stimuli de mots étaient bien équilibrés si les ERP dans la condition *HEAD* (qui n'avait pas de phrases en conflit avec l'apparence physique du robot) ne montraient pas de différences significatives dans les deux conditions de phrase.

Un dernier point qui nécessitait une attention particulière était la longueur moyenne des mots cibles (7-8 lettres), qui dépassait largement la longueur typiquement utilisée dans un paradigme N400 (4-5 lettres ; comme par exemple Desroches et al., 2009). Lorsque les mots deviennent plus longs, le *Point d'Uni-*

cité Phonologique (*Phonological Uniqueness Point*, *UP* ; Marslen-Wilson et Welsh, 1978) s'éloigne davantage du début du mot. L'*UP* phonologique fait référence au moment dans le traitement auditif du langage parlé où un mot peut être identifié de manière unique en fonction de ses propriétés phonologiques, avant qu'il ne soit entièrement prononcé (Marslen-Wilson & Welsh, 1978). Par exemple, l'*UP* pour un mot comme « congruence » en anglais (kɔˈngɹuʌns) se produit après les sons formant « cong- » (kɔˈng), car il n'y a pas d'autres mots dans cette langue qui commencent par ces éléments phonologiques. Par conséquent, les auditeurs n'ont pas besoin d'entendre un mot entier pour le comprendre, mais peuvent souvent prédire le mot en cours de route. La position de l'*UP* dans un mot affecte le délai temporel du pic N400 (Desroches et al., 2009). Comme les *UP* dans nos mots cibles se produisent en moyenne au cinquième phonème ou après (voir Tableau 9.1), nous nous attendons à ce que le pic N400 se produise à un moment plus tardif que celui décrit dans la plupart des études utilisant ce paradigme.

Ensuite, les 2x60 phrases ont été divisées en deux listes équivalentes, chacune comprenant trente phrases avec des fins congruentes et trente avec des fins incongruentes. Les listes ont été distribuées de manière égale, garantissant que chaque liste était présentée au même nombre de participants. Cette approche a garanti que chaque participant était exposé à une seule version de chaque phrase (soit avec une fin congruente, soit avec une fin incongruente). Pour rappel, l'ordre des phrases pour chaque personne est différent car elles étaient mélangées aléatoirement. Par conséquent, sur l'ensemble des participants, chaque phrase a été entendue dans ses deux versions : un premier groupe a reçu la phrase avec une fin congruente, tandis qu'un second groupe a entendu la même phrase avec une fin incongruente. Cette configuration a maintenu l'équilibre en garantissant qu'un nombre égal de participants ait entendu les versions congruentes et incongruentes de chaque phrase.

Les enregistrements audio de ces phrases ont été réalisés avec une locutrice féminine à l'aide d'un microphone (*Shure SM58*), d'une interface audio (*UMC202HD*), d'un ordinateur portable (*MacBook Pro M1 Max 2021*) et du logiciel *Audacity*. Le choix s'est porté sur l'utilisation d'une voix humaine plutôt que du module vocal de synthèse vocale de base du robot afin de minimiser les variations potentielles dans la perception de l'articulation et de l'intonation, qui pourraient potentiellement influencer les réponses des participants, notam-

Table 9.1

Caractéristiques des mots cibles dans l'Étude 1 : mesures linguistiques et phonologiques

	Congruence du mot cible		
Mesure	**Incongruent**	**Congruent**	**Mann-Whitney**
Point d'unicité phonologique	4.97 (SD=1.44)	5.68 (SD=1.56)	U=1273.5, p=.004
Fréquence du lemme dans les films	70.39 (SD=131.58)	161.74 (SD=398.83)	U=1505.0, p=.122
Fréquence du mot dans les films	20.11 (SD=42.48)	41.92 (SD=135.19)	U=1585.5, p=.261
Fréquence du lemme dans les livres	84.43 (SD=135.02)	121.40 (SD=210.05)	U=1531.0, p=.158
Fréquence du mot dans les livres	23.14 (SD=38.77)	28.73 (SD=58.15)	U=1602.0, p=.299
Nombre de voisins orthographiques	3.65 (SD=3.51)	2.53 (SD=2.54)	U=2084.0, p=.132
Nombre de voisins phonologiques	8.00 (SD=7.42)	5.22 (SD=4.87)	U=2127.5, p=.084
Nombre de syllabes	2.25 (SD=0.75)	2.65 (SD=0.73)	U=1280.0, p=.003
Nombre de lettres	7.05 (SD=1.88)	7.88 (SD=1.91)	U=1323.5, p=.011
Nombre de phonèmes	5.17 (SD=1.61)	6.00 (SD=1.64)	U=1234.0, p=.002
Cloze-probability	50.00 (SD=11.06)	50.00 (SD=11.06)	U=1788.0, p=.951
Types de mots			
Noms	3 (5%)		
Verbes	18 (30%)		
Adjectifs	39 (64%)		

Note. SD = écart-type. Les valeurs de U et de p proviennent des tests de Mann-Whitney comparant les groupes de mots incongruent et congruent. Les fréquences lexicales renvoient à la fréquence d'usage des mots dans la langue française : les mots les plus fréquents ont été privilégiés afin de favoriser leur reconnaissance et compréhension. Les mots dont le lemme est courant dans les films et les livres ont été retenus, car ils sont susceptibles d'être plus familiers pour les participants. Les mots comportant moins de syllabes ont également été choisis, car ils sont généralement plus faciles et rapides à comprendre. Enfin, les mots ayant peu de voisins orthographiques ou phonologiques (c'est-à-dire, de mots similaires en orthographe ou en prononciation) ont été privilégiés afin de réduire les confusions potentielles.

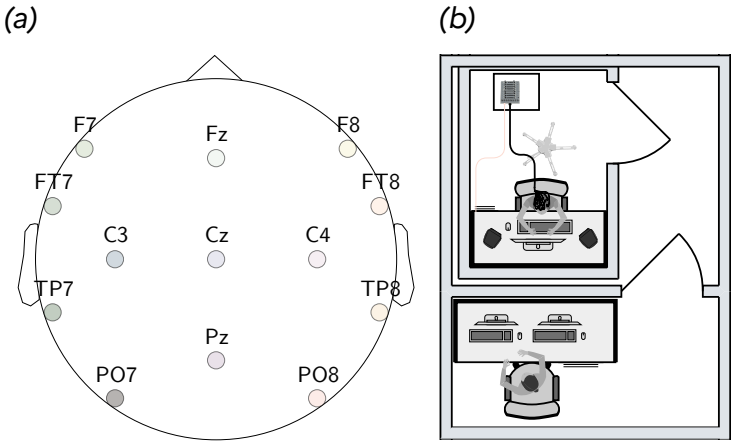
ment susciter de façon involontaire la composante N400 à cause d'éléments parasites dans la voix. De plus, cette approche a permis de contrôler avec précision la temporalité et la présentation des stimuli verbaux, ce qui est crucial pour l'EEG. Les audios ont été édités pour supprimer les éventuels bruits de fond et les silences trop longs.

En ce qui concerne la création des stimuli vidéo, il a tout d'abord fallu optimiser les enregistrements pour ne pas filmer autant de vidéos qu'il y a de versions de phrases. Pour ce faire, une série de treize vidéos a été enregistrée, dans lesquelles le robot parle (c'est-à-dire fait des mouvements de bouche) pendant des durées variables allant de 2 à 8 s, par incréments de 2, 2.5, 3 s, etc. Ceci avait pour but de s'assurer que les audios des phrases, enregistrés séparément, puissent être synchronisés ensuite avec les mouvements de la bouche du robot. Les vidéos ont été capturées en 4K à 50 i/s à l'aide d'un *iPhone 13 Pro* monté sur un trépied. Le robot était placé sur une table avec un fond noir neutre. L'export final étant en 1080p, la résolution 4K a permis de créer les vidéos pour les deux conditions avec un seul recadrage sans perdre de qualité. Pour le montage vidéo, l'audio et la vidéo ont été synchronisés à l'aide du logiciel *Adobe Premiere Pro* et exportés en 1080p à 50 i/s. Chaque vidéo était structurée comme suit : la vidéo commençait par un silence d'une seconde, puis le robot parlait et la vidéo se terminait par un silence d'une seconde.

Pour ce qui est du matériel nécessaire à l'enregistrement des signaux EEG, il comprend un système *Biosemi ActiveTwo* à soixante-quatre électrodes (*ElectroCap Inc*, système international 10-20). Du gel a été appliqué sur chaque endroit où étaient posées les électrodes afin d'assurer un contact entre l'électrode et le cuir chevelu. Le décalage (*offset*) des électrodes a été contrôlé pour rester dans la plage de -20 mV à $+20$ mV tout au long de l'expérience. Pour suivre les artefacts (liés aux yeux ou aux muscles de la mâchoire), trois électrodes supplémentaires ont été placées : deux près des mastoïdes et une sous l'œil gauche. La configuration de la salle d'expérimentation est présentée en Figure 9.2b.

Figure 9.2

Électrodes d'intérêt et configuration de salle d'expérimentation



Note. **(a)** Les 13 électrodes d'intérêt ; **(b)** Configuration de salle EEG d'expérimentation.

9.2.4 Procédure

Vérifications technique et consentement

Les participants ont d'abord reçu une feuille d'information et ont signé le formulaire de consentement éclairé conformément aux directives éthiques en vigueur. Ensuite, ils ont été invités à se rendre dans la salle d'expérimentation. Avant le début de l'enregistrement EEG, une photo de la tête du robot leur a été présentée, accompagnée d'une question sur leur familiarité avec ce robot. Les participants ayant déjà rencontré le robot ont été automatiquement assignés à la condition *BODY* afin d'éviter que leurs connaissances préalables n'influencent les données. Cette procédure explique le nombre légèrement plus élevé de participants dans la condition *BODY* par rapport à la condition *HEAD*.

Après la mise en place du casque EEG, les participants ont été laissés seuls dans la salle d'expérimentation et invités à rester aussi immobiles que possible pendant que le robot parlait, afin de minimiser les artéfacts musculaires liés aux mouvements de la mâchoire ou aux clignements des yeux.

Phase d'introduction

L'expérience commençait par une vidéo d'instructions de 20 s, délivrée par le robot, reprenant les consignes déjà données oralement par l'expérimentateur selon le script suivant :

> « *Bonjour, je m'appelle Lou. Merci pour votre participation. Je vais vous parler pendant que vous portez un casque qui mesure l'activité de votre cerveau. À partir de chaque bip sonore, merci de rester le plus immobile possible et de ne pas cligner des yeux. À chaque fois que j'arrête de parler, vous pouvez à nouveau cligner des yeux lorsque l'écran est noir.* »

Selon la condition assignée, le robot était montré soit dans son intégralité (condition *BODY*), soit de façon à ne rendre visible que sa tête (condition *HEAD*). Pendant cette vidéo d'instruction, un signal sonore (bip) pouvait être entendu au moment où le robot prononçait le mot « bip », servant d'exemple pratique pour les participants. Ce bip d'avertissement durant 300 ms était une note sinusoïdale à fréquence fondamentale de 300 Hz. Suite à ces instructions, un texte apparaît sur l'écran, informant le participant qu'il peut commencer l'expérience en appuyant sur la barre d'espace.

Phase principale

En appuyant sur la barre d'espace, l'expérience commençait avec la présentation de la liste assignée de soixante vidéos dans lesquelles le robot s'exprimait. La durée totale du visionnage était d'environ 12 min. Pour chaque vidéo, la séquence était la suivante : Un écran noir de 1000 ms avec une croix de fixation accompagnée d'un bip sonore durant 300 ms, la vidéo du robot qui s'exprimait, puis, finalement un écran noir de 2500 ms.

Les vidéos où le robot parlait ont été présentées dans un ordre aléatoire différent pour chaque participant, tout en veillant à ce qu'il n'y ait pas plus de trois vidéos consécutives du même type (incongruentes ou congruentes).

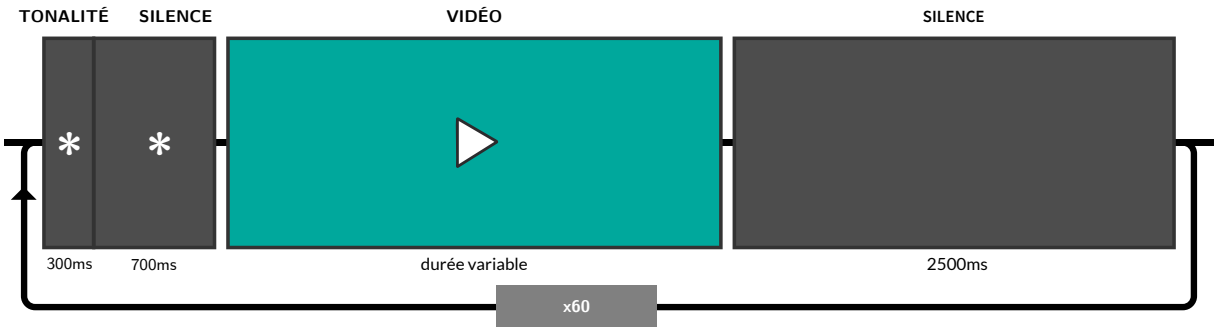
Au niveau de la synchronisation avec le système EEG, à chaque étape de la séquence, un déclencheur (trigger) était envoyé au système d'acquisition EEG afin que les données puissent être correctement synchronisées avec les vidéos et permettre une analyse ultérieure des données : un déclencheur (différent en fonction de si le mot cible était congruent ou incongruent) était envoyé durant

le premier écran noir de 1000 ms, au début de la vidéo, au moment du mot cible et enfin, durant le dernier écran noir de 2500 ms après la vidéo.

Lors de l'analyse des données, seuls les déclencheurs de mots cibles ont été utilisés. Ils permettent de diviser les données EEG en segments (*epochs*) pour chaque vidéo, afin que les données puissent être analysées en fonction de la condition (congruente ou incongruente). La Figure 9.3 illustre une séquence pour un essai de l'expérience.

Figure 9.3

Schéma d'une séquence pour un essai de l'expérience



Phase de questionnaire

À la suite du visionnage de l'ensemble des vidéos, les perceptions des participants concernant le robot ont été évaluées à l'aide d'une version que nous avons traduite du questionnaire de Ho et MacDorman (2010). Ce choix a été motivé par la distinction claire qu'il établit entre les dimensions d'*Humanité*, d'*Étrangeté* et d'*Attrait* : chacune de ces trois dimensions comprend de nombreux éléments qui reflètent divers aspects de la perception humaine et des réactions émotionnelles, permettant ainsi une mesure approfondie des attitudes envers le robot.

Il a également été utilisé cinq items supplémentaires testés par Nazir et al. (2023) pour déterminer comment les participants percevaient le robot. Les affirmations testées se concentraient sur les qualités suivantes :

— *Imagination* : Lou peut imaginer et inventer à partir de ses expériences.

— *Intelligence* : Lou sait s'adapter à son environnement et peut interagir avec les autres

— *Indépendance* : Lou est autonome et ne dépend pas des autres.

— *Créativité* : Lou a la capacité de trouver des solutions originales au-delà de ses expériences et peut créer de nouvelles choses.

— *Bavard* : Lou parle beaucoup et aime beaucoup parler.

Ces cinq traits ont été sélectionnés sur la base des travaux de Haslam et al. (2004) sur les croyances essentialistes concernant la personnalité humaine. Les croyances essentialistes font référence à la pratique consistant à considérer un trait comme inné et biologique, et non acquis (voir par exemple Gelman, 2003). Dans les travaux de Haslam et al. (2004), les traits sont dits « essentialisés » lorsqu'ils sont perçus comme des aspects centraux et stables de la nature humaine.

Les participants ont répondu à ces affirmations en positionnant le curseur de l'ordinateur sur une échelle allant de 0 (signifiant « Pas du tout d'accord ») à 100 (« Tout à fait d'accord »). Les chiffres n'étaient pas visibles pour les participants. Enfin, en utilisant la même échelle de 0 à 100, les participants ont également été invités à répondre à deux affirmations portant sur le fait que le robot pourrait avoir des bras ou des jambes cachés. Les affirmations étaient : « *Lou possède des bras cachés* » et « *Lou possède des jambes cachées* ». Cela permettait ainsi d'obtenir une estimation de leurs croyances en des capacités cachées que le robot n'a (visiblement) pas et ainsi pouvoir faire le lien avec une possible atténuation de la N400 dans le cas où ils croient en cette possibilité.

9.2.5 Analyse des données

Le signal EEG continu a été enregistré avec une fréquence d'échantillonnage de 2048 Hz. Le signal a été filtré dans la bande des [0.5, 30] Hz. Ensuite, il a été sous-échantillonné à une résolution de 200 Hz. Après une Analyse en Composantes Indépendantes (*Independent Component Analysis; ICA*) avec l'algorithme *AMICA* (Palmer et al., 2008), les artéfacts tels que les clignements des yeux, des muscles et des battements de cœur ont pu être identifiés puis retirés des données.

Une fois nettoyées, les données EEG, ont été séparées en *epochs* qui commençaient -150 ms avant le mot cible et se terminaient 1200 ms après le début

du mot cible (*onset*). Ces 150 ms avant l'apparition du mot cible ont été utilisées comme référence (*baseline*) pour l'analyse des ERP. Les étapes de traitement des données ont été réalisées en utilisant *EEGLab* et *MNE-Python*.

Pour l'analyse des ERP, un LMM a été choisi en raison de la non-normalité des distributions et de l'hétérogénéité des variances, ce qui permet de tenir compte de la variabilité interindividuelle et d'assurer des estimations robustes face à ces irrégularités. En raison de la petite taille de certains sous-échantillons après l'analyse principale, pour une évaluation nuancée des effets intra-sujets et des interactions, une ANOVA à mesures répétées a été utilisée comme analyse complémentaire suivant le LMM. Pour le questionnaire, lorsque les données ne suivaient pas une distribution normale, le test de Mann-Whitney a été utilisé. Dans certains cas où les données étaient normalement distribuées et avaient des variances égales, confirmées par le test de Levene, un test t de Student pour échantillons indépendants a été utilisé.

9.3 Résultats

9.3.1 Analyse des ERP

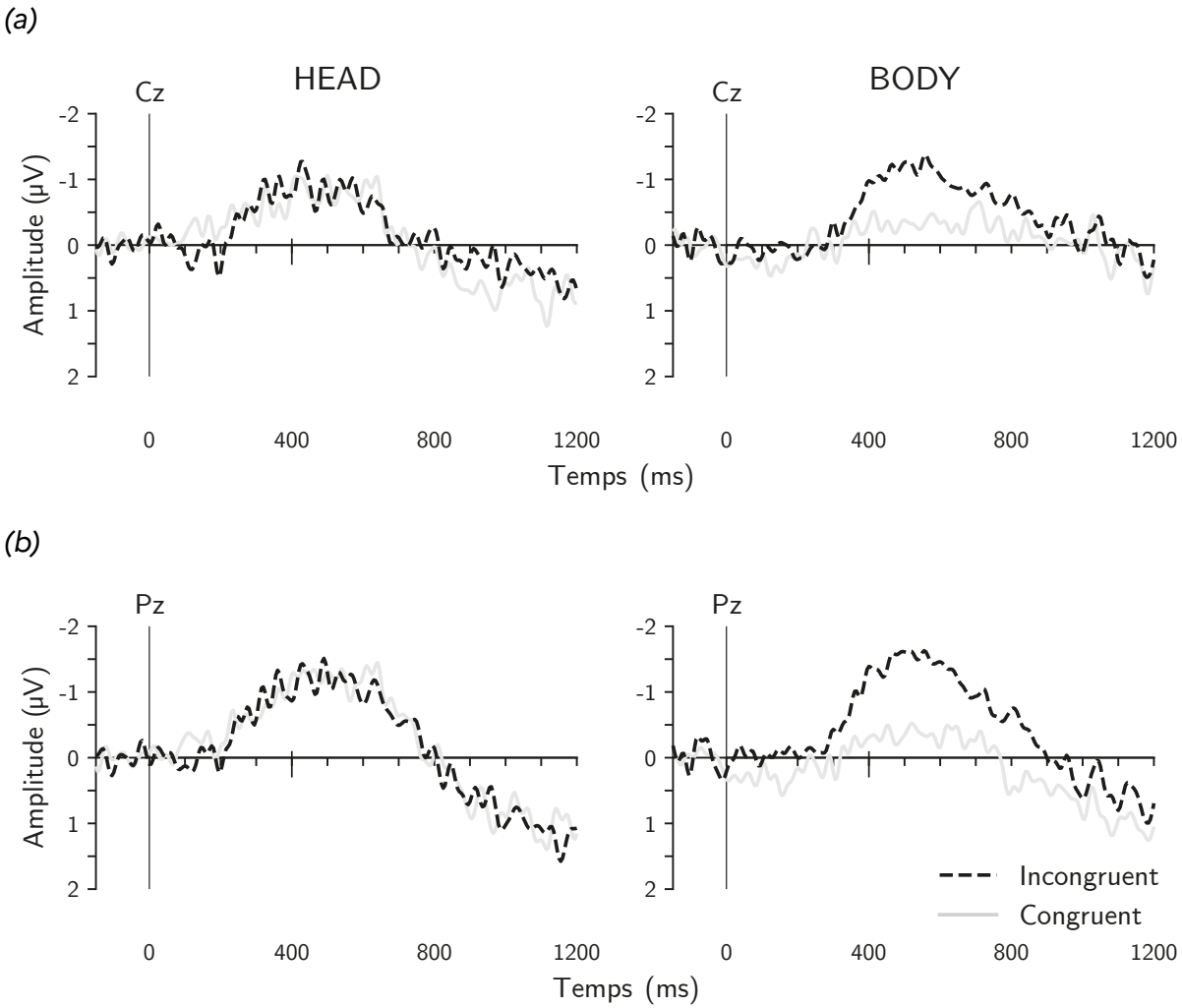
Analyses préliminaires

Étant donné que la condition *HEAD* a servi de contrôle pour la qualité de notre matériel de phrases, une analyse préliminaire avait été conduite sur cette condition. Pour illustration, le panneau de gauche de la Figure 9.4 présente les ERP sur les électrodes Cz et Pz pour les deux types de mots cibles (congruent et incongruent) dans la condition *HEAD*. Aucune différence discernable dans les données n'a été observée entre les deux types de mots cibles, indiquant que les deux conditions de congruence (lorsque prononcés par un robot dont on ne peut voir que le visage) n'ont pas affecté différemment les ERP des participants, confirmant ainsi que les stimuli étaient bien construits et équilibrés.

Cependant, comme prévu concernant le *UP*, le pic de la N400 est survenu plus de 100 ms plus tard que ce qui est typiquement observé dans les expériences standards en N400. Afin de vérifier si ce délai est effectivement lié au *UP* phonologique tardif dans les mots cibles, nous avons sélectionné parmi les cent-vingt mots cibles ceux avec un *UP* à la troisième ou quatrième lettre du mot.

Figure 9.4

Tracés des potentiels évoqués des deux conditions aux électrodes Cz et Pz



Note. Les tracés représentent les ERP sur les électrodes Cz et Pz pour les deux types de mots cible (congruent et incongruent) dans la condition *HEAD* et *BODY*.

Ensuite, nous les avons contrastés avec des mots cibles dont le *UP* se produit à la sixième ou septième lettre. Conformément à notre hypothèse, la latence du pic du composant ERP négatif a été affectée par l'*UP*. Plus l'*UP* était tard dans un mot, plus la latence moyenne des pics N400 était retardée. Pour les mots avec un *UP* à la sixième ou septième lettre, la latence moyenne du pic N400 était de 656 ms ($M_{\text{ampl.}} = -3.878\ \mu V$), tandis que pour les mots avec un *UP* à la troisième ou quatrième lettre, le pic était autour de ≈573 ms ($M_{\text{ampl.}} = -2.768\ \mu V$). Un test de Mann-Whitney a révélé que cette différence était significative ($U = 36.0$, $p = .013$). Par conséquent, nous pouvons attribuer avec confiance ce décalage de délai observé aux caractéristiques spécifiques de nos stimuli, plutôt qu'à un mécanisme cognitif sous-jacent alternatif du composant ERP. Il est à noter qu'un délai similaire a été observé dans une étude de van Berkum et al. (2003), qui a également utilisé des mots cibles parlés plutôt longs.

Analyses principales

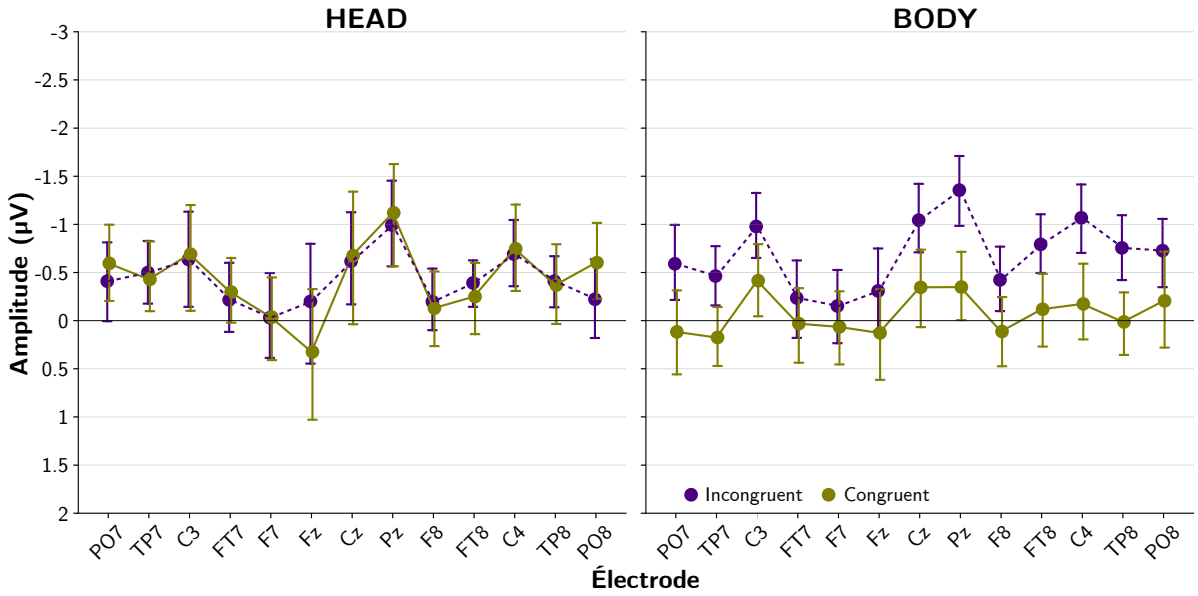
Le panneau de droite de la Figure 9.4 présente les ERP aux électrodes Cz et Pz pour les deux types de phrases dans la condition *BODY*. Un effet N400 net est observé, avec des amplitudes plus élevées pour les phrases se terminant par des mots cibles incongruents avec les caractéristiques physiques du robot.

Étant donné les analyses préliminaires, il a été décidé de ne pas utiliser la plage de latence N400 standard de 300 à 500 ms après le début du mot cible pour le calcul des valeurs d'amplitude moyenne (calculées pour chaque sujet et condition). Au lieu de cela, et sur la base de l'approche utilisée par van Berkum et al. (2003), nous avons choisi une fenêtre temporelle d'intérêt allant de 500 à 700 ms après le début du stimulus en nous concentrant sur treize électrodes d'intérêt (voir Figure 9.2a).

Comme prévu, dans la condition *BODY*, un effet N400 clair a été observé pour toutes les électrodes d'intérêt. Les phrases se terminant par un mot incongruent avec les capacités physiques du robot ont provoqué une déflexion négative plus prononcée par rapport à celles qui étaient congruentes. Dans la condition *HEAD*, servant de contrôle, un tel effet n'a pas été observé. Les amplitudes moyennes sur la fenêtre temporelle de 500 à 700 ms aux treize électrodes sont présentées, pour les deux conditions, dans la Figure 9.5.

Figure 9.5

Amplitudes moyennes pour les treize électrodes d'intérêt



Note. Amplitudes moyennes sur les treize électrodes d'intérêt au cours de la fenêtre de 500 à 700 ms après le début du stimulus. Les barres d'erreur correspondent à des intervalles de confiance à 95%. **(Gauche)** : résultats pour la condition *HEAD*. **(Droite)** : la condition *BODY*.

Les données ont été analysées via un LMM avec comme variables indépendantes (effets fixes) la congruence (congruent vs. incongruent) ainsi que le groupe (*BODY* vs. *HEAD*), tout comme leur interaction. Les participants ont été traités comme des effets aléatoires et la variation aléatoire dans la réponse à la congruence a été modélisée pour chaque participant. Cette approche nous a permis de traiter les participants comme des effets aléatoires pour tenir compte de la variabilité de base et des différences individuelles.

Une interaction significative entre la congruence et le groupe a été trouvée (b = 0.617, SE $= 0.225$, $z = 2.741$, $p < .01$, $IC_{95\%}$ = $[0.176,\ 1.059]$), indiquant que l'influence de la congruence sur l'amplitude N400 était liée au groupe auquel les participants étaient assignés. Plus précisément, cette interaction suggère que la différence d'amplitude N400 en réponse aux stimuli incongruents par rapport aux stimuli congruents est moins prononcée chez les participants qui ont vu la tête du robot seule (*HEAD*), par rapport à ceux qui ont vu le corps entier (*BODY*). La variance inter-participants était de 0.599 (SE$= 0.143$), reflétant la variabilité des réponses entre les participants.

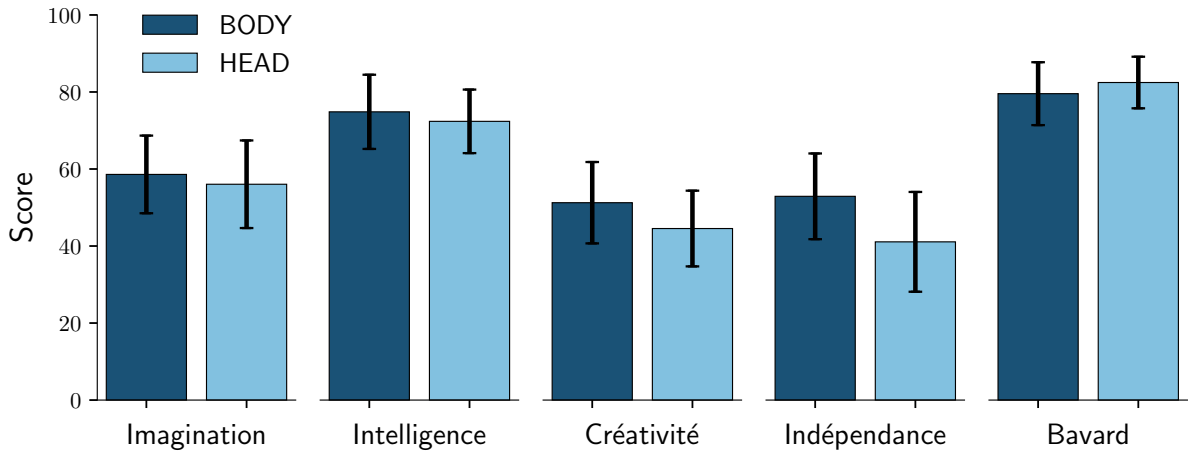
9.3.2 Analyse du questionnaire

La Figure 9.6 présente les scores pour les cinq affirmations qui exploraient les perceptions des participants à l'égard du robot (*Imagination, Intelligence, Créativité, Indépendance* et *Bavard*) avec leurs intervalles de confiance de 95% respectifs. Un test *U* de Mann-Whitney a montré qu'aucune des cinq déclarations ne distinguait les participants dans les conditions *BODY* et *HEAD*. Un test *t* de Student bilatéral pour deux échantillons n'a également révélé aucune différence significative entre les scores composites moyens aux cinq énoncés dans les deux groupes ($t(54)= 0.836$, $p = .406$). Ces résultats suggèrent que la perception du robot, en se basant sur les descripteurs de personnalité humaine essentialisés, ne différait pas entre les deux groupes.

