Computational modeling applied to strategic social decision making and Theory of Mind across different species of primates

Toan Nong

THESE de DOCTORAT DE L'UNIVERSITE CLAUDE BERNARD LYON 1

Ecole Doctorale de NEUROSCIENCES ET COGNITION ED 476

Discipline : Neurosciences Cognitives

Soutenue publiquement le 15/11/2023, par :

Toan NONG

La modélisation computationnelle appliquée à la prise de décision sociale et à la théorie de l'esprit dans des jeux stratégiques de coordination chez différentes espèces de primates

---

Devant le jury composé de :

Directeurs de thèse :
DREHER Jean-Claude - Directeur de recherche CNRS, UCBL1, ISC-MJ
DERRINGTON Edmund - Professeur, CNRS, UCBL1, ISC-MJ

Présidente du jury :
CROCKFORD Catherine - Directrice de recherche CNRS, ISC-MJ

Jury :
DAUNIZEAU Jean - Chargé de recherche INSERM, ICM - examinateur
DUFOUR Valérie - Maitresse de Conférence HDR, CNRS, Université de Tours, IFCE, INRAE - examinatrice
GOMEZ Alice – Maitresse de Conférence HDR, UCBL1, ISC-MJ - examinatrice
KHAMASSI Mehdi – Directeur de recherche CNRS, Université Sorbonne, ISIR - Rapporteur
RAKOCZY Hannes – Professeur, Université de Göttingen - Rapporteur

La modélisation computationnelle appliquée à la prise de décision sociale et à la théorie de l'esprit dans des jeux stratégiques de coordination chez différentes espèces de primates

Computational modeling applied to social decision making and Theory of Mind in strategic coordination games across different species of primates

Toan NONG

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*

Neurosciences cognitives

Novembre 2023

Preliminary notes

The present thesis is a publication-based (cumulative) dissertation. It is based on one original article that is ready to be submitted to international peer-reviewed journals, and two original research articles that are currently in preparation. These articles are:

T. Nong, S. Möller, R. Philippe, E. Derrington, I. Kagan, J.-C. Dreher. Differences in computational decision processes between humans and macaques in a transparent Bach-or-Stravinsky Game. *In preparation*.

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Mentalising underly strategic coordination in Guinea baboons (Papio papio). *In preparation*.

T. Nong, Y. Li, C. Qu, J. H. Woo, Y. Wang, C. Miao, X. Liu, Q.Liu, R. Philippe, J.-B. Van der Henst, E. Derrington, A. Soltani, J.-C. Dreher. Computational mechanisms underlying the emergence of theory-of-mind in children. *Submitted*. Psyarxiv: https://psyarxiv.com/y876r

In this dissertation, I introduce the context needed to understand my research topic and the approach taken in the articles to reach their results. I also summarize the main findings, relate my research journey for this PhD thesis, and finally provide a general conclusion. My thesis supervisors, Jean-Claude Dreher and Edmund Derrington, as well as my co-authors, supported me with helpful advice and discussion throughout the whole thesis.

I hereby declare that all parts of this dissertation were written by myself, and co-authors when relevant. All sources have been quoted.

Résumé

L'une des difficultés pour comprendre l'Autre vient de l'inaccessibilité à son état mental : ses buts, croyances, désirs sont la majorité du temps inférés par l'observation de son comportement. L'ensemble des capacités cognitives qui permettent à un individu de se représenter l'état mental d'autrui est défini par la Théorie de l'Esprit. Longtemps considérée comme spécifique à l'espèce humaine, celle-ci reste encore peu comprise des chercheurs sous de nombreux aspects. En particulier, les mécanismes computationnels sous-jacents à la Théorie de l'Esprit restent encore mal caractérisés.

Le but de cette thèse est de proposer une analyse computationnelle des comportements sociaux, et plus particulièrement de montrer l'utilité de cette approche afin de caractériser spécifiquement différents aspects de la prise de décision sociale et de la Théorie de l'Esprit. Dans un premier temps, nous avons étudié l'apprentissage de la coordination dans un jeu de « guerre des sexes » simultané et transparent, i.e. qui permet l'accès visuel aux mouvements et décisions de l'autre joueur, chez les humains et les macaques. Nous cherchions à caractériser les paramètres d'influence ainsi que la nature des processus d'apprentissage et de décision mis en jeu dans une tâche sociale de coordination chez différentes espèces. Nous montrons que la nature du processus de décision diffère fondamentalement entre les 2 espèces.

Notre seconde expérience avait pour but de comprendre comment l'apprentissage de la coordination pouvait émerger dans un contexte non transparent chez des primates non-humains, des babouins. Grâce à un design expérimental novateur, nous avons montré que leur processus de coordination était mieux modélisé par un modèle d'apprentissage d'influence (« influence learning ») qui prenait en compte une représentation de l'influence du choix du sujet sur celui de l'autre, ainsi que la différence de dominance entre les 2 babouins en présence dans la tâche en condition sociale. En condition non-sociale, leur processus de décision se rapprochait plus d'une heuristique simple de « si je gagne, je continue/si je perds, je change » (Win-Stay/Lose-Switch).

Finalement, dans la dernière approche présentée dans ces travaux, nous avons cherché à étendre notre approche à l'apprentissage de la coordination chez les enfants. Nous avons étudié le développement de la capacité à s'adapter au changement d'intentions d'autrui lors d'une tâche de décision dans différents groupes d'enfants de différents âges. A partir de 8 ans environ, la majorité des enfants acquièrent cette capacité qui est décrite par le même modèle computationnel que celui retrouvé chez les adultes : le modèle d'apprentissage d'influence à intentions mixed (« Mixed-Intentions Influence Learning »).

Nos travaux s'appliquent essentiellement à l'étude de la coordination stratégique dans un contexte social, mais le principe de notre approche s'étend à tout type de prise de décision stratégique et sociale. La modélisation computationnelle nous a permis de tester et d'expliquer plus précisément les mécanismes décisionnels impliqués dans la Théorie de l'Esprit dans un panel très varié de participants et de situations. Cette thèse souligne la pertinence de cette approche à l'étude des comportements sociaux, et plus particulièrement

de la Théorie de l'Esprit, pour répondre aux nombreuses questions encore débattues dans la communauté scientifique.

Mots-clés : Modélisation computationnelle ; prise de décision sociale ; Théorie de l'Esprit ; Primate ; Enfants ; coordination

Abstract

One of the main difficulties to understand others stems from the inaccessibility to their minds: their goals, beliefs, desires are often inferred from their behavior. The set of cognitive abilities that are used by an individual to mentalize the internal state of mind of others is known as Theory of Mind (ToM). Considered as specific to humans for a long time, this ability is still puzzling researchers of various fields around the world. In particular, the mechanisms underlying ToM are still to be computationally characterized.

The aim of this work is to study social behavior under the lens of computational modeling. More specifically, we want to provide a solid and rigorous approach to characterize different aspects of Theory of Mind and strategic social decision making. Our first study deals with the modeling of coordination learning in a simultaneous and transparent version of the game "Battle of the Sexes" (BoS) in adult humans and rhesus macaques. We wanted to characterize computationally the parameters of influence and the nature of the learning process undergoing in a coordinative social task within different species. We show that, even with the possibility to see the movements and decisions of the other player at current trial, the decision process between the two species differs in nature.

Our second experiment's goal was to understand the computational mechanism underlying coordination learning, this time in a non-transparent setup in baboons (Papio papio). With a novel experimental design, we found that baboons could learn to coordinate by mentalizing their influence on the other's choice and that this process was modulated by their relative difference in dominance with the other baboon in the social condition. They were indeed better modeled computationally by an influence learning model that was modulated by their difference in ELO scores. In the solo condition, their decision process was closer to a simpler "Win-Stay/Lose-Switch" heuristic.

Finally, in the last project presented in this thesis, we extended our computational approach to coordination learning in children. More precisely, we studied the development of the ability to adapt to non-signaled changes of cooperative vs competitive intentions of others in different groups of children of different ages. Starting 8 years old, most children acquired the same computational decision process as the one previously found in adults: the Mixed-Intentions Influence Learning model.

Our work was mainly applied to strategic coordination learning, but the principle behind our method can be generalized to all types of interactions in the field of strategic social decision making. We used computational modeling to test and explain more precisely the decision mechanisms implicated in some aspects of ToM in a diverse panel of participants and situations. This thesis highlights the relevance of a computational approach to study ToM and provide fundamental elements to help answer the still ongoing debates on ToM in the scientific community.