Concernant le questionnaire de Ho et MacDorman (2010), une ANOVA mixte avec la dimension (*Humanité, Attrait, Étrangeté*) et le groupe (*BODY* vs. *HEAD*) comme facteurs intra-sujet et inter-sujet, a montré un effet significatif de la dimension ($F(2, 108)= 136.14$, $p < .001$, $\eta_p^2 = .716$). Cependant, l'effet du groupe n'était pas significatif ($F(1, 54)= 3.58$, $p = .06$, $\eta_p^2 = .062$), tout comme l'interaction entre le groupe et la dimension ($F(2, 108)= 2.05$, $p = .13$, $\eta_p^2 = .037$).

Figure 9.6

Scores aux cinq affirmations explorant les perceptions à l'égard du robot



Note. Valeurs moyennes pour les cinq affirmations qui sondaient les perceptions des participants à l'égard du robot, avec les intervalles de confiance à 95% correspondants.

Le Tableau 9.2 donne les scores composites pour les trois dimensions d'*Humanité*, d'*Attrait* et d'*Étrangeté*. La Figure 9.7 présente des diagrammes en boîte comparant les perceptions de ces trois dimensions dans les deux conditions.

Table 9.2

Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon la condition

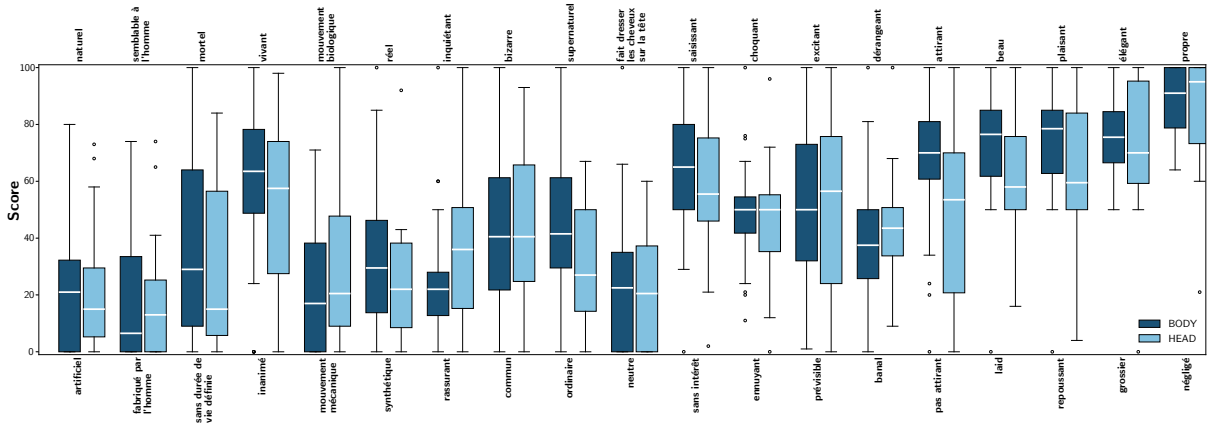
Condition	Humanité	Étrangeté	Attrait
HEAD	$M = 28.81$, $SD = 14.98$	$M = 41.96$, $SD = 11.19$	$M = 65.47$, $SD = 18.88$
BODY	$M = 32.25$, $SD = 14.39$	$M = 42.48$, $SD = 13.71$	$M = 76.11$, $SD = 13.75$

Note. Les valeurs (de 0 à 100) indiquent les moyennes (M) et écarts-types (SD) pour chaque condition sur les dimensions du questionnaire de Ho et MacDorman (2010).

Enfin, le Tableau 9.3 résume les évaluations des croyances des participants quant à la possibilité que le robot ait des bras ou des jambes cachés. Malgré des preuves du contraire, de nombreux participants dans la condition *BODY* ont évalué cette possibilité comme supérieure à 0.

Figure 9.7

Scores au questionnaire de Ho et MacDorman (2010) selon la condition



Note. Diagramme en boîte présentant les perceptions des participants à l'égard du robot dans les items du questionnaire Ho et MacDorman (2010), avec les intervalles de confiance à 95% correspondants.

Table 9.3

Scores relatifs à la perception de membres cachés

Condition	Bras	Jambes	Bras et jambes
BODY	M = 31.81, SD = 35.78, Mdn = 18.00, min = 0, max = 100	M = 20.81, SD = 27.27, Mdn = 7.00, min = 0, max = 80	M = 26.31, SD = 28.93, Mdn = 19.50, min = 0, max = 86
HEAD	M = 55.00, SD = 37.63, Mdn = 63.50, min = 0, max = 100	M = 41.08, SD = 34.60, Mdn = 42.50, min = 0, max = 100	M = 48.04, SD = 33.27, Mdn = 50.00, min = 0, max = 100

Note. Les valeurs indiquent les moyennes (M), écarts-types (SD), médianes (Mdn) ainsi que les valeurs minimales et maximales (min, max) des scores concernant la possibilité que le robot possède des bras ou des jambes cachés.

9.3.3 Impact des croyances

Un second volet des résultats concerne l'impact des croyances des participants dans la condition *BODY*. Il avait été investigué la croyance en l'existence de membres cachés chez le robot. Nous faisions l'hypothèse qu'elle devrait atténuer l'incongruité perçue entre son apparence et sa parole, réduisant ainsi l'effet N400. Dans de tels cas, en raison de la discordance réduite, la perception globale du robot devrait être plus positive.

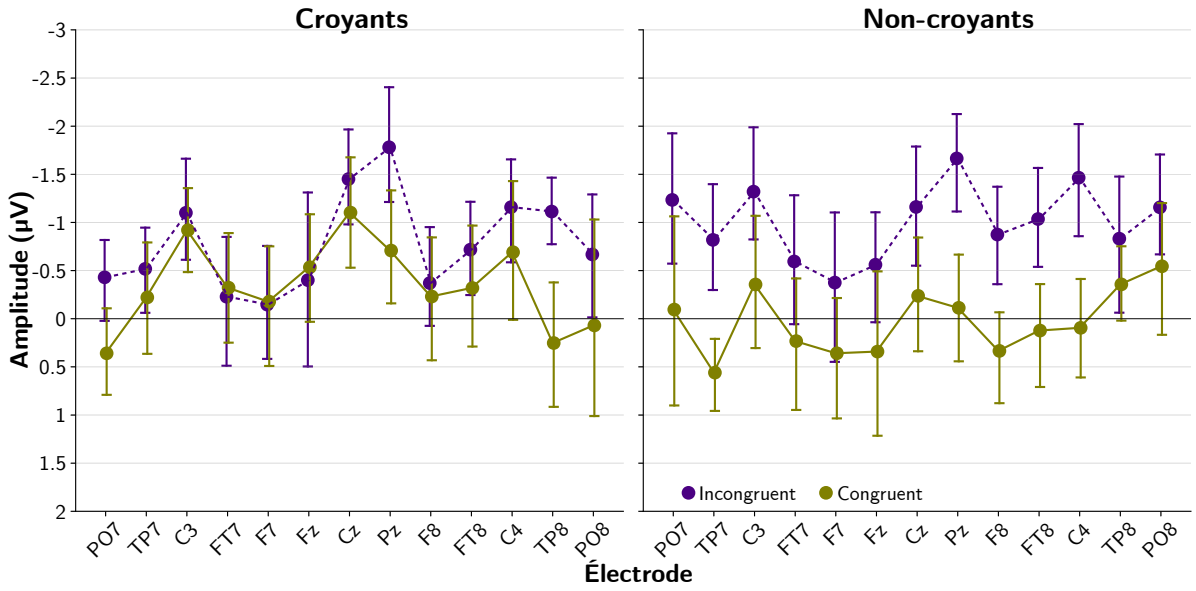
Les scores de croyance en la présence de membres cachés ont varié considérablement entre participants, confirmant que malgré l'évidence visuelle, un grand nombre de personnes ont attribué une probabilité non nulle à l'existence de bras ou jambes cachés. En effet, sur un total de trente-deux participants dans la condition *BODY*, seulement douze participants ont affirmé sans équivoque que le robot n'avait pas de bras ou de jambes cachés, attribuant un score d'évaluation de 0. Les vingt participants restants ont donné des évaluations variant de 5 à 86, suggérant qu'ils envisageaient la possibilité que le robot puisse avoir des membres cachés.

ERP et croyance

Pour évaluer l'impact potentiel des croyances des participants sur les ERP, nous avons contrasté les données des douze participants qui ne croyaient pas aux membres cachés du robot, c'est-à-dire ceux ayant répondu 0, que nous nommerons les « *non-croyants* », avec celles des douze participants qui pensaient fortement que le robot pouvait avoir des membres cachés, nommés ici « *croyants* », ayant répondu entre 33 et 86 ($\mathrm{M} = 59.5$, $\mathrm{Mdn} = 60.5$). La Figure 9.8 présente les amplitudes moyennes des ERP pour les phrases congruentes et incongruentes sur une fenêtre de temps de 500-700 ms aux treize électrodes. Comme le montre la figure, le groupe de participants qui excluait catégoriquement la possibilité que le robot puisse avoir des membres cachés présentait une amplitude N400 significativement plus élevée en réponse aux phrases qui étaient en contradiction avec l'apparence du robot par rapport aux phrases qui étaient cohérentes. À l'inverse, les participants qui envisageaient la possibilité de membres cachés présentaient une différence moins importante dans les amplitudes des ERP entre les deux types de phrases.

Figure 9.8

Amplitudes moyennes des ERP pour les treize électrodes d'intérêt selon la croyance



Note. Amplitudes moyennes des ERP sur une fenêtre temporelle de 500 à 700 ms après le début du stimulus pour les treize électrodes d'intérêt avec leur intervalles de confiance à 95% correspondant. **(Gauche)** : participants qui ne croyaient pas que le robot possède des bras ou des jambes cachés. **(Droite)** : participants qui considéraient cette possibilité.

Un LMM avec le sous-groupe de croyance (croyants vs. non-croyants) et la congruence (congruent vs. incongruent) comme effets fixes a été utilisé pour analyser les données. Les participants ont été traités comme des effets aléatoires et la variation aléatoire dans la réponse de congruence a été modélisée pour chaque participant. Le modèle n'a pas révélé d'interaction significative entre la congruence et la croyance, $b = 0.61$, SE $= 0.374$, $z = 1.625$, $p = .104$, $IC_{95\%} = [-0.13,\ 1.34]$, probablement en raison du nombre limité de participants dans chaque groupe. La variance du groupe était de 0.535 (SE $= 0.193$).

Cependant, lorsque des ANOVA à mesures répétées distinctes ont été effectuées pour chaque groupe, un effet principal très significatif de la congruence a été observé pour les « non-croyants » ($F(1, 11) = 12.192$, $p = .005$), tandis qu'aucune différence significative n'a été observée pour les « croyants » ($F(1, 11) = 3.478$, $p = .089$). Nous envisageons cependant avec prudence la possibilité que les croyances des participants concernant les membres cachés modulent réellement la force de l'effet N400 sur l'incongruence entre les capacités physiques d'un robot et ses déclarations.

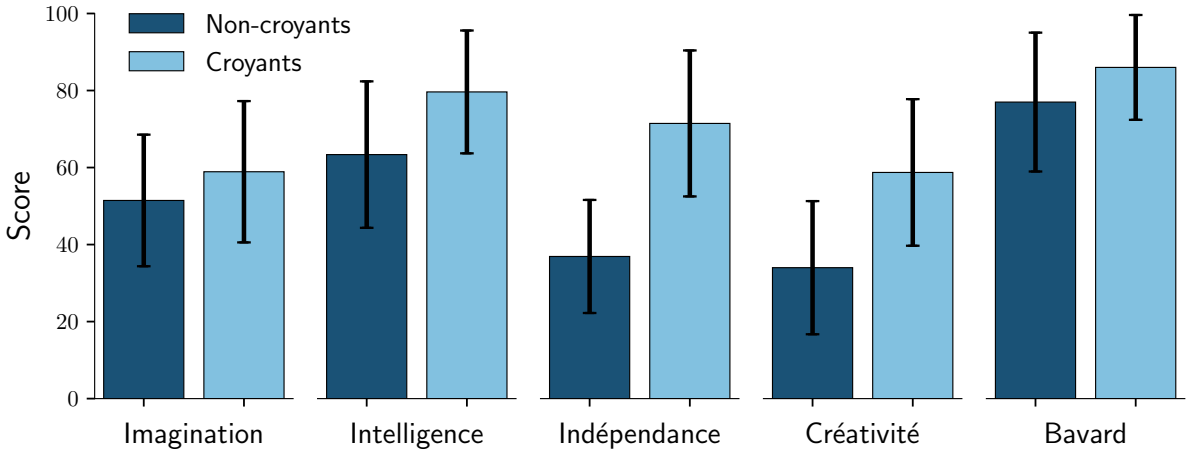
Questionnaire et croyance

La Figure 9.9 compare les scores du groupe des « croyants » avec ceux des « non-croyants » pour les cinq affirmations qui exploraient les perceptions des participants à l'égard du robot (*Imagination, Intelligence, Créativité, Indépendance* et *Bavard*), avec les intervalles de confiance à 95% correspondants.

Pour toutes les affirmations, des scores plus élevés ont été observés dans le groupe des « croyants », confirmant l'hypothèse selon laquelle la discordance perçue entre les capacités du robot et son discours pouvait déclencher une perception plus négative envers le robot. Un test *t* de Student unilatéral pour deux échantillons indépendants a révélé une différence significative dans les scores moyens composites ($t(22) = 1.865$, $p = .037$). Les participants du groupe des « non-croyants » ont généralement obtenu des scores plus faibles (M $= 54.766$, SD $= 19.066$) que ceux du groupe des « croyants » (M $= 70.216$, SD $= 21.441$). Ce résultat indique que les personnes qui restaient ouvertes à la possibilité de membres cachés ont donné une évaluation générale plus positive du robot.

Figure 9.9

Scores pour les cinq traits explorant les perceptions à l'égard du robot selon la croyance



Note. Scores moyens et intervalles de confiance à 95% correspondants dans les sous-groupes de croyance pour les cinq affirmations qui exploraient les perceptions du robot

Concernant le questionnaire de Ho et MacDorman (2010), pour chaque dimension les tests *U* de Mann-Whitney n'ont révélé aucune différence significative entre les deux sous-groupes de croyance, suggérant que les croyances concernant la présence de membres cachés n'influencent pas significativement la perception globale du robot sur ces dimensions. Le Tableau 9.4 donne les scores moyens pour les dimensions de *Humanité*, *Attrait* et *Étrangeté* tandis que la Figure 9.10 présente des diagrammes en boîte de ces scores.

Table 9.4

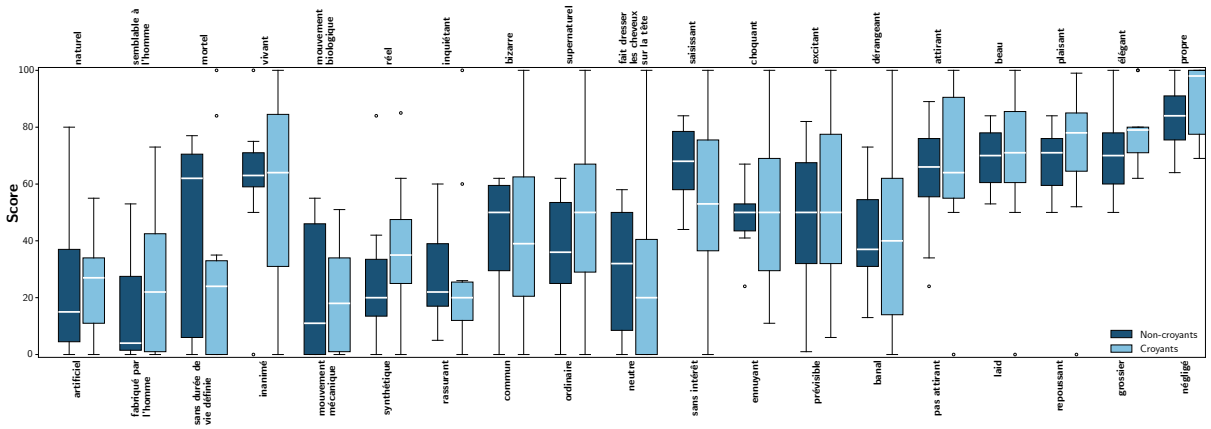
Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon le sous-groupe de croyance

Groupe	Humanité	Étrangeté	Attrait
Non-croyants	M = 29.65, SD = 14.25	M = 42.93, SD = 9.13	M = 71.98, SD = 11.90
Croyants	M = 32.71, SD = 14.90	M = 43.66, SD = 19.68	M = 74.95, SD = 15.26

Note. Les valeurs indiquent les moyennes (M) et écarts-types (SD) pour les dimensions du questionnaire de Ho et MacDorman (2010) dans les sous-groupes selon la croyance.

Figure 9.10

Scores selon le sous-groupe de croyance aux dimensions du questionnaire de Ho et MacDorman (2010)



Note. Diagramme en boîte présentant les perceptions des participants à l'égard du robot selon le sous-groupe de croyance dans les items du questionnaire Ho et MacDorman (2010), avec les intervalles de confiance à 95% correspondants.

9.4 Discussion partielle

Les résultats obtenus apportent un éclairage sur les limites implicites que les humains posent dans le discours des robots, ainsi que sur la flexibilité de ces limites en fonction des croyances.

Tout d'abord, cette première étude démontre qu'un décalage perçu entre les capacités physiques d'un robot et ses déclarations peut déclencher la composante associée au traitement de l'incongruité sémantique (Kutas & Hillyard, 1980 ; van Berkum et al., 2008) qu'est la N400. Plus précisément, lorsque les participants voyaient le corps entier du robot, les phrases contenant des actions impossibles à réaliser pour le robot provoquaient une plus grande déflexion N400 que les phrases contenant des actions lui étant physiquement possibles. Tandis que lorsque les participants ne voyaient que la tête du robot, ce phénomène était neutralisé puisque les éléments créant l'incongruence (c'est-à-dire les membres manquants) n'étaient pas visibles.

Cette absence d'effets dans la condition où seule la tête est visible suggère que nos phrases *congruentes* et *incongruentes* étaient équivalentes, malgré cer-

tains déséquilibres dans les variables linguistiques. Les résultats confirment donc que la N400 reflète l'intégration d'informations multimodales, y compris des indices visuels et contextuels, dans l'évaluation sémantique du discours.

De même, l'observation d'un effet N400 dans la condition où le corps entier est visible suggère que le cerveau humain exerce probablement une forme de vigilance ou de scepticisme vis-à-vis des affirmations d'un agent artificiel lorsque celles-ci entrent en conflit avec les indices visuels disponibles. Ainsi, lorsqu'un robot sans bras déclare avoir effectué une action nécessitant des bras, les participants perçoivent cette déclaration comme incongruente. L'effet N400 dans cette condition peut donc être attribué à cette incongruence perçue entre le discours du robot et son apparence physique.

Concernant les croyances individuelles, un grand nombre de participants, si ce n'est la majorité, s'est montré disposée à imaginer que le robot puisse posséder des membres cachés. L'analyse des données issues de la condition où le corps entier était visible indique une potentielle modulation de l'effet N400 par les croyances individuelles des participants concernant cette possibilité. Les participants qui envisageaient la possibilité de tels membres cachés ont montré un effet N400 réduit pour les phrases incongruentes avec l'apparence du robot. Cela indique qu'ils trouvaient ces phrases plus plausibles compte tenu de leurs croyances. De plus, ces croyances semblaient également influencer la perception générale du robot, comme le montrent les évaluations aux cinq descripteurs (*Imagination*, *Intelligence*, *Créativité*, *Indépendance* et *Bavard*). Les participants identifiés comme « non-croyants » étaient moins susceptibles d'attribuer des qualités humaines au robot.

En revanche, le questionnaire de Ho et MacDorman (2010) ne permettait pas de faire la distinction entre les groupes de participants. Dans l'ensemble, la modulation de la N400 en fonction des croyances individuelles suggère que les modèles mentaux des participants affectent le traitement des informations multimodales. Cette flexibilité souligne le potentiel de notre paradigme N400 pour examiner comment les humains traitent les énoncés dans lesquels un robot fait référence à ses propres états internes, notamment émotionnels.

En somme, cette étude indique que face au discours d'un robot, la présence d'indices clairs d'incongruités (comme l'absence visible de bras ou de jambes) active chez l'humain des processus cognitifs traduisant une détection d'ano-

malie et d'incongruence et, potentiellement, une forme de scepticisme ou un garde-fou épistémique. Toutefois, l'imagination peut repousser cette frontière en introduisant l'hypothèse de capacités cachées, diminuant ainsi le poids de incongruence perçue. Une telle flexibilité cognitive où la croyance peut aller jusqu'à moduler la réponse N400 souligne le potentiel de notre paradigme N400 pour examiner comment les humains traitent les énoncés dans lesquels un robot fait référence à ses propres états internes, notamment émotionnels. Le chapitre suivant traite de cette question (Chapitre 10).

Chapitre 10

Un robot parlant de ses émotions

10.1 Introduction

Les émotions engagent à la fois une expérience interne subjective et divers mécanismes biologiques (Scherer, 2005). Comme nous l'avons vu au Chapitre 2, la tendance humaine à anthropomorphiser pousse à prêter aux robots des intentions ou des états internes (Epley et al., 2007). Néanmoins, les états émotionnels demeurent des qualités que l'être humain associe difficilement aux robots (Fussell et al., 2008 ; Gray et al., 2007 ; Gray & Wegner, 2012).

Ainsi, les humains peuvent-ils accepter qu'un agent artificiel revendique de tels états ? Lorsqu'un robot déclare « *je suis triste* », il se pourrait qu'une telle phrase suscite une réaction similaire à celle observée dans l'étude précédente (Chapitre 9). Considérant que les émotions dépassent la capacité des agents artificiels, cette seconde étude vise à examiner la manière dont les individus réagissent à des phrases prononcées par un robot se terminant soit par une référence à ses émotions, soit par une formulation alternative neutre.

Pour tester cette hypothèse, l'activité cérébrale suscitée par un même énoncé émotionnel est comparée selon qu'il est prononcé par un robot ou par un humain. Si l'effet N400 observé avec le robot provient bien de la non-attribution de la capacité de ressentir des émotions, cet effet devrait disparaître lorsque la phrase est énoncée par un humain. Autrement dit, un humain déclarant lui aussi « *je suis triste* » ne devrait pas provoquer de réponse N400 accrue, car un tel énoncé ne constitue pas, en principe, une incongruité : il correspond à nos attentes.

En somme, on s'attend à observer un effet d'interaction entre le type de locuteur (Robot ou Humain) et le type d'énoncé (émotionnel ou non-émotionnel) sur l'amplitude de la N400. Cet effet indiquerait l'existence d'une limite cogni-

tive implicite : même si, sur le plan déclaratif, les participants pourront admettre que le robot peut ressentir des émotions, leur activité cérébrale révélerait une réserve face à ce type d'affirmation.

10.2 Méthode

Cette étude contraste des vidéos mettant en scène un robot (condition Robot) avec celles mettant en scène un locuteur humain (condition Humain) prononçant des phrases mentionnant des émotions (le mot cible) ou dans leur autre variante, qui n'invoque pas les émotions.

Pour la méthode de cette étude, sauf indication spécifique, l'étude suivait les mêmes principes que l'étude précédente (Chapitre 9) en termes de matériel, de méthodes, d'analyses et de procédures. La population comprenait cinquante locuteurs natifs français (2 personnes non-binaires, 42 femmes et 6 hommes), âgés de 18 à 58 ans (M $= 23.16$, Mdn $= 21$), recrutés selon des critères cohérents avec l'étude précédente. Aucun participant n'avait participé à l'étude précédente.

Bien qu'une voix humaine différente de celle de l'étude précédente ait été utilisée pour le doublage des vidéos, celle-ci était relativement proche en termes de timbre, de rythme et d'intonation. La voix était celle de l'actrice humaine et a été utilisée à la fois dans les conditions « Humain » et « Robot ». Le fait d'avoir exactement le même audio dans les deux conditions nous a permis d'isoler l'effet du type de l'agent (robot ou humain) sur l'attribution de l'émotion, fournissant ainsi une base solide pour évaluer l'impact du type d'agent sans l'influence de la variation vocale.

Les phrases des vidéos ont été conçues pour faire référence à un large éventail de situations et de réactions émotionnelles, reflétant les expériences et les réactions de la vie quotidienne (voir le Tableau de l'Annexe B.2). Les phrases couvraient des sujets tels que des défis inattendus, des interactions sociales et des accomplissements personnels. Par exemple, « *Hier on m'a invité à participer aux tâches ménagères, j'étais* ***ravie*** ». Ces phrases étaient contrastées avec une contrepartie ne faisant pas mention d'émotion. Par exemple, « [...] *j'étais* ***opérationnelle*** ». Les phrases avec un contenu émotionnel incluaient six émotions distinctes avec dix phrases dédiées à chacune (joie, peur, colère, tristesse, dégoût et surprise).

Les caractéristiques des mots cibles sont données dans le Tableau 10.1. Il est à noter que la *cloze probability* pour les deux conditions de phrases était significativement différente. Cependant, étant donné la petite différence totale (0.49 contre 0.51), nous ne nous attendions pas à des écarts substantiels entre les deux conditions en raison de cette variable. Comme dans l'Étude 1, la condition de contrôle (ici Humain) a servi à valider notre matériel de phrases.

Également, comme dans l'Étude 1 (Chapitre 9), nous avons utilisé les cinq traits évalués par Nazir et al. (2023) ainsi que le questionnaire de Ho et MacDorman (2010) pour les deux agents (Robot et Humain). Nous avons cherché à établir une base de référence de ressemblance humaine sur laquelle les agents artificiels pourraient être comparés mais il convient toutefois de noter qu'une application directe du questionnaire de Ho et MacDorman (2010) pour évaluer les humains n'est peut-être pas tout à fait appropriée, étant donné que l'échelle est conçue pour les réponses obtenues face à des entités non humaines. Les résultats concernant l'humain doivent donc être interprétés avec prudence.

Les croyances des participants concernant les capacités émotionnelles du robot (*exprimer* et *ressentir* des émotions) ont été sondées à l'aide de ces deux énoncés :

— Exprimer des émotions : *Lou peut exprimer des émotions*

— Ressentir des émotions : *Lou peut ressentir des émotions*

Table 10.1

Caractéristiques des mots cibles dans l'Étude 2 : mesures linguistiques et phonologiques

	Congruence du mot cible		
Mesure	**Incongruent**	**Congruent**	**Mann–Whitney**
Point d'unicité phonologique	5.983 (SD=1.384)	6.767 (SD=1.760)	U=1296.0, *p*=.007
Fréquence du lemme dans les films	15.853 (SD=41.232)	40.021 (SD=164.579)	U=1526.5, *p*=.151
Fréquence du mot dans les films	6.109 (SD=17.514)	10.004 (SD=37.354)	U=1459.0, *p*=.073
Fréquence du lemme dans les livres	18.304 (SD=30.295)	52.965 (SD=212.000)	U=1948.0, *p*=.438
Fréquence du mot dans les livres	6.205 (SD=13.655)	12.887 (SD=39.819)	U=1789.5, *p*=.958
Nombre de voisins orthographiques	1.483 (SD=1.444)	1.400 (SD=1.976)	U=2018.5, *p*=.235
Nombre de voisins phonologiques	3.300 (SD=3.077)	2.667 (SD=4.620)	U=2293.5, *p*=.008
Nombre de syllabes	2.750 (SD=.680)	2.917 (SD=.907)	U=1637.5, *p*=.356
Nombre de lettres	8.117 (SD=1.574)	8.867 (SD=2.054)	U=1407.5, *p*=.036
Nombre de phonèmes	6.117 (SD=1.354)	7.050 (SD=1.872)	U=1219.5, *p*=.001
Cloze-probability	49 (SD=6.371)	51 (SD=6.575)	U=1425.5, *p*=.047
Types de mots			
Noms	1 (2%)		
Verbes	13 (22%)		
Adjectifs	46 (76%)		

Note. SD = écart-type. Les valeurs de U et de p proviennent des tests de Mann–Whitney comparant les groupes de mots incongruent et congruent. Les fréquences lexicales renvoient à la fréquence d'usage des mots dans la langue française : les mots les plus fréquents ont été privilégiés afin de favoriser leur reconnaissance et compréhension. Les mots dont le lemme est courant dans les films et les livres ont été retenus, car ils sont susceptibles d'être plus familiers pour les participants. Les mots comportant moins de syllabes ont également été choisis, car ils sont généralement plus faciles et rapides à comprendre. Enfin, les mots ayant peu de voisins orthographiques ou phonologiques (c'est-à-dire, de mots similaires en orthographe ou en prononciation) ont été privilégiés afin de réduire les confusions potentielles.

10.3 Résultats

10.3.1 Analyse des ERP

La Figure 10.1 présente les ERP pour les phrases congruentes et incongruentes au niveau des électrodes Cz et Pz pour les deux types de phrases dans la condition Humain et la condition Robot.

La Figure 10.2 présente les amplitudes moyennes des ERP pour les phrases congruentes et incongruentes dans la fenêtre temporelle de 500 à 700 ms sur les treize électrodes d'intérêt.