Key-words: Computational modeling ; Social decision-making ; Theory-of-Mind ; Children development ; Primate ; coordination ;

Synthèse

L'une des difficultés à comprendre l'Autre vient de l'inaccessibilité à son état mental : ses buts, croyances, désirs sont la majorité du temps inférés par l'observation de son comportement. L'ensemble des capacités cognitives qui permettent à un individu de se représenter l'état mental d'autrui est défini par la Théorie de l'Esprit. Longtemps considérée comme spécifique à l'espèce humaine, cette capacité de méta-représentation, i.e. à se représenter les représentations d'autrui, reste encore peu comprise des chercheurs sous de nombreux aspects. En particulier, les mécanismes computationnels sous-jacents à la Théorie de l'Esprit restent encore mal caractérisés.

Par exemple, la présence de cette capacité chez d'autres espèces animales, et plus particulièrement le degré de sophistication dont elles seraient capables, fait encore débat au sein de la communauté scientifique : même si de plus en plus de comportements sociaux complexes sont rapportés chez les grands singes et, plus récemment, chez certaines espèces de singes, caractériser clairement la Théorie de l'Esprit reste une tâche complexe et piégeuse, qui peut parfois être soumise à de nombreux biais, parfois inconscients (e.g. d'interprétation, d'anthropomorphisme, de conviction personnelle, etc.) (Call & Tomasello, 2008; Martin & Santos, 2016; van der Vaart & Hemelrijk, 2014).

Certains des challenges rencontrés par les comportementalistes chez les animaux sont d'ailleurs partagés par les psychologues développementalistes chez l'humain. Ceux-ci s'interrogent notamment sur le développement de la Théorie de l'Esprit chez l'enfant, le processus par lequel ils l'acquièrent et/ou l'améliorent, la typologie des capacités cognitives mises en jeu, etc. Grâce à des batteries de tests qui requièrent toutes plus ou moins de capacités de méta-représentation, les principales étapes du développement de la Théorie de l'Esprit chez l'humain font désormais à peu près consensus (Hughes & Devine, 2015; Osterhaus et al., 2016; Rakoczy, 2022; Wellman et al., 2011; Wellman & Liu, 2004). L'émergence de la forme la plus basique de Théorie de l'Esprit commence un peu avant 1 an. Puis, une seconde étape clé se produit à partir d'environ 4 ans avec l'acquisition progressive d'une forme plus aboutie de Théorie de l'Esprit. Il s'agit de ce niveau de Théorie de l'Esprit qui fait régulièrement débat dans la communauté scientifique qui étudie la cognition sociale chez les animaux non-humains. Enfin, à partir de 7 ans environ, apparaît la Théorie de l'Esprit de haute sophistication, incluant notamment les raisonnements récursifs, la compréhension de l'ironie, du sarcasme, et des émotions les plus subtiles qui peuvent parfois être cachées ou fausses.

L'étude de la Théorie de l'Esprit est soumise à de nombreux challenges. L'une des principales difficultés est de dissocier la lecture de comportements (i.e. la capacité à déchiffrer les comportements des autres sur la base d'apprentissages associatifs) de la Théorie de l'Esprit qui fait usage de représentations d'états mentaux (van der Vaart & Hemelrijk, 2014). Une autre difficulté est d'établir comment mesurer cette capacité de manière valide (i.e. contrôlée et dénuée de tous facteurs confondants), réplicable (i.e. reproductible à grande échelle par différentes équipes), et cela sans nécessiter de verbalisation (en tout cas pour les études chez les animaux non-humains et les jeunes enfants) (Rakoczy, 2022). D'autant plus que la Théorie de l'Esprit est en réalité composée de plusieurs aspects. La représentation d'états mentaux de différentes natures, e.g. les désirs, les connaissances, les intentions, les perceptions, les

émotions, les croyances, etc…, ne font pas nécessairement appel aux mêmes processus cognitifs, d'où la nécessité grandissante de trouver une taxonomie précise permettant de distinguer entre les aspects de la Théorie de l'Esprit qui sont basés sur des mécanismes cognitifs différents (Beaudoin et al., 2020).