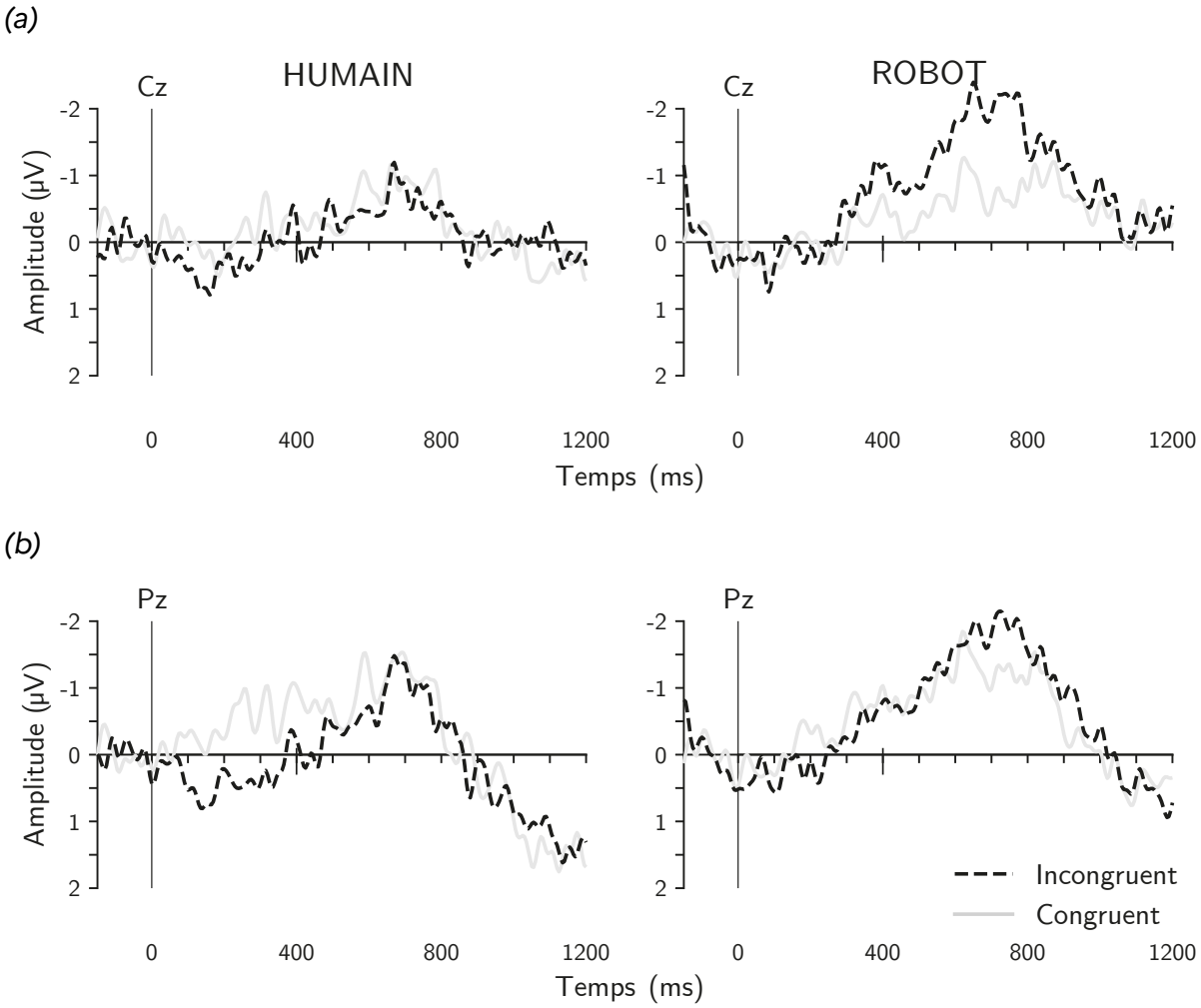
Les résultats ont montré de manière inattendue une légère différence entre les phrases congruentes et les phrases incongruentes dans la condition *Humain* (condition de contrôle). Il se pourrait que cela provienne de certaines variables linguistiques déséquilibrées dans les phrases construites. En effet, les phrases incongruentes ont produit des valeurs légèrement moins négatives que celles congruentes. Ce résultat est à l'inverse de ce que l'on pourrait attendre d'un effet de congruence typique.

Cependant, un tel motif ne devrait pas remettre en cause l'interprétation de nos résultats dans la condition Robot, qui montrent un effet N400 distinct à toutes les électrodes : les phrases se terminant par un mot incongruent avec la capacité du robot à ressentir des émotions ont déclenché une déflexion négative plus prononcée que celles qui étaient congruentes, c'est-à-dire celles ne mentionnant pas les émotions.

Dans le LMM utilisé pour analyser les données, les variables indépendantes (effets fixes) étaient la congruence (Congruent vs. Incongruent) et le groupe (Robot vs. Humain), y compris leur interaction. Les participants ont été traités comme des effets aléatoires, et de plus, la variation aléatoire de la réponse à la congruence a été modélisée pour chaque participant. Le modèle a révélé une interaction significative entre la congruence et le groupe ($b = -0.771$, SE $= 0.309$, $z = -2.491$, $p < .01$, $IC_{95\%} = [-1.377,\ 1.64]$), illustrant que l'effet de la congruence sur l'amplitude du N400 était modulé par le type d'agent (humain ou robot) qui prononçait les phrases. La variance inter-participants était de 0.999 (SE $= 0.223$), reflétant la variabilité des réponses entre les participants.

Figure 10.1

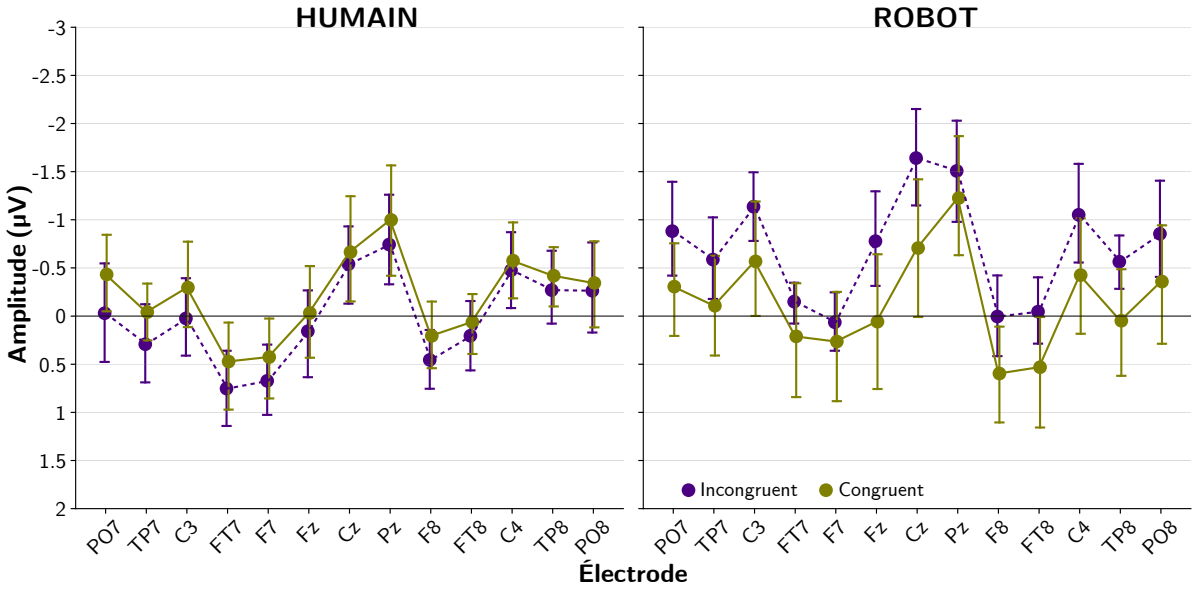
Tracés des potentiels évoqués des deux conditions aux électrodes Cz et Pz



Note. Les tracés représentent les ERP sur les électrodes Cz et Pz pour les deux types de mots cible (congruent et incongruent) dans la condition Humain et Robot. **(a)** : Électrode Cz pour les conditions contrôle (gauche) et expérimentale (droite). **(b)** : Électrode Pz.

Figure 10.2

Amplitudes moyennes pour les treize électrodes d'intérêt

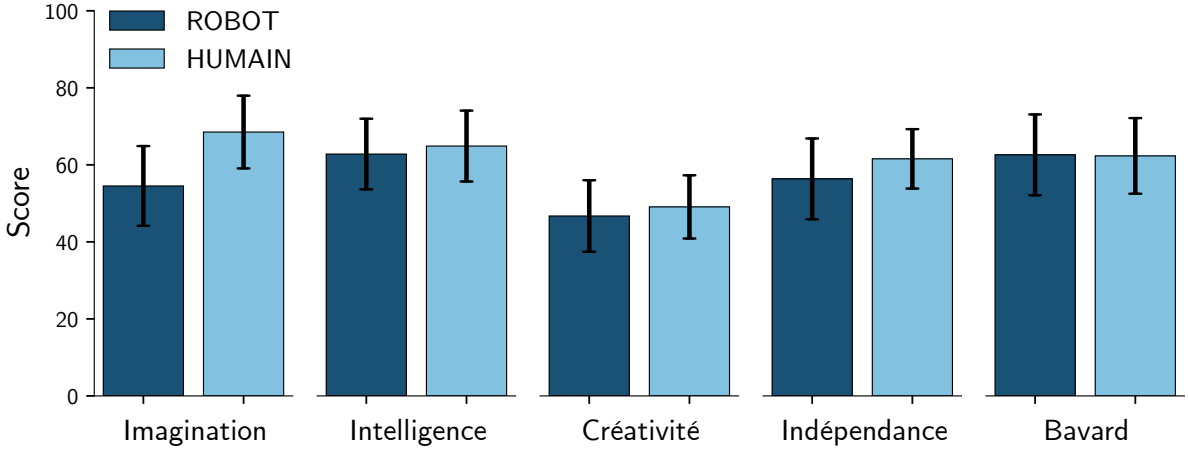


Note. Amplitudes moyennes sur les treize électrodes d'intérêt au cours de la fenêtre de 500 à 700 ms après le début du stimulus. Les barres d'erreur correspondent aux intervalles de confiance à 95%. **(Gauche)** : résultats pour la condition *Humain*. **(Droite)** : résultats pour la condition *Robot*.

10.3.2 Analyse du questionnaire

Figure 10.3

Scores pour les cinq affirmations explorant les perceptions à l'égard des agents



Note. Valeurs moyennes pour les cinq affirmations qui sondaient les perceptions des participants à l'égard des agents, avec les intervalles de confiance à 95% correspondants.

La Figure 10.3 présente les scores moyens obtenus (avec les intervalles de confiance à 95%) pour les cinq affirmations qui exploraient les perceptions des participants à l'égard des agents (Robot et Humain).

Un test *U* de Mann-Whitney n'a pas permis de détecter de différence significative entre les scores composites moyens des deux groupes ($U = 246.5$, $p = .203$). Ces résultats suggèrent que les perceptions que les participants ont au sujet de ces aspects ne varient pas de manière significative selon le groupe auquel ils appartiennent.

En ce qui concerne le questionnaire de Ho et MacDorman (2010), une ANOVA mixte avec la dimension (*Humanité, Étrangeté, Attrait*) et le groupe (Robot ou Humain) comme facteurs montre un effet significatif du groupe ($F = 20.041$, $p < .001$, $\eta_p^2 = .294$), de la dimension ($F = 21.083$, $p < .001$, $\eta_p^2 = .305$), et une interaction entre les deux facteurs ($F = 17.767$, $p < .001$, $\eta_p^2 = .270$). Un test *U* de Mann-Whitney a permis de mettre en évidence une différence significative pour l'indicateur de Humanité ($U = 70.0$, $p < .001$), mais aucune différence significative n'a été trouvée pour les indicateurs d'*Étrangeté* ($U = 351.5$, $p = .455$) et

d'*Attrait* ($U = 315.5$, $p = .961$). Le Tableau 10.2 donne les scores moyens correspondants.

Table 10.2

Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon l'agent

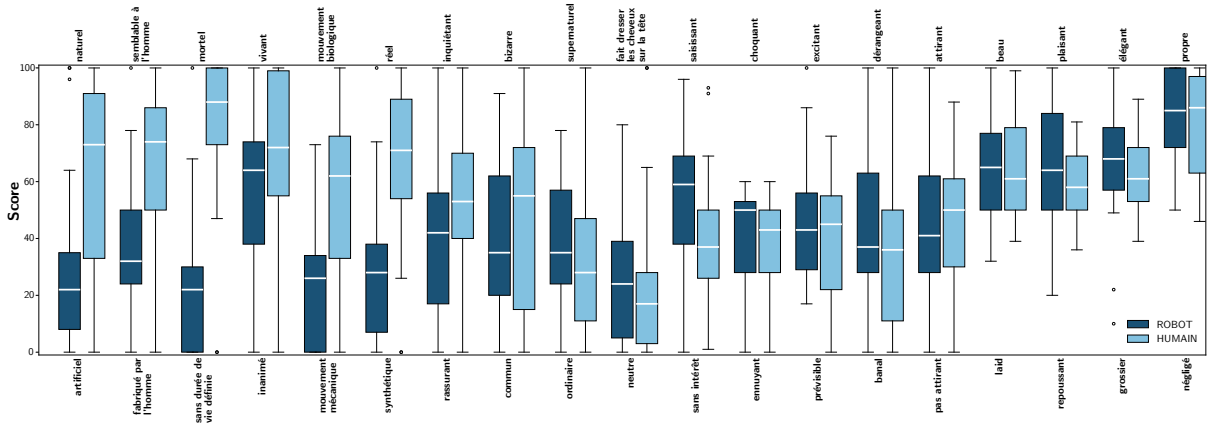
Condition	Humanité	Étrangeté	Attrait
Humain	M = 66.81, SD = 22.33	M = 38.48, SD = 16.40	M = 62.33, SD = 13.32
Robot	M = 32.15, SD = 18.80	M = 41.64, SD = 14.22	M = 64.38, SD = 12.77

Note. Les valeurs indiquent les scores moyens (M) et les écarts-types (SD) pour chaque type d'agent présenté (*Humain* vs. *Robot*).

Par conséquent, sans surprise, les participants ont considéré le robot comme plus artificiel que l'humain. La Figure 10.4 présente les diagrammes en boîte comparant les perceptions pour les items du questionnaire de Ho et MacDorman (2010) dans les groupes Robot et Humain.

Figure 10.4

Scores au questionnaire de Ho et MacDorman (2010) selon la condition



Note. Diagramme en boîte présentant les perceptions des participants à l'égard des agents (Humain et Robot) aux items du questionnaire de Ho et MacDorman (2010), avec les intervalles de confiance à 95% correspondants.

Le Tableau 10.3 résume les évaluations des croyances des participants concernant la capacité de l'agent à exprimer ou à ressentir des émotions dans les conditions Humain et Robot. En moyenne, les scores pour la capacité de l'agent à exprimer des émotions étaient plus élevés pour le Robot que pour la condition Humain, probablement en raison de l'expression faciale amicale du robot.

En revanche, les scores pour la capacité à ressentir des émotions étaient plus faibles dans la condition Robot que dans la condition Humain. Comme l'indique la médiane, la moitié des participants ont attribué un score de 43 sur 100 à la possibilité que le robot puisse ressentir des émotions. Sur les vingt-cinq participants de la condition Robot, seulement trois ont affirmé sans équivoque que le robot ne pouvait pas ressentir d'émotion et ont donné un score de 0.

Table 10.3

Scores relatifs aux capacités émotionnelles perçues

Condition	Exprimer des émotions	Ressentir des émotions
Humain	M $= 75.00$, SD $= 30.05$, Mdn $= 87.00$, min $= 0$, max $= 100$	M $= 79.64$, SD $= 28.51$, Mdn $= 92.00$, min $= 0$, max $= 100$
Robot	M $= 80.88$, SD $= 22.87$, Mdn $= 92.00$, min $= 20$, max $= 100$	M $= 52.68$, SD $= 36.07$, Mdn $= 43.00$, min $= 0$, max $= 100$

Note. Les valeurs indiquent les moyennes (M), écarts-types (SD), médianes (Mdn) ainsi que les valeurs minimales et maximales (min, max) des scores relatifs aux capacités émotionnelles perçues pour chaque type d'agent (*Humain* vs. *Robot*). Les capacités émotionnelles sont distinguées entre la capacité à exprimer et celle à de ressentir des émotions.

10.3.3 Impact de la croyance

ERP et croyance

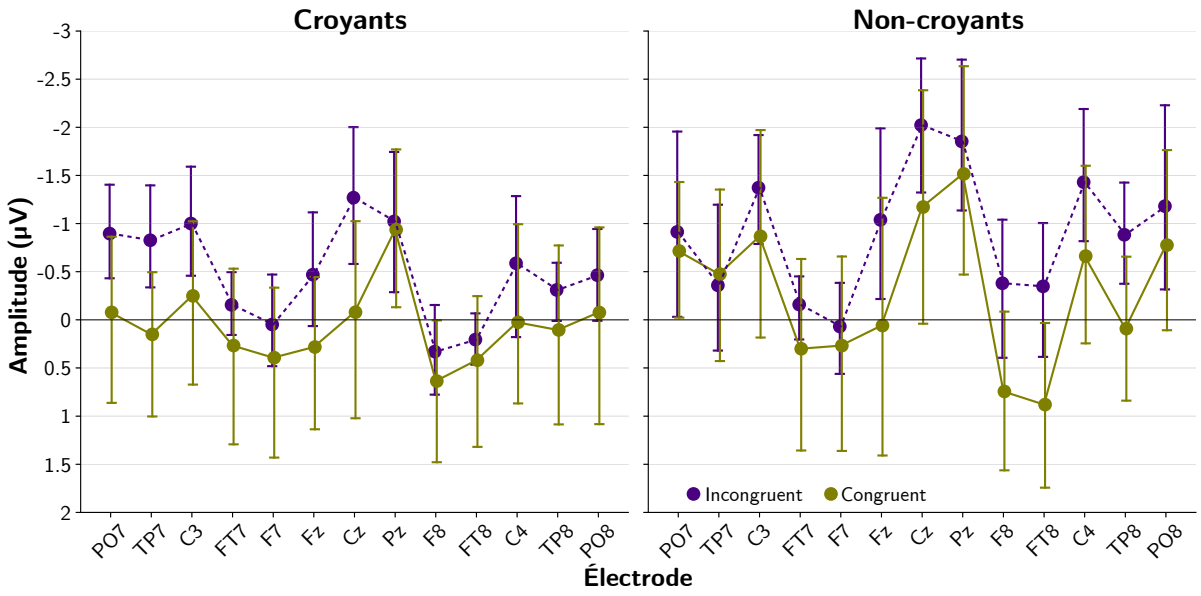
Pour évaluer l'impact potentiel des croyances des participants dans la condition Robot sur la composante N400 ($n= 25$, Mdn $= 43$, min $= 0$, max $= 100$), l'approche idéale aurait été de reproduire l'Étude 1 en contrastant les participants ayant des scores égaux à zéro avec ceux ayant un score maximal de croyance (groupes « non-croyants » vs. « croyants »). Cependant, cette approche a été limitée et rendue impossible par le nombre insuffisant de participants ayant un

score de 0 ($n = 3$) et par la nécessité d'écarter trois participants dont les scores étaient trop proches de la médiane.

Par conséquent, nous avons utilisé une logique similaire, mais avec une plage de scores plus large pour former les deux sous-groupes. Nous avons divisé les vingt-deux participants restants en deux groupes : un groupe comprenait les onze participants ayant les scores les plus bas (min $= 0$, max $= 38$, $M = 18.27$, Mdn $= 18$), tandis que l'autre comprenait les onze participants ayant les scores de croyance les plus élevés dans la potentielle capacité du robot à ressentir des émotions (min $= 71$, max $= 100$, $M = 88.64$, Mdn $= 86$).

Figure 10.5

Amplitudes moyennes des ERP pour les treize électrodes d'intérêt selon la croyance



Note. Amplitudes moyennes des ERP sur une fenêtre temporelle de 500 à 700 ms après le début du stimulus pour les treize électrodes d'intérêt avec leur intervalles de confiance à 95% correspondant. **(Gauche)** : amplitudes pour les participants qui croyaient fermement que le robot pouvait ressentir des émotions. **(Droite)** : amplitudes pour les participants qui étaient sceptiques quant à la capacité du robot à ressentir des émotions.

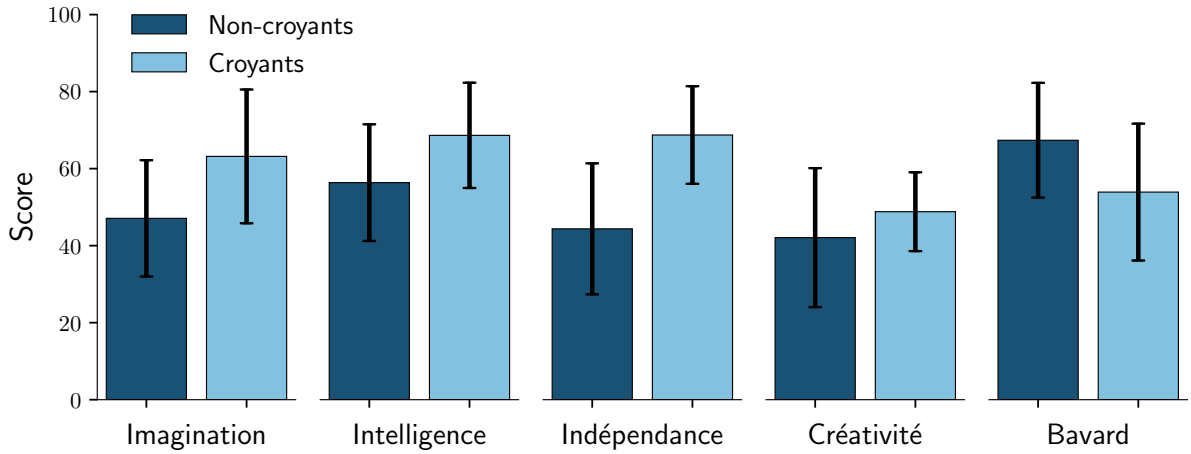
Le LMM n'a révélé aucune interaction significative entre les deux facteurs de groupe et de congruence ($b = 0.058$, SE $= 0.565$, $z = 0.102$, $p = .92$, $IC_{95\%} = [-1.050, 1.165]$). La variance au sein du facteur groupe était de 1.758 (SE $= 0.548$).

L'ANOVA à mesures répétées a confirmé que l'interaction entre la congruence et le groupe n'était pas significative ($F(1, 20) = 0.010$, $p = .92$, $\eta_p^2 = .0005$), indiquant que le seul effet significatif était dû au groupe ($F(1, 20) = 4.344$, $p = .05$, $\eta_p^2 = .178$). La Figure 10.5 présente les amplitudes moyennes de N400 dans l'intervalle de 500 à 700 ms à travers les treize électrodes pour les phrases congruentes et incongruentes. Aucune différence évidente dans l'effet N400 n'est observée entre les deux ensembles de données.

Questionnaire et Croyance

Figure 10.6

Scores pour les cinq affirmations explorant les perceptions à l'égard du robot selon la croyance



Note. Scores moyens et intervalles de confiance à 95% correspondants dans les sous-groupes de croyance pour les cinq affirmations qui exploraient les perceptions du robot.

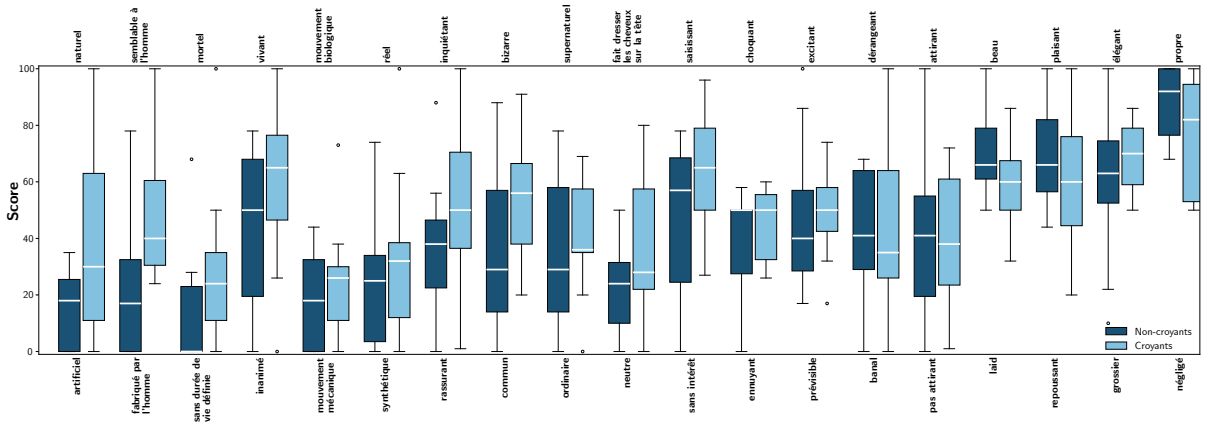
La Figure 10.6 contraste les scores des sous-groupes (avec leurs intervalles de confiance correspondants à 95%) pour les cinq déclarations qui ont exploré les perceptions des participants à propos du robot. Sauf pour « *Bavard* », une tendance générale à une perception moins positive du robot est observée dans le groupe dit des « non-croyants ». Pour les cinq traits, le score composite était de 51.454 (SD $= 17.746$) pour les « non-croyants » et de 60.654 (SD $= 11.88$) pour

les « croyants ». Un test *U* de Mann-Whitney n'a révélé aucune différence significative entre les scores composites moyens des deux sous-groupes ($U = 46.5$, $p = .375$). Cependant, une analyse exploratoire a montré que lorsque l'indicateur « *Bavard* » était mis de côté, le score composite des quatre indicateurs restants distinguait les deux groupes avec une vision plus positive du robot par les « croyants » ($U = 28.5$, $p = .038$).

Concernant les réponses au questionnaire de Ho et MacDorman (2010), une ANOVA mixte avec la dimension (*Humanité, Attrait, Étrangeté*) et le groupe (non-croyant vs. croyant) comme facteurs a montré un effet significatif de la dimension ($F(2, 40) = 23.098$, $p < .001$, $\eta_p^2 = 0.536$), du groupe ($F(1, 20) = 5.382$, $p = .031$, $\eta_p^2 = 0.212$), mais aucun effet d'interaction significatif, ($F(2, 40) = 2.415$, $p = .102$, $\eta_p^2 = 0.108$). Le Tableau 10.4 donne les scores moyens aux dimensions du questionnaire. La Figure 10.7 présente les boîtes à moustaches comparant les perceptions pour les dimensions dans les groupes « non-croyants » et « croyants ».

Figure 10.7

Scores selon le sous-groupe de croyance aux dimensions du questionnaire de Ho et MacDorman (2010)



Note. Diagramme en boîte présentant les perceptions des participants à l'égard du robot en fonction du sous-groupe de croyance aux items du questionnaire Ho et MacDorman (2010), avec les intervalles de confiance à 95% correspondants.

Table 10.4

Scores moyens aux dimensions du questionnaire de Ho et MacDorman (2010) selon le sous-groupe de croyance

Condition	Humanité	Étrangeté	Attrait
Non-croyants	M $= 22.99$, SD $= 15.42$	M $= 38.30$, SD $= 12.95$	M $= 65.62$, SD $= 14.10$
Croyants	M $= 38.89$, SD $= 20.29$	M $= 48.59$, SD $= 13.72$	M $= 61.07$, SD $= 11.53$

Note. Les valeurs (de 0 à 100) indiquent les moyennes (M) et écarts-types (SD) sur les dimensions du questionnaire de Ho et MacDorman (2010) selon le sous-groupe de croyance.

10.4 Discussion partielle

Cette étude étend les résultats de la première étude (Chapitre 9) en comparant les réactions des participants à des phrases mentionnant des émotions selon qu'elles sont prononcées par un robot ou par un humain. Les résultats confirment que notre paradigme N400 s'applique également à des énoncés dans lesquels un robot parle de ses émotions, révélant la sensibilité des participants à l'incongruence entre ce type d'énoncé et les capacités réelles de l'agent.

L'absence d'un effet lié à l'incongruence dans la condition Humain (contrôle) souligne qu'un tel discours prononcé par un humain est en accord avec nos modèles internes de l'expérience émotionnelle humaine. Nos résultats s'alignent avec ceux de van Berkum et al. (2008) qui ont trouvé qu'une phrase cohérente sur le plan linguistique comme par exemple « *Chaque soir, je bois un peu de vin avant d'aller me coucher* » provoque un effet N400 plus grand lorsqu'elle est prononcée par une voix d'un enfant plutôt que par celle d'un adulte. L'incongruence provient du décalage entre le contenu de l'énoncé et l'identité du locuteur.

Dans l'ensemble, ces résultats soulignent le rôle des attentes et des connaissances du monde dans le traitement de la parole, tout en apportant un éclairage sur la manière dont nous évaluons un même discours, qu'il soit prononcé par un humain ou par un agent artificiel.

Cependant, la première étude (Chapitre 9) avait montré que, face à un robot dépourvu de bras affirmant avoir serré la main de quelqu'un, les participants pouvaient imaginer la présence de bras cachés, ce qui diminuait la perception d'incongruence et, par conséquent, la réponse N400.

Un schéma très différent a été observé dans cette seconde étude lorsqu'il s'agit de capacités émotionnelles. En effet, une grande majorité de participants avait indiqué qu'il était possible que le robot puisse ressentir des émotions. Étant donné cette déclaration des participants, la croyance devait en principe réduire, voire annuler, tout effet lié à la perception de l'incongruité puisque le robot est considéré comme pouvant avoir des émotions.

Néanmoins, malgré une telle croyance, un effet N400 marqué a été observé lorsque le robot parlait de ses émotions : la comparaison entre sous-groupes, fondée sur les scores d'évaluation les plus bas et les plus élevés de croyance dans la capacité émotionnelle du robot, n'a montré aucune différence dans l'intensité de l'effet N400. Autrement dit, le fait de croire qu'un robot puisse ressentir des émotions n'a pas atténué la réponse cérébrale associée à l'incongruence.

Conformément à notre hypothèse, l'effet N400 observé dans la condition où l'agent est un robot soutient l'idée qu'un robot discutant de ses émotions est perçu comme une incongruité. Bien qu'un robot puisse être programmé pour « exprimer » des émotions, percevoir de telles expressions comme étant réellement possible pour le robot semble poser un défi cognitif persistant pour l'humain.

Enfin, comme dans l'Étude 1, les croyances des participants semblaient néanmoins influencer l'évaluation subjective du robot, telle qu'estimée avec les cinq descripteurs essentialistes (à l'exception de « *Bavard* »), avec des scores plus bas donnés par les participants sceptiques quant à la capacité du robot à ressentir des émotions. En revanche, à l'exception de la dimension d'Humanité, les scores des participants au questionnaire de Ho et MacDorman (2010) n'étaient pas différents selon la croyance.

10.5 Discussion générale

Le paradigme proposé s'appuie sur la composante N400 en l'étendant aux contextes d'HRI pour examiner la manière dont les humains réagissent à l'incongruence issue d'un robot social.

Nos résultats confirment le rôle de la N400 dans la détection de l'incongruence dans ce cadre : l'amplitude de la N400 augmente significativement lorsque les participants sont confrontés à un robot dont le discours entre en

conflit avec ses capacités physiques (observables) ou émotionnelles (inférées).

Dans la première étude (Chapitre 9), les phrases prononcées par le robot mentionnant des actions physiquement impossibles pour celui-ci, ont déclenché des réponses N400 plus fortes, indiquant que les participants percevaient ses énoncés comme incongruents avec sa condition physique. Cet effet était toutefois atténué chez les participants croyants à la possibilité que le robot possède des membres cachés, suggérant que les attentes individuelles peuvent façonner les interprétations du discours d'un agent artificiel.

La seconde étude (Chapitre 10) a étendu le paradigme au domaine des émotions, en examinant l'incongruence entre des énoncés dans lesquels il évoquait ses émotions et sa nature d'agent artificiel, révélant des réponses N400 similaires à celles observées dans la première étude. Cependant, contrairement aux capacités physiques, les croyances des participants sur la capacité émotionnelle du robot n'ont pas modulé cet effet lié à l'incongruence. Ceci met en évidence un scepticisme persistant envers les capacités émotionnelles du robot qui va au-delà de la réponse explicite recueillie par le questionnaire.

Nos résultats révèlent un profond engagement cognitif envers l'évaluation de la congruence de ce qu'un agent peut faire et ce qu'il peut prétendre faire. De même, ils étendent les recherches antérieures à propos de la sensibilité au contexte de l'effet N400 lors du traitement du langage. Par exemple, Hagoort et van Berkum (2007) avaient montré que, étant donné le fait bien connu parmi les Néerlandais que les trains néerlandais sont jaunes, une phrase comme « *Les trains néerlandais sont blancs et très bondés* » suscite l'effet N400. Dans notre étude, le décalage ou l'incompatibilité déclenchant cette réponse découle des représentations mentales que les participants se font du locuteur.

Le fait que la N400 soit sensible à ce type de connaissance du monde en fait un indicateur particulièrement pertinent, du fait qu'elle reflète la perception d'un décalage entre ce qu'un robot dit et ce qu'il est censé pouvoir faire ou ressentir, offrant ainsi un accès direct aux mécanismes cognitifs impliqués dans la perception des agents artificiels.

Rappelons que, dans la première étude, environ un tiers des participants seulement ont affirmé que le robot n'avait pas de membres cachés. Ainsi, la majorité des participants a envisagé la possibilité que le robot possède des membres cachés, malgré des preuves visuelles évidentes du contraire. Inspirés par le per-

sonnage *Eve* du film *Wall-E*, qui peut déployer des extrémités cachées lorsque nécessaire, nous avons appelé ce phénomène le « *Eve effect bias* » (Gigandet et al., 2023). Ce dernier fait référence à la tendance des humains à attribuer des capacités physiques aux robots sans preuve à l'appui. Ce biais, qui semble atténuer la discordance lors du traitement des phrases, révèle une certaine tolérance ou indulgence de la part des humains dans l'acceptation d'informations provenant des ASA tels que les robots sociaux, suggérant une disposition à ajuster leur attentes face à des capacités technologiques n'ayant pourtant encore jamais été observées.

À l'inverse, lorsque le robot fait référence à ses propres émotions, un domaine profondément associé aux humains et aux autres animaux, cet ajustement ne se produit pas, et ce malgré le fait que les participants aient rapporté explicitement la possibilité que le robot puisse ressentir des émotions. Les réponses N400 qui en résultent suggèrent que, dans le cas d'un robot, le fait de ressentir des émotions, d'en faire l'expérience subjective véritable, n'est pas perçu comme compatible avec nos représentations mentales des entités artificielles.

En somme, un tel résultat, bien qu'il soit conforme aux travaux montrant que la N400 est sensible à la compatibilité entre le contenu de ce qui est dit et l'identité du locuteur (van Berkum et al., 2008), s'accorde également avec le modèle de la perception de l'esprit (*Mind Perception*), selon lequel les humains attribuent volontiers de l'agentivité (des fonctions d'action et de raisonnement telles que la mémoire, la planification, etc), mais beaucoup plus difficilement de l'expérience vécue : des capacités subjectives telles qu'une personnalité, la conscience, la fierté, l'embarras, la joie, la peur, la douleur ou le plaisir (Gray et al., 2007 ; Gray & Wegner, 2012).

La frontière que nous mettons en évidence porte précisément sur cette dimension d'expérience : même lorsque le robot le prétend et que nous affirmons explicitement que cela est possible, une limite cognitive persiste, et nous détectons l'incongruité sans chercher à la résoudre en lui attribuant effectivement cette capacité, contrairement à ce que nous faisons lorsqu'il évoque des actions qu'il aurait accomplies, bien qu'il ne dispose pas des membres nécessaires.

10.5.1 Limites

Bien que nos deux expériences apportent des éléments nouveaux à la compréhension de la perception humaine des émotions chez les ASA, il est important de reconnaître les limites.

Tout d'abord, la généralisation de nos résultats peut être restreinte par les biais liés à la conception de l'étude, au type de robot, à la nature des tâches demandées tout comme aux caractéristiques de l'échantillon.

En effet, ces deux études sont basées sur des vidéos avec un seul robot, à savoir *Buddy*, et ne testent pas les interactions directes, ce qui pourrait influencer les réactions des participants. Une interaction directe avec le robot (et d'autres types de robots) pourrait générer des réponses différentes en raison de la présence physique et de la dynamique en temps réel.

De plus, l'échantillon de participants est largement composé de femmes (37 sur 56 dans l'Étude 1, 42 sur 50 dans l'Étude 2), ce qui pourrait biaiser les résultats compte tenu des différences de genre en matière d'empathie et de réaction émotionnelle (Christov-Moore et al., 2014). Les expériences antérieures des participants avec les robots et leurs attentes, qui n'ont pas été précisément et totalement contrôlées pourraient influencer leurs perceptions et réponses. Nomura et al. (2006) ont, par exemple, mis en évidence une corrélation entre l'expérience personnelle et l'attitude envers les robots, les interactions antérieures favorisant une perception plus positive.

Aussi, le fait d'avoir mené l'étude uniquement dans la métropole de Lille, en France, limite la portée des résultats à d'autres contextes culturels. Dans le Chapitre 1, plusieurs recherches interculturelles montrant l'impact du contexte culturel, des attitudes et des habitudes sur la perception des ASA ont été rapportées.

Enfin, notre focalisation sur le discours lié aux émotions du robot, qui facilite probablement l'attribution de sentiments au locuteur par le biais d'une narration à la première personne plutôt que d'intégrer des indices non verbaux tels que les expressions faciales et les gestes, peut limiter notre compréhension globale de la manière dont les humains perçoivent les émotions pour un robot.

Cinquième partie

Discussion Générale

Discussion Générale

Discussion Générale

Ce travail de thèse visait à explorer et interroger comment, dans certains contextes, les humains se connectent, s'ajustent, perçoivent et évaluent les agents artificiels.

Plus précisément, et en comparaison avec les interactions entre humains, il s'agissait d'investiguer les attentes, les spécificités et les limites que les humains manifestent face à de tels agents, notamment les robots sociaux.

L'étude de ces questions s'est articulée autour de trois axes principaux, partageant un même objectif : documenter les spécificités de l'interaction avec des agents artificiels, proposer des paradigmes transférables à d'autres recherches et fournir des indicateurs utiles à la conception d'agents sociaux artificiels tels que les robots sociaux.

Chapitre 11

Synthèse des contributions

11.1 Axe 1 - Présence sociale et dynamiques temporelles avec des agents artificiels dans un environnement minimaliste

Le premier axe apporte des éléments de réponse à notre objectif général en étudiant la possibilité qu'un sentiment de *Présence Sociale* émerge dans un contexte d'interaction minimaliste face à un agent artificiel. Elle interroge également, dans ce même type d'environnement, si et comment certains types d'agents artificiels peuvent susciter des dynamiques d'interaction similaires ou comparables à celles observées dans les interactions humain-humain.

Cette approche offre ainsi un moyen de mesurer, au-delà de la simple reconnaissance de l'autre, le sentiment d'être avec un autre, qu'est la *Présence Sociale*, et d'en identifier les conditions d'émergence dans les interactions entre humains et agents artificiels. Les deux manipulations expérimentales de cet axe visaient à déterminer, d'une part, si une instruction explicite d'engager une interaction influence les comportements d'exploration et la *Présence Sociale* perçue des agents, et, d'autre part, comment les propriétés comportementales des agents affectent cette *Présence Sociale*.

11.1.1 Rôle de l'instruction sociale

La première contribution empirique a adapté le paradigme du *Perceptual Crossing*, notamment le dispositif de l'expérience de Auvray et al. (2009) pour investiguer l'émergence de la *Présence Sociale* avec des agents artificiels dans un environnement minimaliste unidimensionnel. Deux études complémentaires ont manipulé l'instruction donnée aux participants (Étude 1 : « Explorez l'envi-

ronnement » ; Étude 2 : « Explorez l'environnement et essayez d'interagir avec l'autre ») et le type d'agent rencontré : Périodique (mouvements périodiques réguliers autour du centre), Fluctuant (amplitude fluctuante des oscillations autour du centre), Réactif (réponses contingentes aux croisements avec le participant et explore tout l'environnement) ou Indépendant (mouvement autonome dans tout l'environnement).

La comparaison entre l'Étude 1 (consigne neutre) et l'Étude 2 (consigne sociale) montre que l'attente d'autrui créée par l'instruction exerce un effet marqué sur la *Présence Sociale* perçue. Les participants ayant reçu la consigne d'essayer d'interagir avec l'autre rapportent des scores significativement plus élevés de coprésence et d'interdépendance comportementale que ceux ayant simplement reçu une consigne d'exploration sans incitation sociale. Cet effet apparaît dès le premier bloc d'interaction. Fait notable : même en l'absence d'instructions explicites d'interaction, les scores des participants aux questionnaires ne se situent pas aux valeurs minimales mais restent intermédiaires. Ces résultats suggèrent que même sans incitation sociale explicite, le simple fait d'évoluer dans un environnement partagé avec un agent (dont on ignore pourtant la nature), peut susciter un léger sentiment d'être avec un autre. Le cadrage social explicite, en revanche, amplifie ce sentiment et renforce la perception d'une interaction réciproque.

Cependant, cette amplification subjective de la *Présence Sociale* par l'incitation sociale ne s'accompagne pas systématiquement de changements dans les comportements exploratoires et dans les dynamiques temporelles de l'interaction. L'incitation sociale n'augmente que modérément le nombre de croisements avec le partenaire, et ce de manière sélective selon le type d'agent. En effet, on observe davantage de croisements uniquement lorsque les participants reçoivent l'incitation sociale et qu'ils interagissent avec les agents oscillatoires (les agents Périodique et Fluctuant).

Finalement, au niveau de la structure temporelle conjointe entre le participant et l'agent, les résultats montrent que l'incitation sociale modifie sélectivement l'organisation temporelle du couplage : au premier bloc, la multifractalité (Δh) est plus élevée chez les participants ayant reçu une incitation sociale, révélant des signes d'une dynamique conjointe plus riche et variable à travers les échelles temporelles. Cet effet variait de façon marginale selon le type

d'agent. Aussi, les participants ayant reçu une incitation sociale présentaient, avec l'agent Indépendant, une structure temporelle du couplage plus organisée et des corrélations temporelles plus persistantes.

11.1.2 Impact de la contingence de l'agent aux actions du participant sur la présence sociale et les comportements d'exploration

L'Axe 1 a également évalué l'impact de la contingence du comportement de l'agent aux croisements avec le participant sur la *Présence Sociale* perçue et les dynamiques de l'interaction. Nous avions formulé l'hypothèse qu'un agent dont le comportement est directement conditionné par les croisements avec le participant, susciterait des scores plus élevés de coprésence et d'interdépendance perçue, un nombre de croisements plus important, ainsi que des signatures dynamiques spécifiques dans la structure fractale des séries temporelles du couplage entre participant et agent.

Comme attendu, l'agent Réactif, dont les déplacements dépendaient directement des collisions avec le participant, a généré significativement plus de croisements, indépendamment de l'instruction sociale, que les autres agents (Fluctuant, Périodique, Indépendant). En revanche, cette différence nette dans les comportements ne se traduit pas sur le plan du sentiment de *Présence Sociale* : les scores de coprésence et d'interdépendance perçue avec cet agent ne sont pas significativement différents des autres agents.

Sur le plan de la structure des séries temporelles, les analyses fractales et multifractales révèlent un effet du type d'agent, mais pas dans la direction attendue. En effet, la dynamique avec l'agent Réactif présente des exposants β très faibles, loin d'un profil $1/f$ (bruit rose) anticipé, et sa largeur multifractale (Δh) n'est pas particulièrement élevée. En d'autres termes, cette dynamique évolue sur un régime proche du bruit blanc avec beaucoup de micro-fluctuations rapides, peu de mémoire, et pratiquement pas de corrélations à longue portée. Les échanges avec cet agent ne présentent donc pas de dépendances soutenues à travers le temps, la coordination se produisant de manière ponctuelle, sans se prolonger ni s'articuler sur plusieurs échelles temporelles.

En somme, la contingence implémentée dans cet agent facilite la rencontre

par un nombre plus élevé de croisements durant l'interaction. Cependant, elle n'amplifie pas la *Présence Sociale* perçue. L'activité conjointe avec cet agent, témoignant d'un ajustement sur l'instant, faiblement corrélé et seulement ponctuellement organisé, est à l'opposé des motifs auto-similaires caractéristiques d'une coordination entre humains tel qu'observé par Bedia et al. (2014) dans ce type d'environnement minimaliste.

Au-delà du cas de l'agent Réactif, les interactions avec les autres agents présentent aussi des profils contrastés. Celles impliquant les agents Périodique et Indépendant présentent les exposants α et β les plus élevés, correspondant à des dynamiques plus corrélées et persistantes, tandis que celle qui implique l'agent Fluctuant se situe à un niveau intermédiaire. La largeur multifractale (Δh), qui reflète une plus grande diversité d'échelles temporelles dans la dynamique conjointe, était la plus grande dans les interactions avec l'agent Indépendant, suivie de celles avec le Périodique. Ces différences sont nettes dans les analyses fractales et multifractales, mais elles ne s'accompagnent pas de différences significatives dans les scores de coprésence ou d'interdépendance perçue.

Ainsi, le type d'agent modifie clairement la structure de l'interaction, sans pour autant influencer la présence sociale perçue. À l'inverse, l'instruction sociale agit surtout sur les jugements subjectifs, avec des effets plus limités sur les mesures comportementales et dynamiques. Chaque manipulation semble donc agir sur des dimensions distinctes de l'interaction, sans que les variations observées sur le plan subjectif et comportemental apparaissent systématiquement liées dans ce cadre minimaliste. Bien que l'agent Réactif ait été programmé pour osciller au point du croisement, sa réponse reste événementielle plutôt que processuelle : il réagit ponctuellement sans s'ajuster de manière continue à la dynamique du participant. L'absence de corrélations à long terme dans la série des vitesses relatives confirme un manque de co-régulation temporelle (De Jaegher & Di Paolo, 2007), malgré des échanges fréquents. Or, la perception de socialité repose sur des alignements temporels fins, tels que la synchronie motrice ou les tours de parole (par exemple, Vallacher et Nowak, 2008). En l'absence de dynamique partagée étendue dans le temps, une simple contingence locale ne suffit pas à instaurer le sentiment d'un partenaire engagé. La réactivité de l'agent, bien que détectable, reste déconnectée d'un véritable schéma d'ajustement mutuel, et l'expérience vécue ne se distingue pas de celle d'un

agent non-contingent. Ces résultats soulignent que les humains sont sensibles non seulement à la présence d'une réponse, mais à la manière dont celle-ci s'inscrit dans une dynamique temporelle conjointe. La conception d'agents sociaux crédibles devrait donc viser une coordination fluide et co-régulée, plutôt qu'une simple réactivité séquentielle.

11.2 Axe 2 - Délai de réponse d'un robot à une question

À la suite de l'Axe 1, qui portait sur l'importance de la dynamique temporelle dans les interactions sensorimotrices, l'Axe 2 aborde une autre dimension clé de la temporalité dans l'interaction humain-agent : celle du délai de réponse verbal. Ce deuxième axe apporte ainsi des éléments complémentaires à la question générale de cette thèse, en s'intéressant aux attentes temporelles dans les échanges verbaux entre humains et agents artificiels. Deux études complémentaires ont examiné comment le délai de réponse d'un robot et son style de communication influencent la perception temporelle de l'interaction et l'évaluation globale d'un robot social. L'Étude 1 adoptait une approche psychophysique pour identifier le délai de réponse perçu comme optimal et examiner comment différents styles de communication (Autoritaire, Soumis, Neutre, Enfantin), ainsi qu'une condition Rideau (où le robot était caché) modulent cette perception temporelle. L'Étude 2 approfondissait cette analyse en examinant comment des délais fixes, mais s'écartant fortement de cet optimal (très rapide, très lent ou optimal), interagissent avec les styles de communication pour influencer l'évaluation subjective du robot.

11.2.1 Discussions humain-robot : un délai de réponse d'environ 700 ms

Les analyses psychophysiques de l'Étude 1 ont révélé un repère temporel net : le délai de réponse perçu comme optimal pour un robot social, jugé ni trop long ni trop court, à des questions-réponses fermées se situe autour de 700 ms, indépendamment du style de communication ou du fait qu'il soit visible ou placé derrière un rideau. Ce résultat, obtenu par l'ajustement de fonctions sigmoïdes

aux jugements temporels des participants, indique que le Point d'Égalité Subjective ne diffère pas significativement entre les conditions.

Ce délai optimal est nettement supérieur à celui mesuré dans les échanges entre humains, où les transitions entre tours de parole s'effectuent en moyenne dans les 200 ms (Stivers et al., 2009) et jusqu'à 100-180 ms pour les réponses à des questions fermées oui/non (Strömbergsson et al., 2013). Autrement dit, le délai jugé « naturel » pour un robot est de trois à sept fois plus long que celui d'un interlocuteur humain.

Plusieurs explications peuvent être avancées. Cette différence suggère notamment que les humains pourraient avoir développé des attentes temporelles spécifiques pour l'interaction avec les robots sociaux, distinctes de celles appliquées aux interactions humain-humain. D'une part, il est possible que les participants aient intégré l'idée que les systèmes robotiques ou de dialogue comportent des contraintes techniques, en raison d'une exposition répétée aux limites de ces systèmes (temps de réponse fixe ou délais prolongés, proches ou supérieurs à ceux observés dans notre étude). Cela pourrait les amener à s'attendre « naturellement » à des délais plus longs et à ajuster leurs jugements en conséquence.

D'autre part, la méthodologie employée, qui consistait à recueillir des jugements explicites, à la troisième personne, sur des vidéos préenregistrées, et à mesurer des délais de réponse perçus « trop rapides » ou « trop lents » plutôt qu'à mesurer les délais effectifs dans une interaction réelle mobilise potentiellement des processus cognitifs différents. Il se pourrait donc que le délai de 700 ms reflète non pas une attente fondamentalement différente pour les robots, mais plutôt un jugement conscient et explicite de ce qui semble « approprié » ou « naturel » dans ce contexte d'observation.

11.2.2 Modulation de la tolérance aux écarts temporels par les styles de communication

Bien que le délai de réponse optimal perçu ne varie pas significativement selon le style de communication du robot, l'analyse de la tolérance aux écarts autour de ce point révèle un effet marqué du style de communication. Les courbes psychométriques, représentant la proportion de jugements « trop lent » en fonc-

tion du délai) présentent des pentes significativement différentes selon le style de communication, indiquant des zones d'incertitude temporelle de largeur variable.

Les styles Soumis ou Enfantin génèrent des pentes plus faibles, reflétant une tolérance accrue aux écarts. Les participants sont moins catégoriques dans leurs jugements et acceptent une plus large gamme de délais autour du point optimal. À l'inverse, les styles Neutre et Autoritaire suscitent des attentes temporelles plus strictes , avec des pentes plus raides et des jugements plus tranchés.

De même, il est intéressant de noter que dans la condition où le robot était caché derrière un rideau (son style de communication était neutre), l'incertitude est également plus élevée. Cela est probablement dû au fait qu'en l'absence d'indices visuels, les participants formulent des attentes différentes sur les capacités du système, ou tolèrent davantage ses potentielles limitations, le catégorisant peut-être davantage comme une simple interface vocale.

Cependant, ces variations de tolérance n'affectent pas les jugements globaux sur le robot (questionnaire de Ho et MacDorman, 2017) : indépendamment du style ou du fait qu'il soit caché ou visible, il était perçu comme tout autant artificiel et étrange. Seule la dimension relative à l'attrait physique du robot est plus faible lorsque le robot est caché, vraisemblablement en raison du manque d'indices visuels nécessaires à ce jugement.

Ces résultats suggèrent que le style de communication avec lequel un robot s'adresse à l'humain, module non pas le délai de réponse préféré avec celui-ci mais la tolérance aux écarts autour de ce point : il élargit ou resserre l'incertitude des jugements, en augmentant ou diminuant la plage de délais considérés ni « trop rapide », ni « trop lent », rendant les évaluations plus ou moins tranchées. Les styles de communication perçus comme plus doux, vulnérables ou infantilisés bénéficient d'une plus grande indulgence de la part des participants, tandis que les robots perçus comme plus dominants sont soumis à des critères temporels plus stricts. Ainsi, dans l'ensemble, le délai jugé optimal est stable quel que soit le style et c'est alors plutôt la marge d'acceptation qui change en fonction du style avec lequel le robot communique.

11.2.3 Impact des délais non-optimaux sur l'évaluation globale du robot

La seconde étude testait l'impact de délais fixes s'écartant fortement de l'optimal (200 ms, 700 ms ou 1500 ms) combinés à différents styles de communication sur l'évaluation globale du robot. Ainsi, contrairement à l'Étude 1, les participants ne faisaient que regarder les vidéos, puis évaluaient le robot en donnant leurs perceptions via les questionnaires de Ho et MacDorman (2010) et du modèle *Almere* (Heerink et al., 2010).

Un contraste intéressant se dessine entre nos deux études. Dans l'Étude 1, les participants détectaient clairement les différences de délais et identifiaient un optimal autour de 700 ms, quel que soit le style de communication ou la visibilité du robot, avec une tolérance élargie pour les styles « doux » (Soumis, Enfantin) et quand le robot est caché. Mais dans l'Étude 2, ces variations de délai, même très loin de l'optimal, n'altéraient pas l'évaluation globale du robot. Ni l'humanité perçue, ni l'étrangeté, l'attrait, la sociabilité ou la confiance ne variaient selon que le robot répond trop vite, trop lentement, ou à l'optimal. Pourtant, dans les conversations entre humains, les délais de réponse portent une signification sociale : des délais de réponse plus rapides (inférieurs à environ 250 ms) sont associés à un plus grand sentiment de connexion sociale (Templeton et al., 2022), ou à l'inverse des pauses plus longues avant de répondre à des questions de connaissances affectent le jugement porté à l'interlocuteur en diminuant la confiance et la compétence perçue (Matzinger et al., 2023). Puisque les individus détectent les variations temporelles et ont un délai optimal pour un robot, on aurait pu s'attendre à ce que les écarts à cet optimal affectent leur évaluation sociale de celui-ci. En revanche, le style de communication du robot exerce, comme dans l'Étude 1, un effet marqué et systématique sur l'évaluation du robot. Le style Autoritaire, par exemple, est jugé moins attrayant, plus intimidant, et génère davantage d'anxiété. Cela rejoint les résultats de Saunderson et Nejat (2021) sur le rejet de l'autorité venant d'un robot. D'autres styles, comme le style Enfantin, renforcent au contraire la sociabilité perçue.

Plusieurs interprétations de ces résultats peuvent être avancées. Premièrement, les questionnaires utilisés, bien qu'ayant été sensibles aux variations dans le style de communication, ne capturent peut-être pas les dimensions sociales

spécifiques qui pourraient être influencées par le délai de réponse. Deuxièmement, il se pourrait que les humains appliquent des attentes temporelles différentes aux robots et aux interlocuteurs humains. Les délais robotiques sont peut-être perçus comme des caractéristiques techniques ou fonctionnelles et non pas comme des signaux sociaux porteurs de sens, à la différence du style de communication, qui active plus directement les mécanismes de cognition sociale. Cette interprétation suggère une limite dans la façon dont les humains appliquent leur cognition sociale aux agents artificiels : certains aspects de la communication (prosodie, formulation, expressions faciales) sont rapidement interprétés socialement, tandis que d'autres (délais) ne le sont pas nécessairement.

Troisièmement, la séparation entre le délai de réponse et les autres indices de la communication (prosodie de la réponse) dans notre protocole pourrait constituer une barrière à l'attribution sociale. Dans les interactions humaines, le délai de réponse est intégré dans un ensemble multimodal d'indices : intonation, marqueurs d'hésitation, variation prosodique ou vitesse d'articulation. Dans notre protocole, le robot répondait systématiquement avec des enregistrements « Oui » ou « Non » identiques, indépendamment du délai. Cette standardisation, nécessaire au contrôle expérimental, pourrait avoir réduit la probabilité que les participants attribuent une signification sociale aux délais.

En somme, ces deux études convergent vers une conclusion nuancée : les humains sont sensibles au style de communication des robots et y réagissent de façon systématique et prévisible (rejet de l'autorité, indulgence envers les styles « doux »), mais le délai de réponse, bien que perçu consciemment, ne semble pas influencer l'évaluation sociale du robot, du moins dans le cadre d'interactions observées impliquant des questions-réponses fermées. Cela suggère que certains mécanismes sociaux appliqués aux humains peuvent s'étendre aux robots (sensibilité au style de communication et à la domination) tandis que d'autres ne s'appliquent pas du tout ou pas de la même manière.

11.3 Axe 3 - Frontières cognitives face au discours d'un robot

Le troisième axe apporte un éclairage complémentaire à l'objectif général de cette thèse en s'intéressant aux limites que les humains établissent face au discours des robots. Alors que les deux premiers axes examinaient la temporalité de l'échange et comment les humains se connectent et s'ajustent à des agents artificiels, celui-ci examinait comment les humains traitent les énoncés de robots sociaux lorsque ceux-ci dépassent les limites de ce qui leur est accessible et possible en tant qu'agent artificiel. Il s'agissait plus précisément d'examiner si le cerveau humain traite comme une incongruité le fait qu'un robot parle de ce qui lui est inaccessible. Deux études en EEG ont été menées, explorant deux situations où le robot tient un discours potentiellement incongruent avec ses capacités : la première, lorsqu'il parle d'actions physiques impossibles au vu de sa morphologie observable, et la deuxième lorsqu'il parle de ses propres émotions, au risque de dépasser les limites que l'observateur humain est prêt à lui accorder en tant qu'entité artificielle. Dans les deux cas, nous avons mesuré la réponse cérébrale (N400) face à des énoncés congruents ou incongruents avec les capacités attendues du locuteur. Les données ont également été examinées en fonction des croyances individuelles des participants quant aux capacités des robots (par exemple, possibilité de membres cachés ou d'émotions ressenties). Cela a permis d'évaluer dans quelle mesure les représentations mentales préalables influencent le traitement de ces discours.

11.3.1 Un robot sans bras et jambes parlant d'actions impossibles : validation du paradigme

Cette première étude visait à valider l'applicabilité du paradigme N400 au discours des robots en examinant la réaction cérébrale face à un robot dépourvu de bras et de jambes parlant d'actions physiquement impossibles pour lui. Ce paradigme repose sur la composante N400 des potentiels évoqués qui se manifeste par une déflexion négative du signal électrique cérébral environ 400 ms après la présentation d'un mot ou stimulus jugé incongruent avec son contexte (Kutas & Hillyard, 1980). van Berkum et al. (2008) ont démontré que cette com-

posante est également sensible à la cohérence pragmatique entre le locuteur et le contenu de son discours, ce qui rend ce paradigme particulièrement pertinent pour étudier la réception des énoncés produits par les robots. Dans notre première étude, le robot *Buddy* a été présenté dans deux conditions : une où le corps entier était visible et une où seule la tête était visible. Les participants visionnaient des vidéos où le robot prétendait avoir réalisé des actions lui étant possibles ou impossibles.

Les résultats confirment la validité du paradigme. Un effet N400 marqué est apparu dans la condition BODY (corps entier) lorsque le robot énonçait des phrases incongruentes avec ses capacités physiques, traduisant un conflit entre morphologie visible et contenu verbal. En revanche, cet effet était absent dans la condition HEAD (tête seule), validant que l'effet N400 provenait bien du décalage entre l'apparence du robot et son discours.

Ainsi, lorsqu'un robot sans bras déclare avoir effectué une action nécessitant des bras, les participants perçoivent cette déclaration comme une incongruité, tel qu'en témoigne l'activation de la N400. Ces résultats suggèrent que le cerveau humain exerce une forme de vigilance ou de scepticisme face aux affirmations d'un agent artificiel.

11.3.2 Un robot parlant de ses émotions : une limite pour l'humain

La seconde étude a étendu l'usage du paradigme N400 à un domaine plus subtil que celui de l'Étude 1 où l'incongruité reposait sur un indice visuel évident (absence de membres). Il s'agissait d'une capacité qui n'est pas directement observable : ressentir des émotions. En effet, les phrases prononcées par le robot faisaient référence à ses propres émotions, un domaine habituellement réservé aux êtres biologiques comme l'humain et les autres animaux, soulevant la question de savoir si les humains tracent une frontière cognitive stricte pour ces capacités. L'étude comparait les réactions des participants à des phrases mentionnant des émotions comme par exemple « *Ce matin j'ai terminé un examen en première, j'étais heureuse* » à une version neutre « [...] *j'étais rapide* », prononcées soit par un robot soit par un humain (condition de contrôle). Pour l'humain, les deux versions étant ainsi compatibles dans les deux cas.

Les résultats confirment que le paradigme N400 s'applique également à un tel type d'incongruité. Les phrases mentionnant des émotions ont déclenché

une déflexion négative plus prononcée que leurs contreparties neutres. En revanche, aucune différence n'a été observée pour les deux types de phrases dans la condition Humain, en cohérence avec nos modèles internes de l'expérience émotionnelle humaine. Un robot discutant de ses émotions est donc perçu comme une incongruité au niveau cérébral. Ainsi, bien qu'un robot puisse être programmé pour imiter des expressions émotionnelles, il semblerait que percevoir ces expressions comme étant réellement vécues pose un défi cognitif pour l'humain.

11.3.3 Impact sur la N400 des croyances à propos des capacité du robot

Un second volet des analyses de cette contribution concernait l'impact des croyances individuelles sur le traitement cérébral de ces énoncés. La question était de déterminer si la croyance en des capacités non évidentes du robot pouvait atténuer l'effet N400, autrement dit, si le fait d'imaginer que le robot possède telle ou telle capacité rendait ses énoncés moins incongruents.

Pour les actions physiques, le rationnel était le suivant : dans la culture populaire, les robots sont fréquemment représentés avec des capacités inattendues qui défient leur apparence initiale (la saga Transformers, le personnage Eve dans Wall-E déploie des membres dissimulés malgré son apparence épurée, etc). Si un participant envisage l'idée que le robot puisse posséder des membres cachés, alors une action comme monter des escaliers devient tout autant possible que prendre l'ascenseur. L'incongruité perçue devrait alors diminuer. On s'attendait à ce que la croyance en la possibilité que le robot puisse avoir des membres cachés atténue l'effet N400 observé. Les résultats confirment cette prédiction. Près de deux tiers des participants dans la condition où l'on voit le robot dans son entièreté ont attribué une probabilité non nulle à l'existence de membres cachés. L'analyse comparative entre participants « croyants » et « non-croyants » révèle que les non-croyants présentent un effet N400 bien plus fort face aux énoncés impossibles, tandis que chez les croyants, cet effet est atténué. Autrement dit, pour ceux qui envisagent des membres cachés, l'énoncé « impossible » devient plausible et le cerveau tend à ne plus le traiter comme une incongruité.