Plusieurs pistes ont été proposées pour remédier à ces différents problèmes. Afin d'identifier les mécanismes communs à la méta-représentation d'états mentaux de différentes natures, ainsi que ceux qui leur sont spécifiques, seules des méta-analyses exhaustives pourront aider à répondre à cette question (Beaudoin et al., 2020; Schurz et al., 2014). Et dans le but de faciliter la réalisation de telles méta-analyses, il sera indispensable dans chaque expérience de 1) préciser et dissocier clairement les états mentaux étudiés et leurs natures 2) s'intéresser à la méta-représentation d'états mentaux moins souvent étudiés dans la littérature, par exemple les intentions (Beaudoin et al., 2020). L'utilisation d'outils formels mathématiques, comme par exemple la modélisation et simulation informatique ou la théorie des jeux, peut par ailleurs s'avérer être un excellent moyen d'étudier la prise de décision et la Théorie de l'Esprit (Bello & Cassimatis, 2006; Rusch et al., 2020; van der Vaart & Hemelrijk, 2014).

Le but de cette thèse est donc de proposer une analyse computationnelle des comportements sociaux, et plus particulièrement de montrer l'utilité de cette approche afin de caractériser spécifiquement différents aspects de la prise de décision sociale et de la Théorie de l'Esprit. Nous nous intéressons dans cette thèse plus spécifiquement aux intentions coopératives et compétitives et à la coordination dans des jeux stratégiques. Nous avons ainsi été amenés à collaborer avec plusieurs équipes dans le monde : l'équipe d'Igor Kagan et Sébastien Möller en Allemagne pour la 1ère étude, l'équipe de Joël Fagot et de Nicolas Claidière à Aix-Marseille pour la 2ème étude, les équipes de Chen Qu en Chine, de Yansong Li en Chine, d'Alireza Soltani aux Etats-Unis pour la 3ème étude. Les données expérimentales ont notamment été collectées par ces équipes.

Mon rôle au cours de cette thèse a été essentiellement la conceptualisation des designs expérimentaux pour la 2ème et 3ème expérience, ainsi que l'analyse des données et la modélisation computationnelle des 3 projets. L'inspiration des designs expérimentaux de la 2ème et 3ème expérience vient essentiellement de 2 études : celle d'Igor Kagan et de Sébastien Möller et dont nous avons repris les données expérimentales pour notre 1er projet, et celle de Rémi Philippe, dans notre équipe à l'époque. Leur design initial et les limites et questions soulevées à la suite de ces études sont les principales sources qui ont inspiré ces expériences. En ce qui concerne l'analyse des données, le raisonnement et la méthodologie sont les mêmes pour les 3 expériences : après visualisation des données brutes de décision des participants, l'objectif est d'identifier les principaux facteurs d'influence prédisant la décision à l'aide de Régressions Linéaires Généralisées à Effets Mixtes (GLMER). Puis, après identification de ces facteurs, évaluer les potentiels mécanismes computationnels explicatifs du comportement observé dans la littérature, et éventuellement en construire à partir de ceux déjà existants. Après avoir appliqué ces modèles aux données (les "fits" de modèle) et nous être assuré de la discriminabilité de ces modèles les uns par rapport aux autres (en construisant les matrices de confusion), nous avons comparé nos différents modèles par Sélection Bayésienne de Modèle (BMS) afin de déterminer celui qui expliquait le mieux les

données comportementales des différents groupes de participants. Enfin, la dernière étape consistait à analyser, à l'aide des paramètres internes des modèles, les différences éventuelles entre conditions ou groupes de participants afin d'expliquer plus précisément ces différences.

Dans un premier temps, nous avons étudié l'apprentissage de la coordination dans un jeu de « guerre des sexes » simultané et transparent chez les humains et les macaques. Cette étude, basée sur les données acquises par Sébastien Möller et Igor Kagan dans le cadre d'une de leurs études précédentes, utilisait un setup expérimental basé sur un écran tactile transparent. Nous cherchions à caractériser les paramètres d'influence ainsi que la nature des processus d'apprentissage et de décision mis en jeu dans une tâche sociale de coordination chez différentes espèces. Celle-ci mettait en jeu de l'observation de comportement présent et passé afin de prédire des intentions et actions futures. Nous montrons que la nature du processus de décision diffère entre les 2 espèces. Les humains utilisaient un processus de décision, basé sur la nature de la cible (vs le côté), qui exploitait l'accès visuel au choix courant de l'autre grâce à la transparence de notre design expérimental alors que les macaques avaient des processus de décision plus divers. Globalement, nous avons observé chez les macaques 2 types de processus de décision : l'un exploitait l'accès visuel au choix de l'autre mais était basé sur le côté (et non sur la nature de la cible comme observé chez les humains). Le 2nd processus de décision, cette fois basé sur la nature de la cible, n'exploitait pas l'accès visuel, mais plutôt une représentation de l'influence du choix du sujet sur celui de l'autre.