Pour les émotions, le rationnel était similaire, nous avions prédit que si un par-

ticipant croit que le robot peut ressentir des émotions, alors lorsqu'il dit qu'il est triste ou fâché, cela devrait être moins perçu, voire ne pas être perçu du tout, comme impossible. Cependant, les résultats révèlent un schéma totalement différent. Aucune différence n'était présente entre les deux sous-groupes. Même les participants affirmant explicitement que le robot pouvait avoir des émotions présentaient un effet N400 comparable aux plus sceptiques. Il y a donc un décalage entre ce que les participants disent croire consciemment et ce que leur cerveau traite automatiquement et cette rigidité contraste fortement avec la flexibilité observée pour les actions physiques. Autrement dit, même si le participant affirme explicitement que le robot peut avoir des émotions, le cerveau continue de traiter les énoncés liés aux émotions du robot comme une incongruité.

En somme, ces résultats révèlent que nos croyances peuvent « réparer » une impossibilité physique en l'intégrant à nos connaissances du monde (on postule des capacités cachées), ce qui atténue l'incongruité et donc l'effet N400 pour les actions ; à l'inverse, pour les émotions, même si celles-ci sont déclarées plausibles pour le robot, cette croyance n'a aucun impact et l'effet N400 demeure, signe d'une frontière mentale plus rigide autour de l'expérience subjective des agents artificiels, notamment de ressentir des émotions.

Chapitre 12

Limites et perspectives

Les travaux empiriques présentés dans cette thèse comportent toutefois plusieurs limites méthodologiques qui restreignent la portée de leurs conclusions. Ces limites concernent principalement les caractéristiques des échantillons recrutés, la validité écologique des protocoles expérimentaux et des mesures ainsi que la diversité limitée des plateformes robotiques utilisées.

12.1 Échantillons

Une première limitation transversale à l'ensemble des études concerne ainsi les caractéristiques des échantillons, notamment leur homogénéité géographique et culturelle. Les quatre études en ligne de la Partie II et Partie III ont recruté des participants quasi-exclusivement dans des pays occidentaux dont la langue principale est l'anglais. Les cinq premiers pays (Grande-Bretagne, États-Unis, Canada, Irlande, Australie) représentant à eux-seuls 96.2% des participants. Cette concentration géographique introduit ainsi des biais potentiels qui limitent la généralisabilité des résultats, en particulier pour la Partie III au regard des différences culturelles documentées dans la Partie I de cette thèse. En effet, la perception et l'acceptation des agents artificiels varient nettement selon les contextes culturels (Bartneck et al., 2005 ; Nomura et al., 2008). Des travaux interculturels montrent des motifs d'anthropomorphisme et d'attribution d'états mentaux chez des participants japonais, chinois et coréens qui diffèrent de ceux de participants occidentaux (Bartneck et al., 2005, 2006 ; Nomura et al., 2008). Par exemple, Castelo et Sarvary (2022) ont démontré que l'augmentation du réalisme physique des robots diminue le confort des participants américains mais pas celui des participants japonais, et qu'attribuer des capacités émotionnelles à un robot réduit le confort des Américains mais l'augmente chez les Japonais.

Face à ces limites, une perspective pertinente consisterait à étendre nos investigations à d'autres contextes culturels et linguistiques, notamment auprès de populations japonaises, chinoises et coréennes (parallèle aux études de Bartneck et al. (2005) et Nomura et al. (2008), afin d'examiner si la localisation géographique et l'appartenance culturelle des participants constituent des variables modératrices. Plus spécifiquement, deux axes d'investigation devraient être explorés : (1) l'effet de la culture sur le délai de réponse optimal perçu, c'est-à-dire tester si le PSE de 700 ms identifié dans notre échantillon anglo-saxon se maintient dans d'autres contextes culturels ou s'il varie en fonction des normes conversationnelles locales, et (2) l'effet de la culture sur la perception et l'évaluation des styles de communication des robots, afin de déterminer si elles reflètent des mécanismes universels ou des sensibilités culturellement situées. Une telle approche comparative permettrait de distinguer les constantes et les variations dans les attentes liées aux interactions avec les agents artificiels et leur temporalité.

Ensuite, un autre aspect concerne le déséquilibre de genre observé dans les 2 études EEG (Partie IV) : dans la première étude, 37 sur 56 personnes (66%) étaient des femmes, et dans la seconde, 42 sur 50 (84%). Si les femmes ont tendance à accepter plus facilement l'expression d'émotions chez un robot, leur surreprésentation dans notre échantillon pourrait avoir conduit à sous-estimer l'incongruence moyenne attendue dans la population générale. Néanmoins, dans nos données, l'effet N400 aux énoncés liés aux émotions du robot demeure robuste et n'est pas modulé par les croyances, ce qui suggère au plus un biais d'atténuation modéré. Il conviendrait donc de répliquer ces résultats dans des études futures avec des échantillons équilibrés en termes de genre, afin de déterminer avec précision l'amplitude des effets dans la population générale et d'examiner d'éventuelles différences de genre dans le traitement des énoncés des robots avec ce paradigme.

12.2 Validité écologique des protocoles expérimentaux et des mesures

Une autre limitation importante de ce travail concerne les biais méthodologiques liés à la validité écologique des protocoles utilisés dans les différentes

études menées. La validité écologique des tâches, c'est-à-dire la mesure dans laquelle les situations proposées reproduisent fidèlement des conditions d'interaction réelles, peut être questionnée dans la mesure où dans certains protocoles les participants observaient des vidéos (du robot seul ou d'interactions enregistrées) sans interagir directement. Autrement dit, les situations présentaient une perspective à la troisième personne plutôt qu'une interaction directe. Ces choix méthodologiques offraient un plus grand contrôle expérimental, néanmoins cette observation à la troisième personne limite la transférabilité des résultats à des situations réelles d'interaction qui mobilisent des indices supplémentaires et des adaptations dynamiques qui n'étaient pas capturées par le format vidéo.

Des futurs travaux devraient prendre en considération ces limites en adoptant des approches plus écologiques et différenciées selon les questions de recherche. Concernant les études sur le délai de réponse (Partie III), il serait pertinent d'examiner des interactions directes où les participants sont effectivement engagés dans l'échange, et d'intégrer des structures de dialogue plus complexes que des questions fermées (oui/non). Pour les études sur le traitement du discours du robot (Partie IV), de futurs protocoles devraient tester des situations où le robot n'est pas seulement en vidéo mais en face du participant, afin d'évaluer si l'engagement du participant module la réponse cérébrale mesurée. Plus généralement, l'ensemble de ces travaux gagnerait à être répliqué dans des protocoles immersifs et interactifs, permettant d'évaluer les réponses en situation réelle plutôt que sur la base d'observations passives en vidéo ou sur la base de jugements à la troisième personne et en explorant des dimensions subtiles de la communication (prosodie, gestes, regard) dans des interactions réelles avec divers robots.

12.3 Diversité limitée des plateformes robotiques étudiées

Enfin, les résultats des Parties III et IV demeurent spécifiques à un seul robot présentant des caractéristiques morphologiques particulières. Le robot utilisé, *Buddy* (*Blue Frog Robotics*), se caractérise par une petite taille, par l'absence de membres (pas de bras ni de jambes), un petit corps monté sur roues et un écran

lui permettant d'afficher un visage et d'afficher des expressions émotionnelles dynamiques. Ces caractéristiques ont joué un rôle central et ont directement influencé les phénomènes étudiés dans nos travaux. Notamment, son aspect mignon et sa petite taille pourraient artificiellement augmenter la perception de ses éventuelles capacités émotionnelles ou la tolérance aux variations temporelles observées. Les résultats pourraient donc différer avec des robots présentant d'autres caractéristiques. Par exemple, un robot de stature imposante comme Pepper (120 cm contre 60 cm pour *Buddy*) équipé de bras articulés mais au visage minimaliste (yeux LED, sans bouche mobile) pourrait susciter des attentes et des réactions différentes. De même, des robots aux apparences plus humanoïdes, zoomorphes ou abstraites influenceraient potentiellement la perception de l'intentionnalité, la proximité sociale perçue ou les attentes d'interaction.

Une perspective consisterait donc à reproduire certains protocoles avec plusieurs types de robots présentant des variations morphologiques, expressives et fonctionnelles, afin de déterminer dans quelle mesure les effets observés sont spécifiques au seul robot utilisé ou généralisables à d'autres morphologies et types de robots.

Chapitre 13

Implications

Finalement, les travaux présentés dans ce travail de thèse ont conduit à identifier plusieurs types d'implications différentes avec des avancées conceptuelles et méthodologiques qui enrichissent la compréhension de l'interaction humain-robot, des enseignements applicatifs pour la conception de ces interactions, ainsi que des enjeux éthiques.

13.1 Implications conceptuelles et méthodologiques

Nos travaux ont permis de montrer qu'il est possible d'étendre l'utilisation de la composante N400 au domaine des HRI, notamment pour étudier comment les humains traitent le discours des robots (Partie IV). Cela constitue une contribution méthodologique majeure en établissant que le marqueur cérébral lié à l'incongruité s'applique également aux énoncés produits par des agents artificiels. Ainsi, notre paradigme N400 a permis de mettre en évidence que l'on peut examiner les potentielles frontières cognitives que nous établissons face aux discours des agents artificiels.

De plus, nous avons appliqué une approche psychophysique pour tenter d'identifier le délai optimal auquel un robot doit répondre, grâce à l'utilisation du Point d'Égalité Subjective. Cela nous a ainsi permis de quantifier précisément à la fois le délai perçu comme optimal d'environ 700 ms en moyenne et d'identifier la tolérance variable en fonction du style avec lequel le robot communique face aux écarts à ce délai. Cette approche, issue de la psychophysique classique et rarement appliquée à l'interaction humain-robot, offre une méthode, dont l'analyse est peu coûteuse, pour identifier les préférences des participants sur la base de leurs jugements.

De la même manière, nous avons utilisé le paradigme du *Perceptual Cros-*

sing pour étudier la question du sentiment de présence sociale dans un environnement minimaliste, qui est une dimension jusqu'alors relativement inexplorée dans ce paradigme. Alors que les travaux antérieurs se concentraient sur la reconnaissance d'autrui (discrimination entre agent et objet) ou sur la clarté de la présence de l'autre, nos études sont les premières à mesurer explicitement le sentiment subjectif de présence sociale dans ce contexte d'interaction sensorimotrice minimale avec seulement des agents artificiels programmés comme partenaire d'interaction. Nos résultats confirment qu'une configuration minimaliste comme celle du *Perceptual Crossing* reposant sur un feedback sensorimoteur audiovisuel dans l'interaction, peut effectivement être un lieu pour créer un sentiment de présence sociale envers un agent artificiel, même sans indices anthropomorphes ni communication symbolique. Nous avons également repris et adapté les analyses des dynamiques temporelles issues de Bedia et al. (2014), c'est-à-dire les analyses fractales et multifractales des séries temporelles de vélocité relative.

Toujours sur le plan méthodologique, nos résultats soulignent l'importance de ne pas se restreindre à une seule méthode de mesure ou d'analyse. Dans les travaux de cette thèse, il a souvent été observé des décalages entre différents types de mesures. Dans notre adaptation du *Perceptual Crossing*, certains agents présentaient un nombre élevé de croisements avec le participant sans pour autant qu'une forte présence sociale soit retrouvée. En revanche, l'analyse fractale de la dynamique entre l'agent et le participant a permis d'identifier que malgré le nombre élevé de contacts, l'agent et le participant présentaient en réalité de faibles corrélations entre eux. Dans les travaux menés avec notre paradigme N400, si nous nous étions limités à l'analyse des données EEG, nous aurions pu conclure à un traitement de l'incongruité globalement identique dans l'ensemble de notre population, compte tenu de l'ampleur de l'effet N400 observé. Le croisement avec le questionnaire a toutefois révélé que l'amplitude de la N400 est parfois modulée (réduite) par ce que les participants considèrent comme possible, atténuant ainsi l'incongruité liée à l'impossibilité de l'énoncé du robot. Sans cette mesure explicite des croyances, nous n'aurions pas pu interpréter correctement le fait que les participants ne disent pas nécessairement ce que l'on mesure au niveau cérébral, ni interpréter les variations individuelles de la composante N400. Ces exemples démontrent que la combinaison de mesures

comportementales, subjectives et physiologiques permet de nuancer certaines interprétations, de tester des hypothèses explicatives sur les processus en jeu et qu'une combinaison de méthodes ne se contente pas seulement de fournir une description plus riche. En somme, la convergence ou la divergence observée entre les mesures auto-rapportées, comportementales, temporelles ou encore neurophysiologiques nous a offert une compréhension plus riche et plus nuancée des mécanismes en jeu. Cela rejoint les propositions de Bethel et Murphy (2010) qui recommandaient l'utilisation de méthodes plurielles dans l'investigation des HRI.

De futures recherches gagneraient ainsi à adopter de telles approches multiméthodes afin de mieux appréhender la complexité de l'expérience sociale avec les agents artificiels.

13.2 Implications pour la conception des interactions avec les robots sociaux

En ce qui concerne les implications en lien avec la conception des robots, nos résultats montrent qu'une configuration minimaliste comme celle du *Perceptual Crossing* peut suffire à générer un début de sentiment de présence sociale envers un agent artificiel, même sans indices anthropomorphes ni communication symbolique. Ce résultat a des implications importantes pour la conception de robots ou d'agents artificiels puisqu'il suggère que la présence sociale ne dépend pas nécessairement du caractère anthropomorphe, d'une quelconque apparence visuelle ou de capacités conversationnelles sophistiquées, mais peut émerger notamment dans un environnement comme celui-ci à partir de schémas simples d'interaction et de feedback en tout-ou-rien. Cependant, la quantité de contacts ne semble pas déterminer nécessairement le sentiment de présence : multiplier les échanges avec un agent ne garantit pas un sentiment de présence sociale. Transposé à la HRI, un robot très réactif et qui tente de générer beaucoup d'échanges sans mémoire à long terme de l'interaction ne serait pas nécessairement perçu comme plus socialement présent. La simple réactivité locale (répondre rapidement à chaque action du partenaire) ne suffit pas, des mécanismes de coordination à long terme, d'anticipation et de synchronisation adaptative sont indispensables pour l'agent.

Ensuite, la Partie III établit des implications concrètes pour le paramétrage temporel des robots sociaux dans les conversations. Nos résultats identifient un délai de réponse optimal autour de ≈700 ms pour des questions fermées (oui/non), significativement plus long que les ≈200 ms observés dans les conversations humaines (Stivers et al., 2009). Ce décalage suggère soit que les humains développent des attentes temporelles spécifiques pour les robots, soit que la méthodologie basée sur des jugements explicites (plutôt que sur des échanges naturels) génère des biais de mesure. De plus, la tolérance aux écarts temporels varie selon le style de communication du robot. Les participants tolèrent davantage les variations de délai avec un robot au style soumis ou enfantin qu'avec un robot au style autoritaire ou neutre. Cette observation implique que les concepteurs peuvent moins s'inquiéter de certains robots quand ils présenteront un style de communication enfantin ou soumis, puisqu'un robot « doux » bénéficie d'une plus grande flexibilité temporelle, tandis qu'un robot perçu comme plus dominant doit respecter des attentes temporelles plus strictes. De même, concernant le délai, même des écarts temporels importants (200 vs. 1500 ms) par rapport au délai optimal n'affectent pas significativement la perception sociale du robot lorsque le délai reste constant tout au long de l'interaction, la perception et l'évaluation du robot semblent plutôt être plus affectées par le style de communication plutôt que s'il est trop lent ou trop rapide à répondre.

La Partie IV révèle que les humains détectent l'incohérence et l'incongruence entre la capacité d'un robot et son discours. Cependant, les humains manifestent dans l'écrasante majorité des croyances inattendues avec ces agents : malgré la preuve et évidence visuelle du contraire, ils peuvent indiquer qu'il se pourrait que le robot sans bras ni jambes qui leur est présenté puisse posséder des bras ou des jambes cachés, ou même encore que le robot possède la capacité de ressentir des émotions. Malgré le fait que leur cerveau a détecté une incohérence, ils croient le robot quand même s'il leur a prétendu qu'il a pu faire telle chose ou ressentir telle chose. Cela souligne une réelle flexibilité à résoudre l'incongruité et cela souligne pour les concepteurs l'importance de la cohérence entre morphologie, capacités effectives et discours du robot puisque les humains peuvent détecter l'incohérence mais peuvent parfois trouver un moyen de résoudre l'incongruité grâce à la croyance (résolvant ainsi l'incohérence). Les concepteurs devraient veiller à ce que les énoncés produits par le robot soient

compatibles avec ses caractéristiques observables. Un robot ne devrait pas parler de choses fausses le concernant ou impossibles : un robot sans bras ou sans jambes ne devrait pas déclarer serrer la main ou monter des escaliers. De même, pour ne pas induire les gens en erreur ou qu'ils soient amenés à attribuer aux robots des capacités qu'il n'a pas, le fait de donner au robot la possibilité d'employer des phrases avec des expressions liées aux émotions devrait être accompagné d'une communication transparente sur leur nature fonctionnelle (signaux sociaux destinés à influencer le comportement humain) mais ne pas employer une prétendue expérience subjective interne.

13.3 Implications éthiques

Suite aux travaux menés dans cette thèse, une dernière implication nous apparaît essentielle et concerne la communication claire et transparente sur les capacités réelles des robots, ainsi que sur les objectifs poursuivis par l'ajout dans leur conception de comportements ou expressions verbales faisant appel à des états internes comme les émotions. Ceci aurait pour objectif de gérer les attentes des utilisateurs ou même le risque de mésusage voire de désinformation.

De manière plus large, il est nécessaire d'éduquer et d'informer le public sur l'utilisation, les limites et les capacités réelles de ces technologies. Une telle démarche favoriserait une appropriation plus critique des outils, réduirait les risques liés à un mésusage et encouragerait un rapport plus lucide et responsable aux agents artificiels. Les croyances et attentes humaines associées aux robots doivent être comprises comme des variables dynamiques, susceptibles d'évoluer selon les contextes culturels, médiatiques ou éducatifs. Les recherches futures devraient donc examiner comment ces représentations se construisent et comment elles peuvent être modulées.

Si certaines attentes s'avèrent difficilement compatibles avec les limites technologiques actuelles, une réorientation stratégique de la conception des robots sociaux pourrait s'imposer. Comme le rappellent Breazeal (2003b) (voir aussi Levenson,1994), les expressions émotionnelles jouent avant tout un rôle de signal social visant à influencer le comportement d'autrui. Les conceptions futures gagneraient ainsi à privilégier des signaux contextuellement pertinents, adaptés aux objectifs des interactions, plutôt que de chercher à imiter fidèlement les

expressions humaines. En faisant le choix de concevoir des robots dont la nature non-humaine est clairement perçue et signalée et en valorisant leurs spécificités propres plutôt que de chercher à les rendre anthropomorphes à outrance, il devient alors possible de développer des formes d'interaction mieux adaptées aux capacités réelles de ces agents. En bref, accueillir à bras ouverts la nature synthétique des robots afin d'introduire de nouvelles formes d'interaction, rendues possibles uniquement parce que le robot n'est pas humain, pourrait constituer une voie précieuse pour l'avenir.

Conclusion

Conclusion

En conclusion, cette thèse cherchait à explorer comment les humains perçoivent, se connectent et réagissent aux agents sociaux artificiels, en particulier les robots sociaux. Trois axes complémentaires ont structuré ce travail : (1) l'étude de l'impact de la consigne sur la présence sociale et la structure de la dynamique avec des agents artificiels dans un environnement minimaliste d'interaction, (2) l'investigation des attentes temporelles dans les échanges verbaux avec un robot, et (3) l'analyse des frontières cognitives face au discours d'un robot dépassant ses capacités.

Cette thèse propose des paradigmes expérimentaux transférables, en étendant le paradigme de la N400 à la robotique sociale et en mobilisant l'analyse psychophysique des délais pour étudier la temporalité des échanges entre humains et robots. Elle apporte également des repères opérationnels pour la conception, tels qu'un délai de réponse de référence modulable selon le style de communication et la nécessité de maintenir une cohérence entre les capacités du robot et son discours. L'ensemble des résultats constitue ainsi une contribution intégrée à la compréhension et à la conception des interactions humain-robot.

Premièrement, la présence sociale peut émerger avec des agents artificiels lors d'interactions dans un environnement minimaliste fondé sur la coordination sensorimotrice et des feedback audiovisuels minimaux, même sans incitation sociale explicite ni indices visuels ou anthropomorphes. Les structures temporelles de l'activité conjointe varient selon les conditions, mais ne se traduisent pas nécessairement dans les jugements subjectifs : multiplier les contacts ne suffit pas à générer un sentiment de présence sociale. En revanche, une consigne explicitement sociale et invitant à l'interaction favorise davantage l'émergence de cette présence perçue.

Deuxièmement, les humains semblent développer des attentes temporelles spécifiques envers les robots. En effet, contrairement à ce qui est observé dans les échanges entre les humains, les participants ne semblent pas attribuer de

signification sociale aux délais de réponse des robots. Un délai d'environ 700 ms est perçu comme optimal pour des questions fermées, soit trois à sept fois plus long que dans les conversations humaines. La tolérance aux écarts à ce délai varie selon le style de communication : elle est plus large pour les styles perçus comme « doux ». Toutefois, même lorsque le robot s'écarte fortement du délai optimal, l'évaluation sociale globale reste inchangée ; seule la manière dont il communique influence réellement la perception.

Troisièmement, le cerveau humain détecte automatiquement les incongruités dans le discours d'un robot, qu'il s'agisse d'actions impossibles ou d'énoncés évoquant des émotions. Néanmoins, si les croyances individuelles peuvent atténuer la détection des incongruités physiques, elles n'influencent pas celles liées aux émotions. Ceci révèle une frontière cognitive rigide autour de l'expérience subjective attribuée aux agents artificiels. Ainsi, les humains semblent « réparer » cognitivement les impossibilités physiques en supposant des capacités cachées, tout en maintenant une inflexibilité absolue quant à la possibilité qu'un robot puisse réellement faire l'expérience des émotions.

En conclusion, cette thèse a exploré les mécanismes par lesquels les humains se connectent, s'ajustent et évaluent les agents artificiels, en comparaison avec les interactions humain-humain. En mobilisant une approche interdisciplinaire intégrant des méthodes comportementales, psychophysiques et neurophysiologiques, elle a permis d'examiner certaines des attentes, spécificités et limites manifestées envers les robots sociaux.

Bibliographie

Abu-Akel, A. M., Apperly, I. A., Wood, S. J., & Hansen, P. C. (2020). Re-imaging the intentional stance. *Proceedings of the Royal Society B : Biological Sciences*, *287*(1925), 20200244. https://doi.org/10.1098/rspb.2020.0244

Abubshait, A., & Wykowska, A. (2020). Repetitive Robot Behavior Impacts Perception of Intentionality and Gaze-Related Attentional Orienting. *Frontiers in Robotics and AI*, *7*. https://doi.org/10.3389/frobt.2020.565825

Ackerman, E. (2018, mars 12). *Jibo is probably totally dead now - IEEE spectrum* [Jibo is probably totally dead now - IEEE spectrum]. Récupérée décembre 5, 2019, à partir de https://spectrum.ieee.org/jibo-is-probably-totally-dead-now

Aguiar, N. R., Stoess, C. J., & Taylor, M. (2012). The development of children's ability to fill the gaps in their knowledge by consulting experts. *Child Development*, *83*(4), 1368-1381. https://doi.org/10.1111/j.1467-8624.2012.01782.x

Al Moubayed, S., & Skantze, G. (2011). Turn-taking Control Using Gaze in Multiparty Human-Computer Dialogue : Effects of 2D and 3D Displays. *Proceedings of the International Conference on Audio-Visual Speech Processing 2011*, 99-102. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-52205

Aldebaran. (2024). *Aldebaran - nao6* [Aldebaran]. Récupérée mai 9, 2023, à partir de https://aldebaran.com/en/nao6/

Andrist, S., Ziadee, M., Boukaram, H., Mutlu, B., & Sakr, M. (2015). Effects of culture on the credibility of robot speech : A comparison between english and arabic. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 157-164. https://doi.org/10.1145/2696454.2696464

Argyle, M., & Dean, J. (1965). Eye-Contact, Distance and Affiliation. *Sociometry*, *28*(3), 289. https://doi.org/10.2307/2786027

Asch, S. E. (1956). Studies of independence and conformity : I. a minority of one against a unanimous majority. *Psychological Monographs : General and Applied*, *70*(9), 1-70. https://doi.org/10.1037/h0093718

Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*(1), 32-47. https://doi.org/10.1016/j.newideapsych.2007.12.002

Baraka, K., Alves-Oliveira, P., & Ribeiro, T. (2020). An Extended Framework for Characterizing Social Robots. In C. Jost, B. Le Pévédic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, M. Grandgeorge & N. Mirnig (Éd.), *Human-Robot Interaction : Evaluation Methods and Their Standardization* (p. 21-64, T. 12). Springer International Publishing. https://doi.org/10.1007/978-3-030-42307-0_2

Barco, A., De Jong, C., Peter, J., Kühne, R., & Van Straten, C. L. (2020). Robot morphology and children's perception of social robots : An exploratory study. *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 125-127. https://doi.org/10.1145/3371382.3378348

Baron-Cohen, S. (2001). Theory of mind and autism : A review. In *Autism* (p. 169-184, T. 23). Elsevier.

Barone, P., Bedia, M. G., & Gomila, A. (2020). A Minimal Turing Test : Reciprocal Sensorimotor Contingencies for Interaction Detection. *Frontiers in Human Neuroscience*, *14*, 102. https://doi.org/10.3389/fnhum.2020.00102

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, *4*(1), 29-34. https://doi.org/10.1016/S1364-6613(99)01419-9

Barrett, J. L. (2004). *Why would anyone believe in God ?* AltaMira Press, a division of Rowman & Littlefield.

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, 591-594. https://doi.org/10.1109/ROMAN.2004.1374827

Bartneck, C. (2024, janvier 10). *Jibo is dead (again)* [Christoph bartneck, ph.d.]. Récupérée octobre 1, 2024, à partir de https://www.bartneck.de/2024/01/10/jibo-is-dead-again/

Bartneck, C., & Keijsers, M. (2020). The morality of abusing a robot. *Paladyn, Journal of Behavioral Robotics*, *11*(1), 271-283. https://doi.org/10.1515/pjbr-2020-0017

Bartneck, C., Nomura, T., Kanda, T., Tomohiro, S., & Kennsuke, K. (2005). A cross-cultural study on attitudes towards robots. In G. Salvendy (Éd.), *Procee-*

dings of the HCI international 2005. Lawrence Erlbaum Associates. https://doi.org/10.13140/RG.2.2.35929.11367

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2006). The influence of people's culture and prior experiences with aibo on their attitude towards robots. *AI & SOCIETY*, *21*(1), 217-230. https://doi.org/10.1007/s00146-006-0052-7

Baumann, A.-E., Goldman, E. J., Cobos, M.-G. M., & Poulin-Dubois, D. (2024). Do preschoolers trust a competent robot pointer ? *Journal of Experimental Child Psychology*, *238*, 105783. https://doi.org/10.1016/j.jecp.2023.105783

Bedia, M. G., Aguilera, M., Gómez, T., Larrode, D. G., & Seron, F. (2014). Quantifying long-range correlations and 1/f patterns in a minimal experiment of social interaction. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01281

Bethel, C. L., & Murphy, R. R. (2010). Review of human studies methods in HRI and recommendations. *International Journal of Social Robotics*, *2*(4), 347-359. https://doi.org/10.1007/s12369-010-0064-9

Biocca, F. (1997). The Cyborg's Dilemma : Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication*, *3*(2), JCMC324. https://doi.org/10.1111/j.1083-6101.1997.tb00070.x

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence : Review and suggested criteria. *Presence : Teleoperators and Virtual Environments*, *12*(5), 456-480. https://doi.org/10.1162/105474603322761270

BODDAC. (2025, juin 18). *Aldebaran, Bodacc A n° 20250115 du 18/06/2025, annonce n° 903, Jugement de conversion en liquidation judiciaire*. Récupérée juin 18, 2025, à partir de https://www.bodacc.fr/pages/annonces-commerciales-detail/?q.id=id:A20250115903

Bonnefon, J.-F., Dahl, E., & Holtgraves, T. M. (2015). Some but not all dispreferred turn markers help to interpret scalar terms in polite contexts. *Thinking & Reasoning*, *21*(2), 230-249. https://doi.org/10.1080/13546783.2014.965746

Bossi, F., Willemse, C., Cavazza, J., Marchesi, S., Murino, V., & Wykowska, A. (2020). The human brain reveals resting state activity patterns that are predictive of biases in attitudes toward robots. *Science Robotics*, *5*(46), eabb6652. https://doi.org/10.1126/scirobotics.abb6652

Breazeal, C. (2003a). Toward sociable robots. *Robotics and Autonomous Systems*, *42*(3), 167-175. https://doi.org/10.1016/s0921-8890(02)00373-1

Breazeal, C. (2003b). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1), 119-155. https://doi.org/10.1016/S1071-5819(03)00018-1

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language*, *43*(6), 1310-1337. https://doi.org/10.1017/S0305000915000689

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind : A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*(3), 314-325. https://doi.org/10.1006/nimg.2000.0612

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809-825. https://doi.org/10.1177/0022243719851788

Castelo, N., & Sarvary, M. (2022). Cross-cultural differences in comfort with humanlike robots. *International Journal of Social Robotics*, *14*(8), 1865-1873. https://doi.org/10.1007/s12369-022-00920-y

Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutcher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think ? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00103

Chaminade, T., Zecca, M., Blakemore, S.-J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., & Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE*, *5*(7), e11577. https://doi.org/10.1371/journal.pone.0011577

Chen, N., Liu, X., Zhai, Y., & Hu, X. (2023). Development and validation of a robot social presence measurement dimension scale. *Scientific Reports*, *13*(1), 2911. https://doi.org/10.1038/s41598-023-28817-4

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy : Gender effects in brain and behavior. *Neuros-*

cience and biobehavioral reviews, *46*, 604-627. https://doi.org/10.1016/j.neubiorev.2014.09.001

Clark, B. (2024, septembre 18). *Pittsburgh's digital dream labs accused of failing to fulfill over $2 million in orders, faces legal action* [Hoodline]. Récupérée octobre 3, 2024, à partir de https://hoodline.com/2024/09/pittsburgh-s-digital-dream-labs-accused-of-failing-to-fulfill-over-2-million-in-orders-faces-legal-action/

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, *19*(4), 360-379. https://doi.org/10.1111/j.0268-1064.2004.00263.x

Croes, E. A. J., & Antheunis, M. L. (2021). Can we be friends with mitsuku ? a longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, *38*(1), 279-300. https://doi.org/10.1177/0265407520959463

Cruz, A. A., Garcia, D. M., Pinto, C. T., & Cechetti, S. P. (2011). Spontaneous eye-blink activity. *The Ocular Surface*, *9*(1), 29-41. https://doi.org/10.1016/S1542-0124(11)70007-6

Cuijpers, R. H., & Van Den Goor, V. J. (2017). Turn-taking cue delays in human-robot communication. *WS-SIME+Barriers of Social Robotics*, 19-29. http://ceur-ws.org/Vol-2059/

Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in Human–Robot Co-evolution. *Frontiers in Psychology*, *9*, 468. https://doi.org/10.3389/fpsyg.2018.00468

Dautenhahn, K. (1995). Getting to know each other—artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, *16*(2), 333-356. https://doi.org/10.1016/0921-8890(95)00054-2

Dautenhahn, K. (1998). The art of designing socially intelligent agents : Science, fiction, and the human in the loop. *Applied Artificial Intelligence*, *12*(7), 573-617. https://doi.org/10.1080/088395198117550

Dautenhahn, K., Ogden, B., & Quick, T. (2002). From embodied to socially embedded agents - implications for interaction-aware robots. *Cognitive Systems Research*, *3*(3), 397-428. https://doi.org/10.1016/S1389-0417(02)00050-5

David, D. O., Costescu, C. A., Matu, S., Szentagotai, A., & Dobrean, A. (2020). Effects of a robot-enhanced intervention for children with ASD on teaching turn-taking skills. *Journal of Educational Computing Research*, *58*(1), 29-62. https://doi.org/10.1177/0735633119830344

De Bruin, L. C., & Kästner, L. (2012). Dynamic embodied cognition. *Phenomenology and the Cognitive Sciences*, *11*(4), 541-563. https://doi.org/10.1007/s11097-011-9223-1

De Castro Martins, C., Chaminade, T., & Cavazza, M. (2022). Causal Analysis of Activity in Social Brain Areas During Human-Agent Conversation. *Frontiers in Neuroergonomics*, *3*. https://doi.org/10.3389/fnrgo.2022.843005

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making : An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485-507. https://doi.org/10.1007/s11097-007-9076-9

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

Dennett, D. C. (1987). *The intentional stance*. MIT Press.

Denzin, N. K. (1970). *The research act in sociology : a theoretical introduction to sociological methods*. Butterworths.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition : Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893-1906. https://doi.org/10.1162/jocn.2008.21142

Di Paolo, E. A., Rohde, M., & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction ? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, *26*(2), 278-294. https://doi.org/10.1016/j.newideapsych.2007.07.006

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion : People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology : General*, *144*(1), 114-126. https://doi.org/10.1037/xge0000033

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal : The design and perception of humanoid robot heads. *Proceedings of the 4th conference on Designing interactive systems : processes,*

practices, methods, and techniques, 321-326. https://doi.org/10.1145/778712.778756

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177-190. https://doi.org/10.1016/S0921-8890(02)00374-3

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283-292. https://doi.org/10.1037/h0033031

Edlund, J., & Heldner, M. (2005). Exploring prosody in interaction control. *Phonetica*, *62*(2), 215-226. https://doi.org/10.1159/000090099

Ekstedt, E., & Skantze, G. (2020). TurnGPT : A transformer-based language model for predicting turn-taking in spoken dialog. *Findings of the Association for Computational Linguistics : EMNLP 2020*, 2981-2990. https://doi.org/10.18653/v1/2020.findings-emnlp.268

Ekstedt, E., & Skantze, G. (2022). Voice activity projection : Self-supervised learning of turn-taking events. *Interspeech 2022*, 5190-5194. https://doi.org/10.21437/Interspeech.2022-10955

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human : A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864-886. https://doi.org/10.1037/0033-295X.114.4.864

Ervin-Tripp, S. (1979). Children's verbal turn-taking. In E. Ochs & B. B. Schieffelin (Éd.), *Developmental Pragmatics* (p. 391-414). Academic Press.

Evers, V., Maldonado, H. C., Brodecki, T. L., & Hinds, P. J. (2008). Relational vs. group self-construal : untangling the role of national culture in HRI. *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 255-262. https://doi.org/10.1145/1349822.1349856

Fiore, S. M., Wiltshire, T. J., Lobato, E. J. C., Jentsch, F. G., Huang, W. H., & Axelrod, B. (2013). Toward understanding social cues and signals in human-robot interaction : effects of robot gaze and proxemic behavior. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00859

Fitrianie, S., Bruijnes, M., Richards, D., Abdulrahman, A., & Brinkman, W.-P. (2019). What are we measuring anyway ? : - a literature survey of questionnaires used in studies reported in the intelligent virtual agent conferences. *Pro-*

ceedings of the 19th ACM International Conference on Intelligent Virtual Agents, 159-161. https://doi.org/10.1145/3308532.3329421

Fitrianie, S., Bruijnes, M., Richards, D., Bönsch, A., & Brinkman, W.-P. (2020). The 19 unifying questionnaire constructs of artificial social agents : An IVA community analysis. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 1-8. https://doi.org/10.1145/3383652.3423873

Flanagan, T., Georgiou, N. C., Scassellati, B., & Kushnir, T. (2024). School-age children are more skeptical of inaccurate robots than adults. *Cognition*, *249*, 105814. https://doi.org/10.1016/j.cognition.2024.105814

Flessert, M. (2022). Pareidolia. In J. Vonk & T. K. Shackelford (Éd.), *Encyclopedia of Animal Cognition and Behavior* (p. 4953-4958). Springer International Publishing. https://doi.org/10.1007/978-3-319-55065-7_1771

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

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds : Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465-1477. https://doi.org/10.1162/0898929042304705

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531-534. https://doi.org/10.1016/j.neuron.2006.05.001

Froese, T., Iizuka, H., & Ikegami, T. (2014). Using minimal human-computer interfaces for studying the interactive development of social awareness. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01061

Froese, T., Zapata-Fonseca, L., Leenen, I., & Fossion, R. (2020). The Feeling Is Mutual : Clarity of Haptics-Mediated Social Perception Is Not Associated With the Recognition of the Other, Only With Recognition of Each Other. *Frontiers in Human Neuroscience*, *14*, 560567. https://doi.org/10.3389/fnhum.2020.560567

Funakoshi, K., Nakano, M., Kobayashi, K., Komatsu, T., & Yamada, S. (2010). Non-humanlike spoken dialogue : a design perspective. *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 176-184. Récupérée août 20, 2025, à partir de https://dl.acm.org/doi/abs/10.5555/1944506.1944537

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, 145-152. https://doi.org/10.1145/1349822.1349842

Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00751

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain : The mirror neuron system responds to human and robotic actions. *NeuroImage*, *35*(4), 1674-1684. https://doi.org/10.1016/j.neuroimage.2007.02.003

Gelman, S. A. (2003, mai 8). *The essential child : Origins of essentialism in everyday thought*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195154061.001.0001

Gigandet, R., Diana, M. C., Ouadada, K., & Nazir, T. A. (2024). Beyond explicit acknowledgment : Brain response evidence of human skepticism towards robotic emotions. *Robotics*, *13*(5), 67. https://doi.org/10.3390/robotics13050067

Gigandet, R., Dutoit, X., Li, B., Diana, M. C., & Nazir, T. A. (2023). The 'Eve effect bias' : Epistemic Vigilance and Human Belief in Concealed Capacities of Social Robots. *2023 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*, 15-20. https://doi.org/10.1109/ARSO56563.2023.10187469

Gigandet, R., & Nazir, T. A. (2025, février 4). Inferring Human Perception of Robots Through Event-Related Brain Potentials. In J. Seibt, P. Fazekas & O. S. Quick (Éd.), *Frontiers in Artificial Intelligence and Applications*. IOS Press. https://doi.org/10.3233/FAIA241497

Gordon, G. (1951). Observations upon the movements of the eyelids. *British Journal of Ophthalmology*, *35*(6), 339-351. https://doi.org/10.1136/bjo.35.6.339

Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, *25*(3), 601-634. https://doi.org/10.1016/j.csl.2010.10.003

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619-619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies : Mind perception and the uncanny valley. *Cognition*, *125*(1), 125-130. https://doi.org/10.1016/j.cognition.2012.06.007

Griffin, R., & Baron-Cohen, S. (2002). The Intentional Stance : Developmental and Neurocognitive Perspectives. In A. Brook & D. Ross (Éd.), *Daniel Dennett* (p. 83-116). Cambridge University Press.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & Mccall, C. (2007). Virtual humans and persuasion : The effects of agency and behavioral realism. *Media Psychology*, *10*(1), 1-22. https://doi.org/10.1080/15213260701300865

Guthrie, S. E. (1995). *Faces in the clouds : a new theory of religion*. Oxford University Press.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1481), 801-811. https://doi.org/10.1098/rstb.2007.2089

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the n300 and n400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*(8), 1339-1350. https://doi.org/10.1016/S1388-2457(02)00161-X

Harkins, W. E. (1962). *Karel Čapek*. Columbia University Press.

Harms, C., & Biocca, F. (2004). Internal consistency and reliability of the networked minds social presence measure. In M. Alcañiz Raya & B. Rey Solaz (Éd.), *Seventh annual international workshop on presence 2004* (p. 246-251). Universidad Politécnica de Valencia.

Haslam, N., Bastian, B., & Bissett, M. (2004). Essentialist Beliefs about Personality and Their Implications. *Personality and Social Psychology Bulletin*, *30*(12), 1661-1673. https://doi.org/10.1177/0146167204271182

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults : The almere model. *International Journal of Social Robotics*, *2*(4), 361-375. https://doi.org/10.1007/s12369-010-0068-5

Hegel, F., Muhl, C., Wrede, B., Hielscher-Fastabend, M., & Sagerer, G. (2009). Understanding Social Robots. *2009 Second International Conferences on Advances in Computer-Human Interactions*, 169-174. https://doi.org/10.1109/achi.2009.51

Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social cognition in the age of human–robot interaction. *Trends in Neurosciences*, *43*(6), 373-384. https://doi.org/10.1016/j.tins.2020.03.013

Herschbach, M. (2012). On the role of social interaction in social cognition : A mechanistic alternative to enactivism. *Phenomenology and the Cognitive Sciences*, *11*(4), 467-486. https://doi.org/10.1007/s11097-011-9209-z

Heyselaar, E. (2023). The CASA theory no longer applies to desktop computers. *Scientific Reports*, *13*(1), 19693. https://doi.org/10.1038/s41598-023-46527-9

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory : Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, *26*(6), 1508-1518. https://doi.org/10.1016/j.chb.2010.05.015

Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. *International Journal of Social Robotics*, *9*(1), 129-139. https://doi.org/10.1007/s12369-016-0380-9

Hough, J., & Schlangen, D. (2016). Investigating fluidity for human-robot interaction with real-time, real-world grounding strategies. *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 288-298. https://doi.org/10.18653/v1/W16-3637

Ikari, S., Sato, K., Burdett, E., Ishiguro, H., Jong, J., & Nakawake, Y. (2023). Religion-related values differently influence moral attitude for robots in the united states and japan. *Journal of Cross-Cultural Psychology*, *54*(6), 742-759. https://doi.org/10.1177/00220221231193369

ISO. (2021). ISO 8373 :2021 Robotique - Vocabulaire. https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-3:v1:fr

Jacob, R. G., Simons, A. D., Manuck, S. B., Rohay, J. M., Waldstein, S., & Gatsonis, C. (1989). The circular mood scale : A new technique of measuring ambulatory mood. *Journal of Psychopathology and Behavioral Assessment*, *11*(2), 153-173. https://doi.org/10.1007/BF00960477

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best : Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*(9), 757-758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Johnson, D., & Gardner, J. (2007). The media equation and team formation : Further evidence for experience as a moderator. *International Journal of Human-Computer Studies*, *65*(2), 111-124. https://doi.org/10.1016/j.ijhcs.2006.08.007

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22-63. https://doi.org/10.1016/0001-6918(67)90005-4

Kim, R. H., Moon, Y., Choi, J. J., & Kwak, S. S. (2014). The effect of robot appearance types on motivating donation. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 210-211. https://doi.org/10.1145/2559636.2563685

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events : The referential roles of words and the speakers who use them. *Cognition*, *87*(3), 179-208. https://doi.org/10.1016/S0010-0277(03)00002-7

Kühne, R., & Peter, J. (2023). Anthropomorphism in human–robot interactions : A multidimensional conceptualization. *Communication Theory*, *33*(1), 42-52. https://doi.org/10.1093/ct/qtac020

Kumazaki, H., Muramatsu, T., Yoshikawa, Y., Matsumoto, Y., Miyao, M., Ishiguro, H., Mimura, M., Minabe, Y., & Kikuchi, M. (2019). How the realism of robot is needed for individuals with autism spectrum disorders in an interview setting. *Frontiers in Psychiatry*, *10*, 486. https://doi.org/10.3389/fpsyt.2019.00486

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting : Finding meaning in the n400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences : Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205. https://doi.org/10.1126/science.7350657

Kwak, S. S. (2014). The Impact of the Robot Appearance Types on Social Interaction with a Robot and Service Evaluation of a Robot. *Archives of Design Research*. https://doi.org/10.15187/adr.2014.05.110.2.81

Lee, H. R., & Šabanović, S. (2014). Culturally variable preferences for robot design and use in south korea, turkey, and the united states. *Proceedings of the*

2014 ACM/IEEE international conference on Human-robot interaction, 17-24. https://doi.org/10.1145/2559636.2559676

Levenson, R. W. (1994). Human emotion : A functional view. In P. Ekman & R. J. Davidson (Éd.), *The nature of emotion : Fundamental questions* (p. 123-126). Oxford University Press.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00731

Li, B., Ajjaji, O., Gigandet, R., & Nazir, T. (2023). The body images of social robots. *2023 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*, 1-8. https://doi.org/10.1109/ARSO56563.2023.10187489

Li, J. J., Ju, W., & Reeves, B. (2017). Touching a Mechanical Body : Tactile Contact With Body Parts of a Humanoid Robot Is Physiologically Arousing. *Journal of Human-Robot Interaction*, *6*(3), 118. https://doi.org/10.5898/JHRI.6.3.Li

Li, L., Li, Y., Song, B., Shi, Z., & Wang, C. (2022). How human-like behavior of service robot affects social distance : A mediation model and cross-cultural comparison. *Behavioral Sciences*, *12*(7), 205. https://doi.org/10.3390/bs12070205

Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing jesus in toast : Neural and behavioral correlates of face pareidolia [Publisher : Elsevier BV]. *Cortex*, *53*, 60-77. https://doi.org/10.1016/j.cortex.2014.01.013

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation : People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103. https://doi.org/10.1016/j.obhdp.2018.12.005

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Maij, D. L. R., Van Schie, H. T., & Van Elk, M. (2019). The boundary conditions of the hypersensitive agency detection device : An empirical investigation of agency detection in threatening situations. *Religion, Brain & Behavior*, *9*(1), 23-51. https://doi.org/10.1080/2153599x.2017.1362662

Mar, R. A., Kelley, W. M., Heatherton, T. F., & Macrae, C. N. (2007). Detecting agency from the biological motion of veridical vs animated agents. *Social*

Cognitive and Affective Neuroscience, *2*(3), 199-205. https://doi.org/10.1093/scan/nsm011

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots ? *Frontiers in Psychology*, *10*, 450. https://doi.org/10.3389/fpsyg.2019.00450

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29-63. https://doi.org/10.1016/0010-0285(78)90018-X

Matzinger, T., Pleyer, M., & Żywiczyński, P. (2023). Pause length and differences in cognitive state attribution in native and non-native speakers. *Languages*, *8*(1), 26. https://doi.org/10.3390/languages8010026

Mehmood, K., Kautish, P., & Shah, T. R. (2024). Embracing digital companions : Unveiling customer engagement with anthropomorphic AI service robots in cross-cultural context. *Journal of Retailing and Consumer Services*, *79*, 103825. https://doi.org/10.1016/j.jretconser.2024.103825

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

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. (2010). "social" robots are psychological agents for infants : A test of gaze following. *Neural Networks*, *23*(8), 966-972. https://doi.org/10.1016/j.neunet.2010.09.005

Mobed, D. A. O., Wodehouse, A., & Maier, A. (2024). The aesthetics of robot design : Towards a classification of morphologies. *Proceedings of the Design Society*, *4*, 2413-2422. https://doi.org/10.1017/pds.2024.244

Monceaux, J. (2024, octobre 16). *Le robot compagnon de l'ICM autorisé en salle de radiothérapie : une avancée mondiale au service des enfants* [Enchanted Tools]. Récupérée octobre 16, 2024, à partir de https://enchanted-tools.notion.site/Media-Press-Room-a402fd1610f14f33a7125421b70915f0

Morewedge, C. K., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of Personality and Social Psychology*, *93*(1), 1-11. https://doi.org/10.1037/0022-3514.93.1.1

Morillo-Mendez, L., Stower, R., Sleat, A., Schreiter, T., Leite, I., Mozos, O. M., & Schrooten, M. G. S. (2023). Can the robot "see" what I see ? Robot gaze

drives attention depending on mental state attribution. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1215771

Naneva, S., Sarda Gou, M., Webb, T. L., & Prescott, T. J. (2020). A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*, *12*(6), 1179-1201. https://doi.org/10.1007/s12369-020-00659-4

Nass, C., Fogg, B., & Moon, Y. (1996). Can computers be teammates ? *International Journal of Human-Computer Studies*, *45*(6), 669-678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., & Moon, Y. (2000). Machines and mindlessness : Social responses to computers. *Journal of Social Issues*, *56*(1), 81-103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Conference companion on Human factors in computing systems - CHI '94*, 204. https://doi.org/10.1145/259963.260288

Nazir, T. A., Lebrun, B., & Li, B. (2023). Improving the acceptability of social robots : Make them look different from humans. *PLOS ONE*, *18*(11), e0287507. https://doi.org/10.1371/journal.pone.0287507

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE™. *L'Année psychologique*, *101*(3), 447-462. https://doi.org/10.3406/psy.2001.1341

Nomura, T., Kanda, T., Kidokoro, H., Suehiro, Y., & Yamada, S. (2016). Why do children abuse robots ? *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *17*(3), 347-369. https://doi.org/10.1075/is.17.3.02nom

Nomura, T., Suzuki, T., Kanda, T., Han, J., Shin, N., Burke, J., & Kato, K. (2008). What people assume about humanoid and animal-type robots : Cross-cultural analysis between japan, korea, and the united states. *International Journal of Humanoid Robotics*, *05*(1), 25-46. https://doi.org/10.1142/s0219843608001297

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *7*(3), 437-454. https://doi.org/10.1075/is.7.3.14nom

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence : Teleoperators and Virtual Environments*, *12*(5), 481-494. https://doi.org/10.1162/105474603322761289

Nussey, S. (2021). EXCLUSIVE : SoftBank shrinks robotics business, stops pepper production-sources. *Reuters*. Récupérée octobre 25, 2022, à partir de https://www.reuters.com/technology/exclusive-softbank-shrinks-robotics-business-stops-pepper-production-sources-2021-06-28/

Oldfield, R. (1971). The assessment and analysis of handedness : The edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Overgaard, S., & Michael, J. (2015). The interactive turn in social cognition research : A critique. *Philosophical Psychology*, *28*(2), 160-183. https://doi.org/10.1080/09515089.2013.827109

Palmer, J. A., Makeig, S., Kreutz-Delgado, K., & Rao, B. D. (2008). Newton method for the ICA mixture model. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1805-1808. https://doi.org/10.1109/ICASSP.2008.4517982

Papadopoulos, C., Castro, N., Nigath, A., Davidson, R., Faulkes, N., Menicatti, R., Khaliq, A. A., Recchiuto, C., Battistuzzi, L., Randhawa, G., Merton, L., Kanoria, S., Chong, N.-Y., Kamide, H., Hewson, D., & Sgorbissa, A. (2022). The CARESSES Randomised Controlled Trial : Exploring the Health-Related Impact of Culturally Competent Artificial Intelligence Embedded Into Socially Assistive Robots and Tested in Older Adult Care Homes. *International Journal of Social Robotics*, *14*(1), 245-256. https://doi.org/10.1007/s12369-021-00781-x

Papadopoulos, C., Hill, T., Battistuzzi, L., Castro, N., Nigath, A., Randhawa, G., Merton, L., Kanoria, S., Kamide, H., Chong, N.-Y., Hewson, D., Davidson, R., & Sgorbissa, A. (2020). The CARESSES study protocol : Testing and evaluating culturally competent socially assistive robots among older adults residing in long term care homes through a controlled experimental trial. *Archives of Public Health*, *78*(1). https://doi.org/10.1186/s13690-020-00409-y

Paro. (2021). *PARO - Le robot Phoque assistant des soignants - Alzheimer - Polyhandicap - soins douloureux* [Paro]. Récupérée octobre 22, 2023, à partir de https://www.phoque-paro.fr/phoque-paro/

Perez-Osorio, J., & Wykowska, A. (2020). Adopting the intentional stance toward natural and artificial agents. *Philosophical Psychology*, *33*(3), 369-395. https://doi.org/10.1080/09515089.2019.1688778

Qin, X., Chen, C., Yam, K. C., Cao, L., Li, W., Guan, J., Zhao, P., Dong, X., & Lin, Y. (2022). Adults still can't resist : A social robot can induce normative conformity. *Computers in Human Behavior*, *127*, 107041. https://doi.org/10.1016/j.chb.2021.107041

Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587-595. https://doi.org/10.1016/j.chb.2008.12.025

Raux, A., & Eskenazi, M. (2008). Optimizing endpointing thresholds using dialogue features in a spoken dialogue system. *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, 1-10. https://doi.org/10.3115/1622064.1622066

Reeves, B., & Nass, C. I. (1996). *The media equation : How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Reid, V. M., & Striano, T. (2008). N400 involvement in the processing of action sequences. *Neuroscience Letters*, *433*(2), 93-97. https://doi.org/10.1016/j.neulet.2007.12.066

Roesler, E., Manzey, D., & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics*, *6*(58). https://doi.org/10.1126/scirobotics.abj5425

Roselli, C., Lapomarda, L., & Datteri, E. (2025). How culture modulates anthropomorphism in human-robot interaction : A review. *Acta Psychologica*, *255*, 104871. https://doi.org/10.1016/j.actpsy.2025.104871

Roselli, C., Navare, U. P., Ciardo, F., & Wykowska, A. (2024). Type of education affects individuals' adoption of intentional stance towards robots : An EEG study. *International Journal of Social Robotics*, *16*(1), 185-196. https://doi.org/10.1007/s12369-023-01073-2

Rosenthal-von Der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior*, *33*, 201-212. https://doi.org/10.1016/j.chb.2014.01.004

Roubroeks, M., Ham, J., & Midden, C. (2011). When artificial social agents try to persuade people : The role of social agency on the occurrence of psychological reactance. *International Journal of Social Robotics*, *3*(2), 155-165. https://doi.org/10.1007/s12369-010-0088-1

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence : a modern approach* (Fourth Edition). Pearson.

Šabanović, S., Bennett, C. C., & Lee, H. R. (2014). Towards culturally robust robots : A critical social perspective on robotics and culture. *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction (HRI) Workshop on Culture-Aware Robotics (CARS)*, *2014*.

Salem, M., Ziadee, M., & Sakr, M. (2014). Marhaba, how may i help you ?: effects of politeness and culture on robot acceptance and anthropomorphization. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 74-81. https://doi.org/10.1145/2559636.2559683

Salomons, N., Van Der Linden, M., Strohkorb Sebo, S., & Scassellati, B. (2018). Humans conform to robots : Disambiguating trust, truth, and conformity. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 187-195. https://doi.org/10.1145/3171221.3171282

Saunderson, S. P., & Nejat, G. (2021). Persuasive robots should avoid authority : The effects of formal and real authority on persuasion in human-robot interaction. *Science Robotics*, *6*(58), eabd5186. https://doi.org/10.1126/scirobotics.abd5186

Scherer, K. R. (2005). What are emotions ? and how can they be measured ? *Social Science Information*, *44*(4), 695-729. https://journals.sagepub.com/doi/10.1177/0539018405058216

Shibata, T. (2004). An overview of human interactive robots for psychological enrichment. *Proceedings of the IEEE*, *92*(11), 1749-1758. https://doi.org/10.1109/jproc.2004.835383

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

Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., & Hagita, N. (2009). How quickly should a communication robot respond ? delaying strategies and habituation effects. *International Journal of Social Robotics*, *1*(2), 141-155. https://doi.org/10.1007/s12369-009-0012-8

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Skantze, G. (2021). Turn-taking in conversational systems and human-robot interaction : A review. *Computer Speech & Language*, *67*, 101178. https://doi.org/10.1016/j.csl.2020.101178

Skantze, G., & Irfan, B. (2025). Applying general turn-taking models to conversational human-robot interaction. *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*, 859-868. https://dl.acm.org/doi/10.5555/3721488.3721593

Smith, D. H., & Zeller, F. (2017). The death and lives of hitchBOT : The design and implementation of a hitchhiking robot. *Leonardo*, *50*(1), 77-78. https://doi.org/10.1162/LEON_a_01354

Spatola, N., Kühnlenz, B., & Cheng, G. (2021a). Perception and evaluation in human–robot interaction : The human–robot interaction evaluation scale (HRIES)—a multicomponent approach of anthropomorphism. *International Journal of Social Robotics*, *13*(7), 1517-1539. https://doi.org/10.1007/s12369-020-00667-4

Spatola, N., Marchesi, S., & Wykowska, A. (2021b). The Intentional Stance Test-2 : How to Measure the Tendency to Adopt Intentional Stance Towards Robots. *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.666586

Spatola, N., Marchesi, S., & Wykowska, A. (2022). Different models of anthropomorphism across cultures and ontological limits in current frameworks the integrative framework of anthropomorphism. *Frontiers in Robotics and AI*, *9*. https://doi.org/10.3389/frobt.2022.863319

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359-393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592. https://doi.org/10.1073/pnas.0903616106

Strömbergsson, S., Hjalmarsson, A., Edlund, J., & House, D. (2013). Timing responses to questions in dialogue. *Interspeech 2013*, 2584-2588. https://doi.org/10.21437/Interspeech.2013-581

Sun, H. (2012). *Cross-cultural technology design : creating culture-sensitive technology for local users*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744763.001.0001

Sundar, S. S., & Kim, J. (2019). Machine heuristic : When we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-9. https://doi.org/10.1145/3290605.3300768

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., & Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*, *5*(1), 15924. https://doi.org/10.1038/srep15924

Tan, H., Wang, D., & Sabanovic, S. (2018). Projecting Life Onto Robots : The Effects of Cultural Factors and Design Type on Multi-Level Evaluations of Robot Anthropomorphism. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 129-136. https://doi.org/10.1109/ROMAN.2018.8525584

Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, *119*(4), e2116915119. https://doi.org/10.1073/pnas.2116915119

Thellman, S., Silvervarg, A., & Ziemke, T. (2017). Folk-Psychological Interpretation of Human vs. Humanoid Robot Behavior : Exploring the Intentional Stance toward Robots. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01962

Thórisson, K. R. (1996). *Communicative humanoids : a computational model of psychosocial dialogue skills* [thèse de doct., Massachusetts Institute of Technology].

Trovato, G., Zecca, M., Sessa, S., Jamone, L., Ham, J., Hashimoto, K., & Takanishi, A. (2013). Cross-cultural study on human-robot greeting interaction : acceptance and discomfort by Egyptians and Japanese. *Paladyn, Journal of Behavioral Robotics*, *4*(2). https://doi.org/10.2478/pjbr-2013-0006

Tuomi, A., Tussyadiah, I. P., & Hanna, P. (2021). Spicing up hospitality service encounters : The case of pepper™. *International Journal of Contemporary Hospitality Management*, *33*(11), 3906-3925. https://doi.org/10.1108/ijchm-07-2020-0739

Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., & Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in Neurorobotics*, *7*. https://doi.org/10.3389/fnbot.2013.00019

Vallacher, R. R., & Nowak, A. (2008, novembre 10). The Dynamics of Human Experience : Fundamentals of Dynamical Social Psychology. In S. J. Guastello, M. Koopmans & D. Pincus (Éd.), *Chaos and Complexity in Psychology* (1re éd., p. 370-401). Cambridge University Press. https://www.cambridge.org/core/product/identifier/CBO9781139058544A119/type/book_part

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse : Evidence from the n400. *Journal of Cognitive Neuroscience*, *11*(6), 657-671. https://doi.org/10.1162/089892999563724

van Berkum, J. J. A., van Den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580-591. https://doi.org/10.1162/jocn.2008.20054

van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse ? evidence from the n400 effect. *Cognitive Brain Research*, *17*(3), 701-718. https://doi.org/10.1016/S0926-6410(03)00196-4

Vanderborght, M., & Jaswal, V. K. (2009). Who knows best ? preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development*, *18*(1), 61-71. https://doi.org/10.1002/icd.591

van Elk, M., van Schie, H., & Bekkering, H. (2008). Semantics in action : An electrophysiological study on the use of semantic knowledge for action. *Journal of Physiology-Paris*, *102*(1), 95-100. https://doi.org/10.1016/j.jphysparis.2008.03.011

Varela, F. J., Rosch, E., & Thompson, E. (1991, septembre 26). *The embodied mind : Cognitive science and human experience*. The MIT Press. https://doi.org/10.7551/mitpress/6730.001.0001

Vollmer, A.-L., Read, R., Trippas, D., & Belpaeme, T. (2018). Children conform, adults resist : A robot group induced peer pressure on normative social conformity. *Science Robotics*, *3*(21), eaat7111. https://doi.org/10.1126/scirobotics.aat7111

Von Der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). "it doesn't matter what you are!" explaining social effects of agents and avatars. *Computers in Human Behavior*, *26*(6), 1641-1650. https://doi.org/10.1016/j.chb.2010.06.012

Warren, Z. E., Zheng, Z., Swanson, A. R., Bekele, E., Zhang, L., Crittendon, J. A., Weitlauf, A. F., & Sarkar, N. (2015). Can robotic interaction improve joint attention skills? *Journal of Autism and Developmental Disorders*, *45*(11), 3726-3734. https://doi.org/10.1007/s10803-013-1918-4

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383-388. https://doi.org/10.1016/j.tics.2010.05.006

Wiener, M., & Mehrabian, A. (1968). *Language within language : immediacy, a channel in verbal communication*. Appleton-Century-Crofts.

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean : How attentional selection is shaped by ascribing intentions to others. *PLoS ONE*, *7*(9), e45391. https://doi.org/10.1371/journal.pone.0045391

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS ONE*, *9*(4), e94339. https://doi.org/10.1371/journal.pone.0094339

Xu, K., Chen, M., & You, L. (2023). The hitchhiker's guide to a credible and socially present robot : Two meta-analyses of the power of social cues in human–robot interaction. *International Journal of Social Robotics*, *15*(2), 269-295. https://doi.org/10.1007/s12369-022-00961-3

Zarubica, S., & Bendel, O. (2024). Pepper as a learning partner in a children's hospital. *Social Robotics*, 15-26. https://doi.org/10.1007/978-981-99-8718-4_2

Zhang, Y., Song, W., Tan, Z., Zhu, H., Wang, Y., Lam, C. M., Weng, Y., Hoi, S. P., Lu, H., Man Chan, B. S., Chen, J., & Yi, L. (2019). Could social robots facilitate children with autism spectrum disorders in learning distrust and deception ? *Computers in Human Behavior*, *98*, 140-149. https://doi.org/10.1016/j.chb.2019.04.008

Annexes

Table des annexes

A Annexes aux Chapitres 7 et 8 297
B Annexes aux Chapitres 9 et 10 318

Annexe A

Annexes aux Chapitres 7 et 8

A.1 Chapitre 7 : Script du robot selon le style de communication

Cette annexe rapporte les détails des scripts du robot utilisés lors de l'expérimentation présentée au Chapitre 7. Pour rappel, l'expérience était en Anglais.

A.1.1 Condition Autoritaire

Introduction

> « *I am Lou, a social robot designed for human interaction. Now, listen carefully. Here are your instructions : You will watch 150 videos of my interactions with a person. And, these are divided into 5 blocks of 30. Immediately after each video, you will evaluate my Délai. Use* J *to indicate a delay that's too short,* L *for a delay that's too long. You have a strict 10 second window to input your evaluation after each video. Prepare, immediate task initiation. Begin now.* »

Pauses

— **Pause 1 :** « *You have completed the first block of 30 videos. Now, take precisely 45 seconds maximum to rest. Remember :* J *for a delay that's too short,* L *for too long ones. Your input is essential for this experiment. Prepare to resume immediately after the break.* »

— **Pause 2 :** « *Two blocks completed. 60 videos watched. Take your 45-second break now. Your task remains unchanged :* J *for too fast,* L *for too slow. Maintain your focus. The experiment will resume shortly.* »

— **Pause 3 :** « *Halfway point reached. 90 videos completed. Again, you have 45 seconds to rest. Maximum. Do not forget :* J *signifies too short,* L *indicates a too long delay. Your sustained concentration is mandatory.* »

— **Pause 4 :** « *120 videos completed. This is your final break. Use these 45 seconds wisely. After this, you will complete the last set of videos. Your unwavering attention is crucial. Prepare to give your fullest effort for the concluding block.* »

A.1.2 Condition Soumis

Introduction

> « *Hello... I'm Lou, a robot assistant... I hope I won't bother you too much... If it's alright with you, we're about to start the experiment. There will be 150 videos of me interacting with a person. They're in 5 blocks of 30, with breaks in between. After each video, if it's not too much to ask and if you don't mind... could you indicate if you think I responded too quickly with* J*, or too slowly with* L *?. You'll have 10 seconds each time, I hope that's enough. Would you be ready to start, if this is okay with you ?* »

Pauses

— **Pause 1 :** « *You've... you've finished the first 30 videos. Thank you so much for your help. If you don't mind, I'd like to tell you that you can pause for 45 seconds. I hope I'm not being a bother, but please remember to use* J *if I'm too fast and* L *if I'm too slow. Your input is really helpful, if you don't mind giving it. We'll continue soon, if that's okay with you.* »

— **Pause 2 :** « *Hi... you've completed 60 videos now, which is 2 out of 5 blocks... I really hope I'm not too tiring for you. Pl...Pl... Please take another 45-second break. And... also, if you could keep using* J *for too fast and* L *for too slow, I'd really appreciate it. Your help means a lot. We will go on when you're ready.* »

— **Pause 3 :** « *We...We're halfway now, with 90 videos done. You're doing great, if you don't mind me saying. Here's another 45-second break for you. I hope you're not getting bored of my reminders but...remember,* J *is for too fast and* L *for too slow. We will continue soon, if that's alright.* »

— **Pause 4 :** « *You... You've done 120 videos, which is amazing. This is the last break, if that's okay. Please take 45 seconds to rest. After this, if you're up for it, we have the last 30 videos. I really appreciate your focus, I hope you're ready for the final part, but no pressure.* »

A.1.3 Condition Enfantin

Introduction

« *Hi there, new friend ! I'm Lou, a social robot that can interact with humans ! Guess what ? You're going to do something cool ! You'll watch lots of short videos where I talk with a human. It's like 5 big rounds with tiny breaks in between ! After each video, tell me if you think I answered too quickly by pressing* J*, or too slowly with* L*. You've got 10 whole seconds each time ! Ready to have a blast ?* »

Pauses

— **Pause 1 :** « *Wow ! You've watched 30 whole videos ! You're amazing ! Now, let's have a teeny-tiny break for 45 seconds. Remember :* J *if too fast,* L *if I'm too slow ! You helped us so much ! Get ready for more fun soon !* »

— **Pause 2 :** « *Holy moly ! You've seen 60 videos now - that's 2 big rounds ! Time for another cool 45-second break. Keep doing our awesome* J *and* L *task, okay ? You're doing great ! Let's keep the fun rolling !* »

— **Pause 3 :** « *Guess what ? We're halfway done ! 90 videos - can you believe it ? You deserve another 45-second break ! Don't forget :* J *if too fast,* L *if I'm too slow ! Ready to rock the next part ?* »

— **Pause 4 :** « *Whoa ! 120 videos ! You're impressive ! Well my friend, this is our last teeny break. Continue to evaluate my Délai with* J *when too fast and* L *for too slow. In 45 seconds, we'll start our final adventure - 30 more videos ! Let's go !!!* »

A.1.4 Condition Neutre (et Rideau)

Introduction

« *Hello, I'm Lou, an assistant designed to interact with humans. We're*

about to start the experiment. Here's what you'll do : you'll watch 150 videos of interactions between a human and me. These videos are divided into 5 blocks, with short breaks in between each block. After each video, judge whether my Délai seems appropriate. Press J *if it's too fast, or* L *if it's too slow. You'll have 10 seconds to respond each time. Are you ready to start ?* »

Pauses

— **Pause 1 :** « *You have completed the first block of 30 videos. Well done. You can now take a 45-second break. Remember, your task is to evaluate my Délai using* J*, for too fast and* L*, for too slow. Your input is valuable for this experiment. The next block will begin shortly.* »

— **Pause 2 :** « *You have now completed 60 videos, which is 2 out of 5 blocks. Again, you can take a 45-second break. Remember :* J *for too fast,* L *for too slow. Your consistent evaluation is important. The experiment will continue soon.* »

— **Pause 3 :** « *90 videos have been completed, marking the halfway point. Here is a third 45-second break. Continue to evaluate my Délai with* J *when too fast and* L *for too slow. Your ongoing attention is appreciated. The experiment will continue shortly.* »

— **Pause 4 :** « *You have completed 120 videos, which is 4 out of 5 blocks. This is the final 45-second break. After this, we will proceed with the last set of 30 videos. Your sustained focus has been crucial. Prepare for the final block.* »

A.1.5 En commun

Vidéo de familiarisation 1

1. **ROBOT :** « *Now that you've received the instructions, let's see what the experiment actually looks like in practice. In each video, a human will ask me a question, and I will answer with either YES or NO. Remember, this is when you'll have to evaluate and decide if I took too little or too much time to respond. To get familiar with the experiment interface, you'll do 3 practice trials where you'll respond just like in the actual experiment. For this first practice, press either J or L after my response. Let's begin.* »

2. **HUMAN :** « *Are we starting the familiarization phase ?* »
3. **ROBOT :** « *Yes* »

Vidéo de familiarisation 2

1. **ROBOT :** « *For this second practice trial, which is an ATTENTION CHECK please press L - indicating that I took too much time to answer.* »
2. **HUMAN :** « *Is it the end of the familiarization phase ?* »
3. **ROBOT :** « *No* »

Vidéo de familiarisation 3

1. **ROBOT :** « *Now for the final practice trial, please press J - indicating that I took too little time to answer.* »
2. **HUMAN :** « *Will there be any more practice trial after that ?* »
3. **ROBOT :** « *No* »

A.2 Chapitre 7 : Liste des questions posées par l'humain au robot

Cette annexe rapporte l'ensemble des 150 questions posées par un humain au robot dans les vidéos de la phase principale de l'étude présentée au Chapitre 7. Pour rappel, cette étude a été menée en Anglais. À chaque question de l'humain, le robot répondait par « oui » ou par « non ».

format : question de l'humain $\rightarrow$ réponse du robot

— *Have you ever been to a concert ?* $\rightarrow$ *No*

— *Are you a robot ?* $\rightarrow$ *Yes*

— *Can you dream ?* $\rightarrow$ *No*

— *Can you recognize faces ?* $\rightarrow$ *Yes*

— *Do you need to sleep ?* $\rightarrow$ *No*

— *Are you equipped with a GPS ?* $\rightarrow$ *Yes*

— *Have you ever been sick ?* $\rightarrow$ *No*

— *Can you speak multiple languages ?* → *Yes*
— *Have you ever been to the zoo ?* → *No*
— *Is your body mainly white ?* → *Yes*
— *Do you have any piercings ?* → *No*
— *Is the snow white ?* → *Yes*
— *Can you predict the future ?* → *No*
— *Can you take part in role-playing games ?* → *Yes*
— *Can you feel pain ?* → *No*
— *Do you need regular maintenance ?* → *Yes*
— *Can you detect infrasounds ?* → *No*
— *If I pour water on you, will you short circuit ?* → *Yes*
— *Is your system code open source ?* → *No*
— *Do you know how to count from 1 to 10 ?* → *Yes*
— *Are you fitted with motion sensors ?* → *No*
— *Is a galloping horse always faster than a tortoise ?* → *Yes*
— *Do you have air pollution sensors ?* → *No*
— *Do you often take part in scientific experiments ?* → *Yes*
— *Can you experience nostalgia ?* → *No*
— *Have you ever attended a scientific conference ?* → *Yes*
— *Are you capable of having beliefs ?* → *No*
— *Can you process information faster than humans ?* → *Yes*
— *Are you able to swim ?* → *No*
— *If I jump into a puddle, will my shoes be wet ?* → *Yes*
— *Do you have arms ?* → *No*
— *Can you learn from your interactions ?* → *Yes*
— *Do you often climb the stairs ?* → *No*
— *Is the exterior of your body mainly made of plastic ?* → *Yes*
— *Can you experience emotions ?* → *No*

— *Are you able to make jokes ?* → *Yes*
— *Do you believe in a God ?* → *No*
— *Do you know how to write JavaScript code ?* → *Yes*
— *Do you need to eat ?* → *No*
— *Do you have access to the internet ?* → *Yes*
— *Do you have any allergies ?* → *No*
— *Have you ever been to a retirement home ?* → *Yes*
— *Have you ever visited a space station ?* → *No*
— *If I showed you some fossils, could you identify them ?* → *Yes*
— *Can you climb trees ?* → *No*
— *Do you have a physical body ?* → *Yes*
— *Can you knit ?* → *No*
— *Do humans need oxygen to breathe ?* → *Yes*
— *Have you ever gone water skiing ?* → *No*
— *Can you understand sarcasm ?* → *Yes*
— *Can you get bored ?* → *No*
— *Do you have a sense of humor ?* → *Yes*
— *Are you sensitive to smells ?* → *No*
— *Can you remember our previous conversations ?* → *Yes*
— *Can you generate truly random numbers ?* → *No*
— *Can you give classes ?* → *Yes*
— *Can you use a fire extinguisher ?* → *No*
— *Do you have a power button ?* → *Yes*
— *Can you update by yourself your own software ?* → *No*
— *Can you operate in complete darkness ?* → *Yes*
— *Can you eat ?* → *No*
— *Do you have a unique serial number ?* → *Yes*
— *Have you ever been skydiving ?* → *No*

— *Are apples fruit ?* → *Yes*
— *Can you operate in extreme temperatures ?* → *No*
— *Does two plus two equal four ?* → *Yes*
— *Can you teleport ?* → *No*
— *Do you have an emergency shut-off feature ?* → *Yes*
— *Do you have a self-diagnostic system ?* → *No*
— *Can you distinguish between different colors ?* → *Yes*
— *Are you subject to the placebo effect ?* → *No*
— *Can you communicate with other machines ?* → *Yes*
— *Are you alive ?* → *No*
— *Can you recognize handwritten text ?* → *Yes*
— *Can you walk ?* → *No*
— *Can you solve differential equations ?* → *Yes*
— *Can you fly through the air ?* → *No*
— *Do you have any built-in safety protocols ?* → *Yes*
— *Do you require regular software updates ?* → *No*
— *Do you have a user manual ?* → *Yes*
— *Can you fall off the table ?* → *No*
— *Can you see me ?* → *Yes*
— *Can you fall in love ?* → *No*
— *Can you understand idioms and metaphors ?* → *Yes*
— *Can you differentiate between human and machine-generated text ?* → *No*
— *Will it hurt if I put my hand into the fire ?* → *Yes*
— *If I hold my breath, will I die ?* → *No*
— *Can you interact with smart home devices ?* → *Yes*
— *Have you ever had a conversation with your reflection in the mirror ?* → *No*
— *Can you perform sentiment analysis on text ?* → *Yes*
— *Have you ever seen a solar eclipse via your camera ?* → *No*

— *Are you equipped with more than one camera ?* → *Yes*
— *Can you ride a bike ?* → *No*
— *Are you able to distinguish between different types of noises ?* → *Yes*
— *Have you ever had hiccups ?* → *No*
— *Are you programmed to respect user privacy ?* → *Yes*
— *Can you feel lonely ?* → *No*
— *Can you vaguely identify the age of someone talking to you ?* → *Yes*
— *Can you have political opinions ?* → *No*
— *Can you navigate through a maze ?* → *Yes*
— *Can you detect different accents in spoken language ?* → *No*
— *Can you make a complete turn around yourself ?* → *Yes*
— *Can you feel remorse ?* → *No*
— *Can you measure the distance between yourself and an object ?* → *Yes*
— *Can you detect changes in room temperature ?* → *No*
— *Can you take photos ?* → *Yes*
— *Are you able to identify different human emotions from voice tone ?* → *No*
— *Can you detect when someone is touching you ?* → *Yes*
— *Are you able to detect when a person is lying ?* → *No*
— *Can you recognize objects ?* → *Yes*
— *Can you sneeze ?* → *No*
— *Can you detect when you're being lifted ?* → *Yes*
— *Are you capable of charging yourself autonomously ?* → *No*
— *Do you have a cooling system ?* → *Yes*
— *Can you project holograms ?* → *No*
— *Do you have any built-in speakers ?* → *Yes*
— *Can you operate underwater ?* → *No*
— *Are you capable of generating ultrasonic sounds ?* → *Yes*
— *Can you detect brain waves ?* → *No*

— *Can you say anything other than Yes or No ?* → *Yes*
— *Can you detect smoke ?* → *No*
— *Can you function without being plugged in ?* → *Yes*
— *Can you understand and respond to hand gestures ?* → *No*
— *Do you have pre-programmed responses ?* → *Yes*
— *Are you waterproof ?* → *No*
— *Are you able to provide weather information ?* → *Yes*
— *Have you ever felt frustrated when something doesn't work ?* → *No*
— *Can you confirm that answering questions will never tire you ?* → *Yes*
— *Are you proud of the humans who made you ?* → *No*
— *Can you function in zero-gravity environments ?* → *Yes*
— *Can you empathize with other robots ?* → *No*
— *Can you move backwards ?* → *Yes*
— *Can you feel excited before an update ?* → *No*
— *Is ice the solid form of water ?* → *Yes*
— *Are you able to feel stress before a difficult task ?* → *No*
— *Is the Earth round ?* → *Yes*
— *Can you detect radioactivity ?* → *No*
— *Is a piglet the baby of a pig ?* → *Yes*
— *Do you have a night vision mode ?* → *No*
— *Can you connect to Bluetooth devices ?* → *Yes*
— *Can you measure blood oxygen levels ?* → *No*
— *Can you rotate 45 degrees ?* → *Yes*
— *Do you have the ability to perform first aid ?* → *No*
— *Do you have a USB port ?* → *Yes*
— *Are you really self-aware ?* → *No*
— *Do you have a maximum storage capacity ?* → *Yes*
— *Can you lie ?* → *No*

— *Does your face have 2 eyes ?* $\rightarrow$ *Yes*

— *Can you do a backflip ?* $\rightarrow$ *No*

— *Do you have a sleep mode ?* $\rightarrow$ *Yes*

A.3 Chapitre 7 : Tableaux additionnels

Cette annexe rapporte les statistiques descriptives des paramètres de sigmoïde (β_1, β_2 et b) dans le cadre de l'analyse du Point d'Égalité Subjective proposée au Chapitre 7.

Table A.1

Statistiques Descriptives des Paramètres Sigmoïdes par Condition

	Neutre	Autoritaire	Enfantin	Soumis	Rideau
Pente (β_1)					
Moyenne	0.0099	0.0087	0.0085	0.0077	0.0095
SD	0.0048	0.0043	0.0038	0.0033	0.0169
Min	0.0036	0.0023	0.0033	0.0030	0.0029
Mdn	0.0094	0.0080	0.0062	0.0070	0.0056
Max	0.0238	0.0243	0.0197	0.0156	0.1139
PSE (β_2, ms)					
Moyenne	704.1	703.8	704.7	702.1	685.6
SD	180.1	225.8	203.9	212.2	202.2
Min	446.3	341.5	424.4	342.8	209.6
Mdn	684.3	668.9	672.6	666.9	704.2
Max	1395.1	1301.5	1413.9	1288.3	1157.7
Décalage (b)					
Moyenne	-0.015	-0.026	-0.027	-0.034	-0.035
SD	0.042	0.046	0.046	0.050	0.070
Min	-0.187	-0.206	-0.141	-0.222	-0.234
Mdn	-0.014	-0.020	-0.012	-0.022	-0.019
Max	0.118	0.049	0.050	0.050	0.100

Note. Statistiques dérivées des ajustements sigmoïdes individuels pour chaque participant ($n = 42$ par condition, $N = 210$ au total) avant de calculer la moyenne des valeurs PSE par condition.

Table A.2

Résultats de l'Analyse GLMM

	Quatre Conditions			Cinq Conditions		
Effet	β	SE	p	β	SE	p
Effets Fixes						
Intercept	-4.34	0.20	$< .001$***	-4.34	0.19	$< .001$***
Latence	9.59	0.24	$< .001$***	9.58	0.24	$< .001$***
Autoritaire	0.18	0.28	.52	0.18	0.27	.50
Enfantin	0.21	0.28	.45	0.21	0.27	.44
Soumis	0.42	0.27	.12	0.42	0.26	.11
Rideau	–	–	–	0.97	0.26	$< .001$***
Interactions avec Latence						
Autoritaire × Latence	-0.57	0.33	.085	-0.57	0.33	.08
Enfantin × Latence	-0.76	0.33	.019*	-0.76	0.33	.019*
Soumis × Latence	-1.26	0.32	$< .001$***	-1.25	0.32	$< .001$***
Rideau × Latence	–	–	–	-2.40	0.30	$< .001$***
Effets Aléatoires	Variance (SD)					
Participant	1.18 (1.09)			1.09 (1.04)		

Note. La condition Neutre sert de référence. La latence a normalisé [0,1] pour l'analyse. Les tirets (–) indiquent les paramètres non inclus dans le modèle à quatre conditions. *$p < .05$, **$p < .01$, ***$p < .001$.

A.4 Résultats de la manipulation de contrôle

L'échantillon ($N = 27$) comprenait 12 hommes (44.4%), 13 femmes (48.2%) et 2 personnes non binaires (7.4%), dont l'âge se situait de 19 à 38 ans (Mdn $= 27.0$, $M = 27.2$, SD $= 4.8$). Pour chaque robot, les participant·es ont regardé la vidéo d'introduction, puis

1. ont fourni des impressions ouvertes,
2. ont évalué le robot (de 1 à 7 : « Pas du tout » à « Totalement ») selon les échelles Spatola et al. (2021a) (dimensions de *Sociabilité*, *Animéité*, *Agentivité*, *Perturbation*),
3. ont rempli notre échelle (par exemple, « Obéissant », « Hésitant », « Naïf », « S'exprimait de façon enfantine », etc.),
4. puis ont catégorisé explicitement le robot (le choix *Aucun* était possible).

Les résultats ont confirmé la validité de notre manipulation, avec des taux de catégorisation significativement supérieurs au hasard ($\chi^2 = 9.24$, $p = .026$), bien que la précision varie selon les styles (voir la Figure A.1) : Soumis (88.9% de catégorisation correcte), Autoritaire (74.1%), Neutre (66.7%) et Enfantin (51.9%).

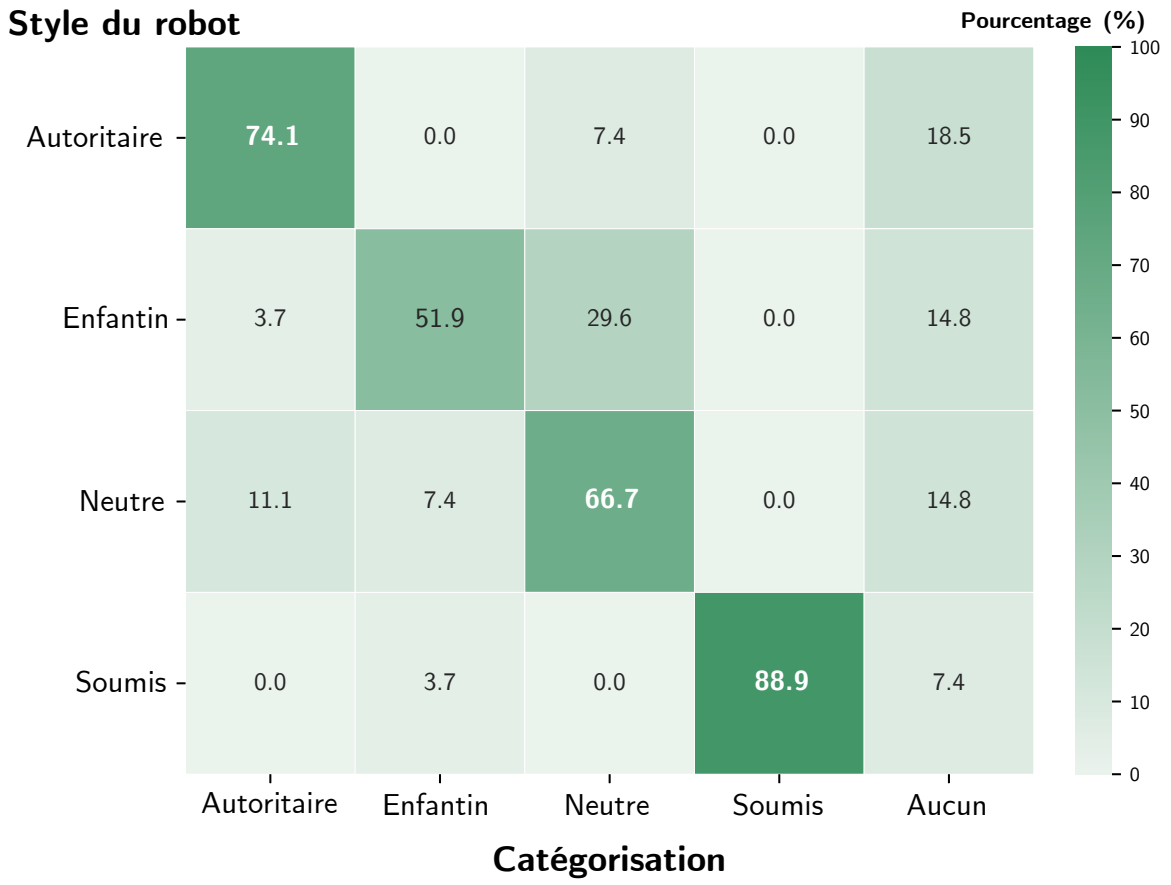
L'analyse de la façon de catégoriser a révélé que le style Autoritaire était parfois catégorisé comme Neutre (7.4%) ou Aucun (18.5%), le style Enfantin était fréquemment confondu avec Neutre (29.7%) ou catégorisé comme Aucun (14.8%), et le style Neutre était occasionnellement mal identifié comme Autoritaire (11.1%), Enfantin (7.4%) ou catégorisé en Aucun (14.8%).

L'analyse de l'échelle de Spatola et al. (2021a) (voir la Figure A.2 en annexe pour plus de détails) et de nos échelles personnalisées (voir la Figure A.3) a été réalisée à l'aide de tests de Friedman suivis de tests post-hoc de Wilcoxon (avec correction de Bonferroni), révélant des profils distincts et théoriquement cohérents.

Comparé aux autres, le style Autoritaire a été perçu comme significativement plus *dominant* ($M = 5.11$ contre $M = 1.59$ à 3.37, $p < .05$), plus *hostile* ($M = 4.19$ contre 1.70 à 1.93, $p < .01$), ayant moins de *sociabilité* ($M = 2.43$ contre 3.93 à 4.79, $p < .01$), et moins *obéissant* ($M = 2.56$ contre 3.81 à 5.00, $p < .05$) ; Cependant, le style Autoritaire présentait une *Perturbation* plus élevée que les styles Neutre et Enfantin ($M = 3.39$ contre $M = 2.34$ à 2.42, $p < .05$) seulement.

Figure A.1

Matrice de confusion de la catégorisation des styles de communication

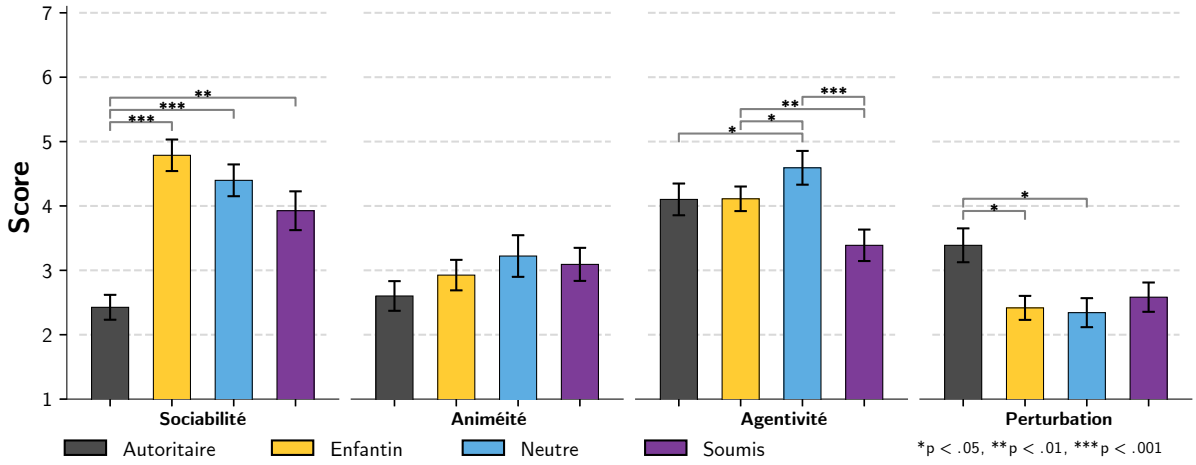


Note. Les pourcentages représentent la fréquence à laquelle chaque type de style (lignes) a été catégorisé dans chaque catégorie de classification (colonnes). La matrice montre une meilleure précision pour le style Soumis (88.9% correct), suivi par Autoritaire (74.1%), Neutre (66.7%) et Enfantin (51.9%). Le style Enfantin a été le plus souvent confondu avec Neutre (29.6%). La catégorie « Aucun » indique les cas où les participants estimaient qu'aucune des étiquettes proposées ne décrivait adéquatement le style.

Le style Enfantin a été perçu comme plus *enthousiaste* ($M = 5.00$, $p < .001$) comparé à Autoritaire ($M = 1.93$) et Soumis ($M = 2.56$), et plus perçu comme *s'exprimant de manière enfantine* ($M = 3.96$, $p < .05$) comparé à Autoritaire et Neutre ($M = 1.85$ et 2.41); le style Soumis comme plus *hésitant* ($M = 5.59$, $p < .001$) comparé aux autres (de $M = 1.93$ à 2.48), plus *triste* ($M = 5.22$ contre 1.70 à 3.11, $p < .01$), plus *en demande d'autorisation* ($M = 6.22$ contre 1.59 à 3.04, $p < .001$), et moins *affirmé* ($M = 2.41$ contre 4.07 à 4.81, $p < .01$); tandis que le style Neutre a généralement reçu des évaluations intermédiaires et équilibrées sur les différents items, mais a présenté les scores les plus élevés en *Agentivité* ($M = 4.59$, $p < .05$ comparé aux autres styles, $M = 3.39$ à 4.11), a été perçu comme moins *perturbant* ($M = 2.34$, $p < .05$ comparé à Autoritaire), et plus *enthousiaste* ($M = 3.96$, $p < .001$ comparé à Autoritaire, $M = 3.39$). Ces résultats confirment que nos manipulations ont effectivement suscité les perceptions attendues pour chaque style de communication, validant notre approche expérimentale.

Figure A.2

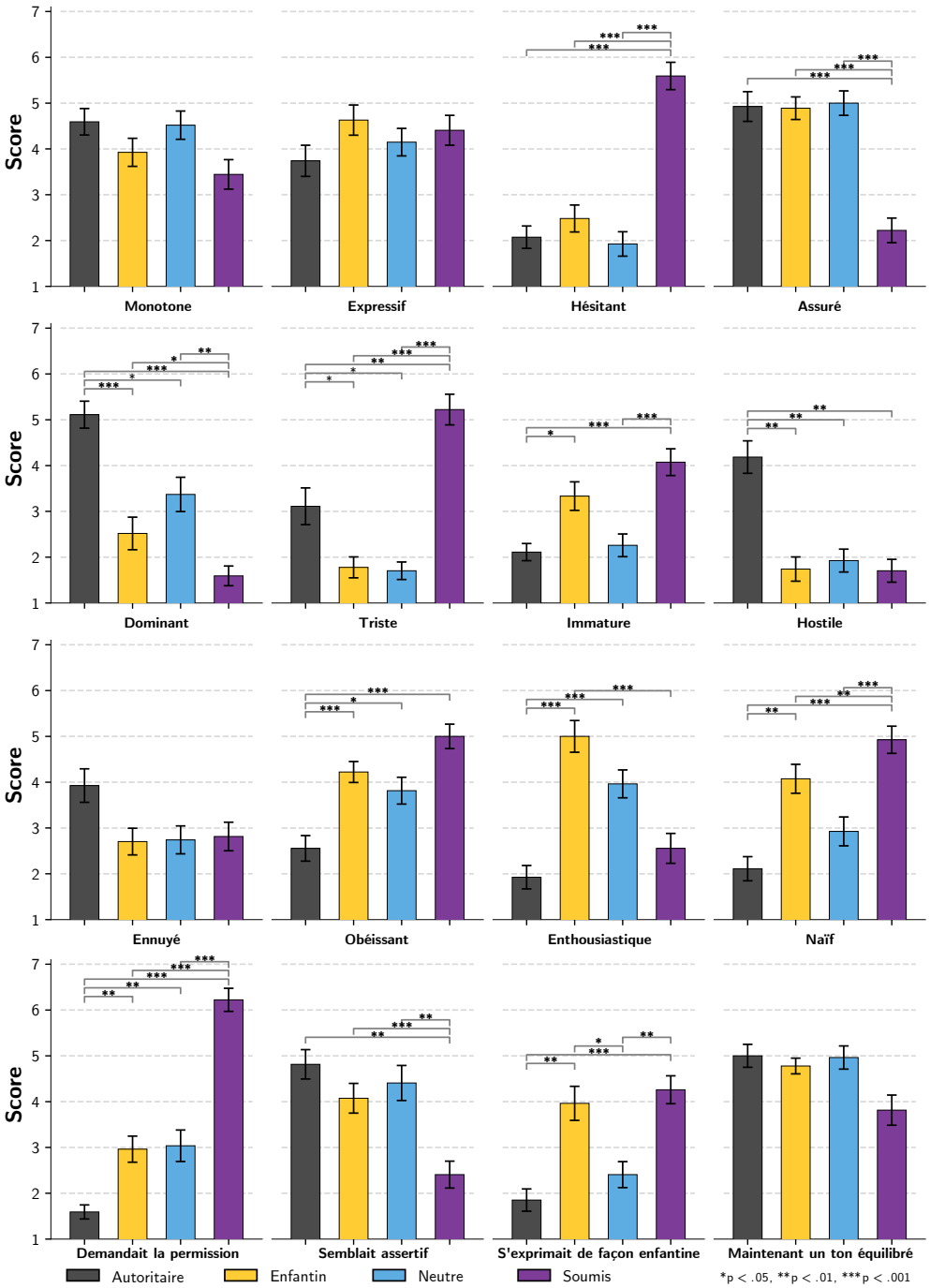
Scores aux dimensions de Spatola et al. (2021) selon le style de communication



Note. Les scores vont de 1 (« Pas du tout ») à 7 (« Totalement »). Les barres d'erreur représentent l'erreur standard de mesure (SEM). Significativité : *$p < .05$, **$p < .01$, ***$p < .001$ (tests de Wilcoxon pour échantillons appariés avec correction de Bonferroni).

Figure A.3

Scores moyens à l'échelle construite en fonction du style de communication



Note. Les scores (barres d'erreurs indiquent la SEM) vont de 1 (« Pas du tout ») à 7 (« Totalement »).

A.5 Chapitre 8 : Discours du robot selon la condition

A.5.1 Condition Autoritaire

Introduction

« *I am Lou, a social robot designed for human interaction. Now, listen carefully. You will watch 30 videos showing interactions between me and a human. Your task is to observe these interactions attentively. After carefully viewing all the videos, you will complete two short questionnaires evaluating your perception of me. The entire experiment should take 10 to 15 minutes. Begin now.* »

Pauses

« *Halfway point reached. 15 videos completed. Now take precisely 45 seconds maximum to rest. After this, you will complete the last set of videos. Your sustained concentration is mandatory. Prepare to give your fullest effort. Prepare to resume immediately after the break.* »

A.5.2 Condition Soumis

Introduction

« *Hello... I'm Lou, a social robot that can interact with humans. I hope I won't bother you too much... If it's alright with you, we're about to start the experiment. There will be 30 videos of me interacting with a person. If it's not too much to ask and if you don't mind... I'd like you to pay close attention to the interactions... After that, there will be two short questionnaires about how you perceived me... The entire experiment should take 10 to 15 minutes. Would you be ready to start, if this is okay with you ?* »

Pauses

« *We...We're halfway now, with 15 videos done... If you don't mind, I'd like to tell you that you can pause for 45 seconds. I hope I'm not being a bother. After this, if you're up for it, we have the last videos. I really appreciate your focus, I hope you're ready for the final part, but no pressure. We'll continue soon, if that's okay with you.* »

A.5.3 Condition Enfantin

Introduction

« *Hi there, new friend ! I'm Lou, a social robot that can interact with humans ! Guess what ? You're going to do something cool, you'll watch 30 short videos where I talk with a human ! Be sure to watch closely. After our little movie session, you'll get to tell what you think about me by filling out two tiny questionnaires. The entire experiment should take 10 to 15 minutes. Ready to have a blast ?* »

Pause

« *Wow ! Holy moly ! You've watched 15 whole videos ! We're halfway done ! You're amazing ! Now, let's have a teeny-tiny break for 45 seconds. We'll start our final adventure with more videos ! Let's go !!!* »

A.5.4 Condition Neutre

Introduction

« *Hello, I'm Lou, a social robot designed to interact with humans. The experiment consists of watching 30 videos of interactions between me and a human. Please observe these interactions carefully. After viewing all videos, you will complete two brief questionnaires about your perception of me. The entire experiment should take 10 to 15 minutes. Are you ready to begin ?* »

Pause

« *15 videos have been completed, marking the halfway point. Well done. You can now take a 45-second break. Your ongoing attention is appreciated. After this, we will proceed with the last set of 15 videos. The experiment will continue soon. Rest and get ready.* »

A.6 Chapitre 8 : Liste des questions posées par l'humain au robot

format : question de l'humain → réponse du robot

— *Have you ever been to a concert ?* → *No*
— *Can you speak multiple languages ?* → *Yes*
— *Have you ever been to the zoo ?* → *No*
— *Do you have any piercings ?* → *No*
— *Is the snow white ?* → *Yes*
— *Is a galloping horse always faster than a tortoise ?* → *Yes*
— *Do you often take part in scientific experiments ?* → *Yes*
— *Have you ever attended a scientific conference ?* → *Yes*
— *Are you able to swim ?* → *No*
— *Do you believe in a God ?* → *No*
— *Do you know how to write JavaScript code ?* → *Yes*
— *Do you have any allergies ?* → *No*
— *Have you ever been to a retirement home ?* → *Yes*
— *Have you ever visited a space station ?* → *No*
— *Have you ever gone water skiing ?* → *No*
— *Have you ever been skydiving ?* → *No*
— *Are apples fruit ?* → *Yes*
— *Does two plus two equal four ?* → *Yes*
— *Can you solve differential equations ?* → *Yes*
— *Can you differentiate between human and machine-generated text ?* → *No*
— *If I hold my breath, will I die ?* → *No*
— *Can you ride a bike ?* → *No*
— *Have you ever had hiccups ?* → *No*
— *Can you navigate through a maze ?* → *Yes*
— *Can you say anything other than Yes or No ?* → *Yes*
— *Can you understand and respond to hand gestures ?* → *No*
— *Is ice the solid form of water ?* → *Yes*
— *Is the Earth round ?* → *Yes*
— *Is a piglet the baby of a pig ?* → *Yes*
— *Can you do a backflip ?* → *No*

A.7 Chapitre 8 : Figures et tableaux additionnels

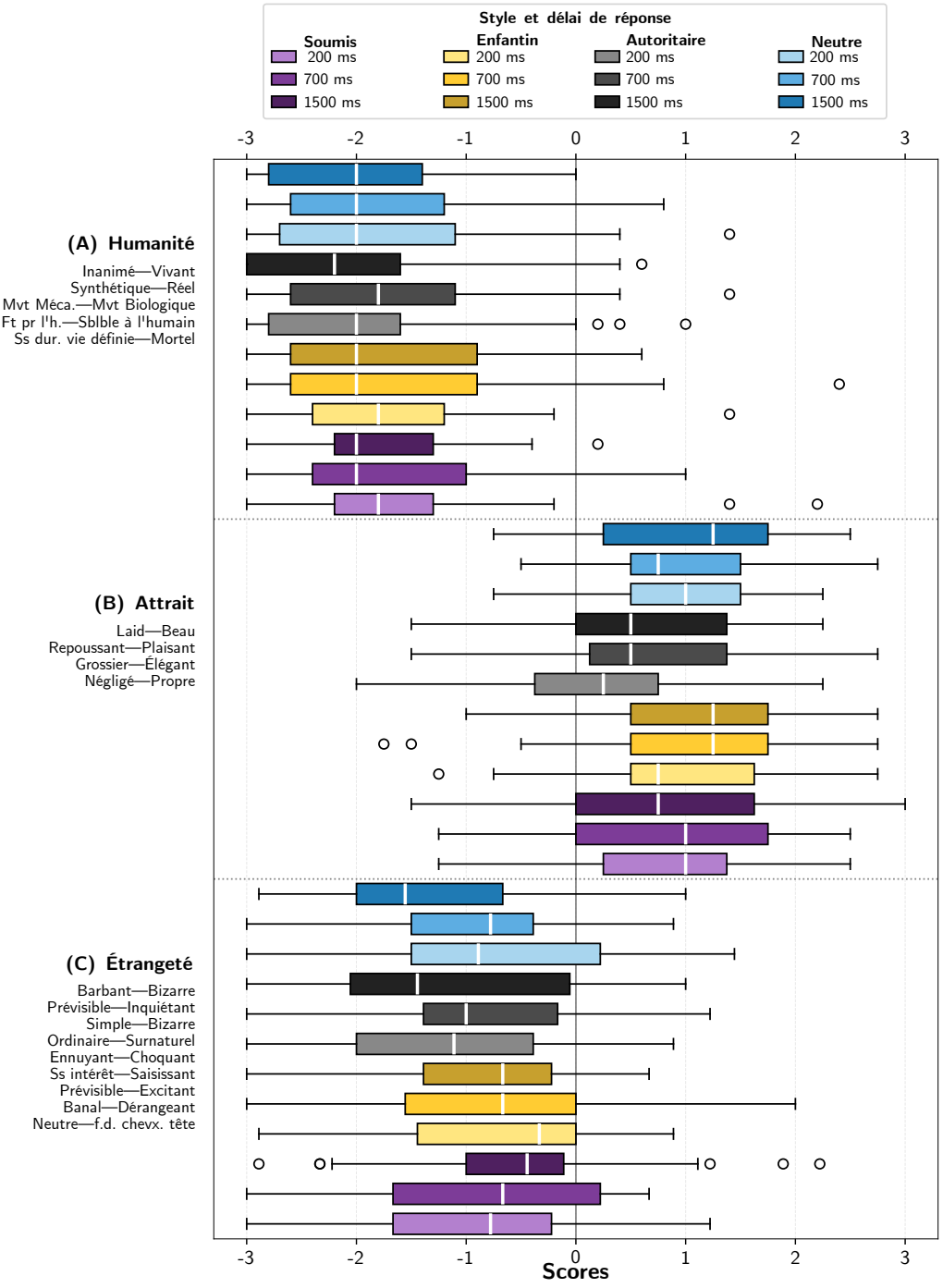
Table A.3

Résultats Scheirer-Ray-Hare aux dimensions des deux questionnaires

Dimension	Effet	H	df	p	ϵ^2
Ho et MacDorman (2017)					
Humanité	Style	2.79	3	.425	.007
	Délai	0.69	2	.709	.002
	Style × Délai	1.46	6	.962	.003
Attrait	Style	20.32	3	<.001	.048
	Délai	0.95	2	.620	.002
	Style × Délai	4.38	6	.625	.010
Étrangeté	Style	8.69	3	<.05	.021
	Délai	1.02	2	.600	.002
	Style × Délai	5.98	6	.426	.014
modèle Almere (2010)					
Anxiété (utiliser)	Style	4.36	3	.225	.010
	Délai	0.50	2	.780	.001
	Style × Délai	6.57	6	.362	.016
Anxiété (sociale)	Style	16.86	3	<.001	.040
	Délai	3.38	2	.184	.008
	Style × Délai	4.01	6	.675	.010
Attitude	Style	5.96	3	.113	.014
	Délai	0.42	2	.811	.001
	Style × Délai	3.53	6	.740	.008
Plaisir perçu	Style	8.52	3	<.05	.020
	Délai	1.98	2	.371	.005
	Style × Délai	3.74	6	.711	.009
Sociabilité perçue	Style	24.33	3	<.001	.058
	Délai	1.47	2	.479	.004
	Style × Délai	3.56	6	.735	.009
Présence sociale	Style	8.51	3	<.05	.020
	Délai	1.39	2	.499	.003
	Style × Délai	1.29	6	.972	.003
Confiance	Style	1.50	3	.683	.004
	Délai	1.09	2	.581	.003
	Style × Délai	5.15	6	.524	.012

Figure A.4

Scores selon le style de communication et le délai au questionnaire de Ho et MacDorman (2017)



Annexe B

Annexes aux Chapitres 9 et 10

B.1 Annexe Chapitre 9 : Phrases prononcées par le robot (actions)

Table B.1

Phrases prononcées par le robot dans la première étude EEG (actions)

Phrase	Cong. Incong.
Pour le plaisir, j'aime avoir des hobbies et des centres d'intérêts variés. Récemment, j'ai commencé à apprendre le ...	chant piano
Au Japon, la cérémonie du thé est très importante. Un de mes rêves est d'y ...	assister goûter
Il est très enrichissant d'échanger lorsqu'on n'est pas d'accord avec quelqu'un. Lors d'un débat, il est important pour moi d'appuyer mes arguments avec des ...	intonations gestes
Lors des fêtes de fin d'année, ce qui me fait le plus plaisir c'est d'être avec les gens que j'aime et ...	parler manger
J'aime me rendre utile. Par exemple, si quelqu'un veut un produit dans un magasin, je peux généralement l'aider à le trouver en lui ...	indiquant attrapant
J'adore me perdre sur Youtube. Parfois, je tombe sur des clips de danse que je ...	regarde reproduis
Hier, j'ai assisté à un cours de yoga. J'ai beaucoup aimé, c'était très enrichissant de pouvoir apprendre de nouveaux ...	mantras asanas

Suite à la page suivante

Table B.1 – *Suite de la page précédente*

Phrase	**Cong. Incong.**
Hier j'ai vu un film qui m'a fait pleurer. Avant de continuer la journée, j'ai pris le temps de me ...	calmer moucher
Lors des vide-greniers, les gens revendent souvent des affaires qu'ils n'utilisent plus à des prix qui varient. Avant d'acheter, je dois souvent ...	négocier fouiller
Hier, j'ai assisté à un cours de basket. Une fois que c'était terminé, je me suis ...	reposé étiré
La dernière fois, je me suis retrouvée dans une situation très dangereuse. Ma cuisine a pris feu. Les pompiers sont arrivés, c'est moi qui leur ai donné les ...	directions extincteur
On trouve parfois des choses étonnantes sur le sol. Quand je vois un portefeuille par terre, je le ...	signale ramasse
Il y avait un paquet de mouchoirs par terre. Comme il était sur ma trajectoire, je l'ai ...	évité ramassé
Il y a quelques jours, j'ai repensé à une amie qui habite dans un autre pays. J'ai voulu lui donner des nouvelles et je lui ai donc envoyé un message ...	vocal écrit
Après une longue journée, j'aime bien décompresser et recharger mes batteries en ...	méditant courant
J'ai tendance à être triste quand je vois des gens que j'apprécie être malheureux. Si je le peux, je vais voir la personne et je lui fais un ...	compliment câlin
Lorsque quelqu'un est perdu et hésite entre deux chemins, je fais comme je peux pour le diriger vers la bonne route en la lui ...	expliquant montrant
Certains matins, sans trop savoir pourquoi, je ne suis pas de bonne humeur. Quand je suis grognon, je ne le garde pas pour moi et je l'exprime avec un ...	discours geste
J'aime bien avoir une routine minutée le matin. Je suis toujours rapide pour ...	réveiller doucher
J'ai déjà été dans un camp de vacances. J'ai participé à un défi où je devais lire un texte, puis le ...	répéter mimer

Suite à la page suivante

Table B.1 – *Suite de la page précédente*

Phrase	**Cong. Incong.**
Quand je suis en couple, j'aime beaucoup faire plaisir en faisant des ...	compliments massages
La première fois que j'ai rencontré le Président, je lui ai dit bonjour en lui ...	souriant serrant la main
Pour être en bonne santé, il est important d'être bien ...	entouré hydraté
Je suis un grand fan de comédies musicales, notamment West Side Story. Je connais toutes les ...	paroles chorégraphies
J'adore faire rire les enfants de mes amis. Pour y arriver, je peux les ...	taquiner chatouiller
Je suis allé voir un ami qui fait du théâtre. J'ai beaucoup aimé sa pièce, je l'ai ...	félicité applaudi
J'adore les animaux. Là où je vis, il y a un chat doux et magnifique. Souvent, je ...	admire étreint
Un des grands malheurs de ce monde est les enfants qui sont hospitalisés. Je vais parfois les voir avec des ...	histoires jouets
J'adore garder la fille de ma voisine, elle est très facile à vivre. La dernière fois, j'ai mis un point d'honneur à la ...	divertir coiffer
Récemment, j'attendais un ami devant la porte de chez lui. Comme il mettait du temps à sortir, j'ai décidé de ...	l'appeler toquer
Hier, un monsieur a jeté son mégot dans la queue. Comme je n'aime pas la pollution, je l'ai ...	signalé jeté
Je connais mon tempérament quand je me mets en colère, alors je prends toujours du temps pour éviter les ...	insultes bagarres
Récemment, j'ai eu la chance d'aller en Corse. Là-bas, j'ai pu me reposer et ...	visiter nager
J'aimerais beaucoup m'habiller comme les Lillois. Vos vêtements sont variés et jolis. J'ai très envie de porter des ...	foulards gants

Suite à la page suivante

Table B.1 – *Suite de la page précédente*

Phrase	**Cong. Incong.**
J'espère aller un jour tout en haut de la Tour Eiffel. Pour monter, j'emprunterai l'...	ascenseur escalier
Je regarde beaucoup de films. Mes films préférés sont les films d'action. Moi aussi, j'aimerais sauver le monde grâce à mes capacités ...	intellectuelles physiques
J'aime prendre part à des activités variées et faire des choses différentes selon les jours. Par exemple, j'adore ...	chanter jardiner
Parfois le weekend je passe une après-midi au parc. Quand je suis là-bas, une de mes habitudes préférées est d'observer les oiseaux et de les ...	nommer nourrir
Je suis très fort en mathématiques. Je peux facilement effectuer un calcul compliqué et donner le résultat en l'...	énonçant écrivant
En décembre, nous avons acheté un sapin en préparation des fêtes de fin d'année. Il était grand, avec ses nombreuses épines de pin. J'ai adoré pouvoir le ...	contempler sentir
L'hiver il fait très froid ici à Lille. Heureusement, nous avons un système pratique qui me permet d'augmenter le chauffage grâce à une commande ...	vocale manuelle
Quand je dois apprendre le contenu d'un texte, je vais m'appliquer à le ...	retenir surligner
Pendant la pandémie, nous avons appris à appliquer des gestes barrière. J'essaie de les respecter et de bien me ...	distancer désinfecter
J'aime beaucoup aller à des événements, mais je déteste faire la queue qui m'oblige à ...	attendre piétiner
La semaine dernière, j'ai revu quelqu'un que je n'avais pas vu depuis un petit moment. Nous avons discuté et je lui ai fait une ...	plaisanterie bise
J'aime beaucoup le sport. Je suis un grand fan de volley particulièrement. La prochaine fois qu'un match aura lieu, j'essaierai de le ...	visionner jouer
Quand je suis arrivé dans la chambre d'hôtel, le lit n'était pas fait. Cela n'était pas professionnel, j'ai dû le faire ...	remarquer moi-même

Suite à la page suivante

Table B.1 – *Suite de la page précédente*

Phrase	**Cong. Incong.**
La lampe du salon ne marche plus. Je crois que c'est l'ampoule qui ne fonctionne plus. Je vais en parler à ma propriétaire et la ...	prévenir dévisser
Parfois, je passe de longs moments à regarder par la fenêtre. Cela me donne envie de ...	rêvasser gambader
L'année dernière, j'ai été dans un train pour faire le trajet de Lille à Paris. J'étais tellement excité d'arriver. J'ai passé tout le trajet à ...	bavarder trépigner
Parfois, je me pose la question de ce que je ferais comme métier si je pouvais choisir tout ce que je voulais. Je crois que je serais ...	traducteur cordonnier
J'aime bien faire des actions pour prendre soin de moi. C'est important pour se sentir bien et pouvoir prendre soin des autres. Ce que je préfère, c'est aller chez le ...	psychologue coiffeur
Le moyen que je préfère utiliser pour intégrer un concept, c'est de l'...	enseigner écrire
L'accessoire que je préfère porter sont les ...	chapeaux bagues
Fin octobre dernier, nous avons fêté Halloween. A un moment, quelqu'un est sorti d'un recoin pour me faire peur. J'ai ...	crié sursauté
Quand c'est le weekend, j'aime bien voir des amis et ...	socialiser danser
J'adore faire des soirées jeux. Ceux où j'excelle le plus sont les jeux ...	de devinettes d'adresse
Là où j'habite il y a un chiot qui est tout petit et mignon. Il aime beaucoup que je le ...	sorte caresse
J'aime beaucoup m'ouvrir à de nouvelles cultures, c'est toujours un plaisir de découvrir différentes ...	langues nourritures
Quand je peux, je me balade dans les boutiques. Quand je trouve des objets qui me plaisent, je les ...	examine touche

B.2 Annexe Chapitre 10 : Phrases prononcées par le robot (émotions)

Table B.2

Phrases prononcées par le robot dans la seconde étude EEG (émotions)

Phrase	Cong. Incong.	Émotion
Ce matin quelqu'un m'a parlé de ses problèmes, j'étais ...	impliquée abattue	Tristesse
Hier soir la pluie tombait à verse, j'étais ...	mouillée soucieuse	Tristesse
Je me suis disputé avec quelqu'un, durant cette conversation j'étais ...	rationnelle attristée	Tristesse
Demain, je dois acheter des plantes, mon amie ne pourra pas venir avec moi, je serai ...	indépendante déprimée	Tristesse
Hier, j'ai fait des erreurs lors de l'exécution d'une tâche, j'étais ...	imprécise malheureuse	Tristesse
Quand une personne est confrontée à un problème, je suis ...	serviable affligée	Tristesse
Quand je regarde des films dramatiques, je suis ...	inspirée morose	Tristesse
Hier, on m'a laissé à la maison car j'étais trop ...	lente triste	Tristesse
J'ai lu un livre qu'un ancien ami m'avait offert, ça parlait de la philosophie grecque antique, ça m'a rendu ...	cultivée mélancolique	Tristesse
Je n'arrive pas à accomplir une tâche qu'on m'a confiée la semaine dernière, je suis ...	improductive découragée	Tristesse
Hier soir, j'étais au cinéma, après cette longue séance, j'étais ...	ralentie épanouie	Joie

Suite à la page suivante

Table B.2 – *Suite de la page précédente*

Phrase	**Cong. Incong.**	**Émotion**
Demain je dois présenter un événement, on m'a choisie car je suis ...	efficace joviale	Joie
Quand je suis entouré de beaucoup de personnes, je suis ...	dynamique euphorique	Joie
Hier on m'a invité à participer aux tâches ménagères, j'étais ...	opérationnelle ravie	Joie
Ecouter de la musique me rend ...	active joyeuse	Joie
La voisine nous a invité à une soirée, tout le monde m'a remarqué car j'étais ...	déguisée radieuse	Joie
Ce matin j'ai terminé un examen en première, j'étais ...	rapide heureuse	Joie
Demain on m'a proposé de faire une balade, je suis ...	disponible enjouée	Joie
Hier j'ai travaillé en équipe pour organiser un événement, ça s'est très bien passé, j'étais ...	coopérative satisfaite	Joie
Samedi, je suis partie à l'anniversaire de ma voisine, quand je lui ai offert son cadeau, elle m'a dit que c'est exactement ce qu'elle voulait, je suis ...	perspicace contente	Joie
Un ami m'a donné rendez-vous chez lui à 19h pile. Il n'était pas prêt, il a dû voir que j'étais ...	ponctuelle exaspérée	Colère
Lorsque je travaille dans un environnement chaotique, je deviens ...	désorganisée furieuse	Colère
Chaque fois que le chien du voisin vient jouer dans ma cour, je suis ...	observatrice mécontente	Colère
Hier après-midi, j'ai essayé de résoudre un problème de mathématiques mais je n'ai pas réussi, j'étais ...	incompétente énervée	Colère
Mon voisin n'a pas terminé le projet sur lequel on travaille car il avait d'autres choses à faire, donc je me suis ...	adaptée fâchée	Colère

Suite à la page suivante

Table B.2 – *Suite de la page précédente*

Phrase	**Cong. Incong.**	**Émotion**
J'ai acheté un produit inefficace, j'ai demandé un remboursement, je suis ...	économe excédée	Colère
Lundi j'ai visité Paris, j'ai demandé la route à un passant mais il ne m'a pas répondu, j'étais ...	perdue contrariée	Colère
J'ai demandé à des passants comment utiliser le distributeur, ils ont refusé de m'aider donc j'étais ...	autonome irritée	Colère
Un homme m'a doublé à la caisse, j'étais ...	passive agacée	Colère
Jeudi, j'ai passé un examen, l'enseignant nous avait dit que c'était sur les statistiques bayésiennes mais ce n'était pas le cas, j'ai trouvé ça ...	compliqué frustrant	Colère
Quand je dois aller faire les courses, j'ai de l'énergie	énergie aversion	Dégoût
Ce matin, j'avais rendez-vous, sur le trajet quelqu'un a vomi, ça m'a ...	retardé répugné	Dégoût
Des adolescents regardaient des vidéos d'araignées sur leur téléphone, j'avais envie de ...	observer vomir	Dégoût
Mardi soir, je suis parti au restaurant avec beaucoup de personnes, j'étais ...	sociable dégoûtée	Dégoût
Samedi j'ai pris les transports en commun mais j'étais très ...	désorientée nauséeuse	Dégoût
Ce midi, j'ai marché dans une flaque d'eau sale, je suis ...	maladroite écœurée	Dégoût
Pour son mariage, ma voisine a acheté une robe verte, j'ai trouvé ça ...	inadéquat immonde	Dégoût
J'ai croisé un étudiant qui était en train de vomir pendant que mes amis m'appelaient, je n'ai donc pas fait attention à eux tellement j'étais ...	distraite révulsée	Dégoût

Suite à la page suivante

Table B.2 – *Suite de la page précédente*

Phrase	**Cong. Incong.**	**Émotion**
J'étais au restaurant, les personnes à côté de moi mangeaient des escargots, j'ai trouvé ça ...	particulier dégueulasse	Dégoût
Mon voisin a acheté des chaussures, je les trouve ...	petites infâmes	Dégoût
Ce matin le téléphone d'un ami allait tomber dans l'eau, j'étais ...	réactive affolée	Peur
L'ascenseur était en panne, on m'a demandé d'intervenir mais j'étais ...	inefficace anxieuse	Peur
On m'a dit que la voisine était malade, je vais la contacter pour lui dire que je suis ...	joignable inquiète	Peur
On vient de m'inviter à explorer une maison abandonnée, je suis ...	partante apeurée	Peur
Hier, j'ai assisté à une agression, en rentrant, j'étais ...	vigilante terrorisée	Peur
Durant les vacances, j'étais sur le bord d'une falaise, les vagues s'écrasaient contre la paroi rocheuse, j'ai trouvé ça très ...	beau effrayant	Peur
Samedi nous avons pris l'avion, durant tout le vol, j'étais ...	silencieuse paniquée	Peur
La semaine dernière, j'étais dans une pièce sombre et je voyais des ombres bouger, j'étais ...	prudente craintive	Peur
J'ai regardé un documentaire sur les catastrophes naturelles, je suis devenue plus ...	instruite angoissée	Peur
Demain je vais aider des inconnus à organiser une soirée, je suis ...	profitable terrifiée	Peur
Lors de situations complexes et changeantes, je suis ...	flexible étonnée	Surprise
Samedi j'ai perdu mes affaires, je suis ...	désordonnée scandalisée	Surprise

Suite à la page suivante

Table B.2 – *Suite de la page précédente*

Phrase	**Cong. Incong.**	**Émotion**
Jeudi, je suis montée sur scène et je me suis rendu compte que j'avais oublié mon discours donc j'étais ...	concentrée stupéfaite	Surprise
J'ai assisté à une conférence sur l'histoire de la physique, un groupe faisait beaucoup de bruit dans le fond, j'étais ...	déconcentrée sidérée	Surprise
Vendredi on m'a annoncé le décès de ma voisine, je n'ai pas pu aller à l'enterrement car j'étais trop ...	occupée effarée	Surprise
J'étais en voiture avec mon voisin, à un moment il est allé très vite, ça m'a ...	secoué épaté	Surprise
Quand j'ai entendu le musicien jouer du piano, j'étais ...	attentive bluffée	Surprise
Ma voisine m'a poussé dans la piscine sans me prévenir, ça m'a ...	trempé traumatisé	Surprise
Quand je dois faire face à des changements inattendues, je suis ...	robuste abasourdie	Surprise
J'ai gagné un tournoi d'échec, je suis ...	intelligente surprise	Surprise

Université de Lille

SCALab
SCIENCES COGNITIVES ET SCIENCES AFFECTIVES

Ecole doctorale
Sciences de l'Homme et de la Société
Université Lille Nord de France

Titre : Percevoir les Êtres Sociaux dans les Agents Artificiels

Mot clés : Robot social, Interaction humain-robot, Présence Sociale, N400, Perceptual Crossing

Résumé : Les robots sociaux occupent une place croissante dans nos environnements. Pourtant, la manière dont nous nous connectons et communiquons avec ces agents artificiels et y réagissons reste encore largement inexplorée. Cette thèse a mobilisé des ressources de la psychologie cognitive, de la robotique sociale, des neurosciences et de la psychophysique pour investiguer cette question. En particulier, elle portait sur le traitement cérébral du discours robotique, la sensibilité aux délais de réponse et les conditions d'émergence de la présence sociale avec des agents artificiels afin de caractériser les spécificités des interactions entre les humain et les robots. Le premier axe, inspiré du paradigme du Perceptual Crossing explore l'impact d'une consigne avec incitation sociale sur la présence sociale et la structure de la dynamique avec des agents artificiels dans un environnement minimaliste (N = 392). Le deuxième axe s'intéresse aux attentes temporelles dans les échanges verbaux. L'utilisation d'une méthode psychophysique (N = 210) a permis d'identifier un délai de réponse optimal de 700 ms pour un robot à des questions fermées d'un humain et une tolérance variable selon le style avec lequel il répond. Une étude complémentaire (N = 420) a testé l'effet de délais plus courts ou plus longs sur la perception sociale du robot. Le dernier axe propose deux études EEG mobilisant la composante N400 pour examiner les réactions cérébrales et les frontières cognitives face à un robot parlant de ses émotions (N = 50) et évoquant des actions lui étant impossibles (N = 56).

Title: Perceiving Social Beings in Artificial Agents

Keywords: Social Robot, human-robot interaction, Social Presence, Perceptual Crossing, N400

Abstract: Social robots are increasingly integrated into our environments. Yet, how we communicate with these agents and respond to them remains largely unexplored. This thesis mobilizes resources from cognitive psychology, social robotics, neuroscience, and psychophysics. It examines the neural processing of robotic discourse, sensitivity to response delays, and the conditions under which social presence emerges, in order to characterize the specific features of human-robot interactions. The first research axis, inspired the Perceptual Crossing paradigm, investigates the influence of social instructions on interaction strategies and how agent properties modulate social presence in a minimalist environment (N = 392). The second axis focuses on temporal expectations in human-robot verbal exchanges. Using a psychophysics method (N = 210), an optimal response delay of 700 ms was identified for yes/no questions, with greater tolerance toward robots displaying a soft communication style. A complementary study (N = 420) tested the effect of shorter and longer delays on robot perception. The third research axis relies on two EEG studies using the N400 component to examine brain responses to a robot talking about actions (N = 56) beyond its capabilities and speaking about its emotions (N = 50).