Notre seconde expérience avait pour but de comprendre comment l'apprentissage de la coordination pouvait émerger dans un contexte non transparent chez des primates non-humains, des babouins dans ce cas précis. Notre design expérimental novateur, rendu possible par l'équipement expérimental de l'équipe de Joël Fagot sur le site de recherche du Rousset, permet l'interaction sociale entre 2 babouins dans des tâches de prise de décision (Claidière et al., 2017; Fagot et al., 2015; Fagot & Bonté, 2010; Fagot & Paleressompoulle, 2009; Gullstrand et al., 2021). Leur participation était libre d'accès et volontaire, avec un équilibre subtil entre la validité écologique et le contrôle des actions et de la répétabilité de l'expérience. L'acquisition automatique des données par un système ingénieux de détection et d'identification s'est révélé particulièrement salvateur pendant la période de pandémie de covid-19. Nous avons montré que le processus de coordination des babouins prenait en compte une représentation de l'influence du choix du sujet sur celui de l'autre modulée par la différence de dominance entre les 2 individus en présence dans la tâche en condition sociale. En condition non-sociale, leur processus de décision se rapprochait plus d'une heuristique simple de « win-stay/lose-switch » (si je gagne, je refais le même choix, sinon je le change).

Finalement, dans la dernière approche présentée dans ces travaux, nous avons cherché à étendre notre approche à l'apprentissage de la coordination chez les enfants. Cette étude, en grande partie inspirée par les travaux de Rémi Philippe (Philippe et al., n.d.) chez les adultes humains, s'intéressait au développement de la capacité à s'adapter au changement d'intentions d'autrui lors d'une tâche de coordination dans différents groupes d'enfants de différents âges : 3-6 ans, 6-7 ans et 8-9 ans. La collecte des données s'est faite en collaboration avec 2 équipes en Chine qui ont testé 192 participants, et l'analyse des données en collaboration avec l'équipe d'Alireza Soltani qui a apporté son expertise dans l'utilisation de

mesures de Théorie de l'Information appliquée à la comparaison des comportements. Ces enfants ont joué dans leur classe sur ordinateur, croyant interagir avec leur camarade placé devant eux sur un autre ordinateur. En réalité, ils interagissaient à leur insu avec un algorithme qui alternait entre 2 modes, compétitif et coopératif, par blocs. Nous avons trouvé que les 2 plus jeunes groupes d'enfants se démarquaient de celui le plus âgé sous plusieurs aspects. En particulier, avant 8 ans environ, leur processus computationnel de décision était basé sur un apprentissage simple bayésien. A partir de 8 ans environ, la majorité des enfants acquièrent cette capacité qui est décrite par le même modèle computationnel que celui retrouvé chez les adultes par Philippe et al.

En conclusion, nos travaux suggèrent que le niveau de sophistication de la Théorie de l'Esprit chez des populations limitées verbalement, comme par exemple les primates non-humains ou les jeunes enfants, peut, avec le bon équipement et un design expérimental cohérent, être testé et caractérisé dans des jeux de décision par l'approche computationnelle décrite dans nos projets. Nos travaux s'appliquent essentiellement à l'étude de la coordination stratégique dans un contexte social, mais le principe de notre approche s'étend à tout type de prise de décision stratégique et sociale. La modélisation computationnelle nous a permis de tester et d'expliquer plus précisément les mécanismes décisionnels impliqués dans la Théorie de l'Esprit dans un panel très varié de participants et de situations. Cette thèse souligne la pertinence de cette approche à l'étude des comportements sociaux, et plus particulièrement de la Théorie de l'Esprit, pour répondre aux nombreuses questions encore débattues dans la communauté scientifique.

Table of content

List of Figures

Acknowledgments

I would like to take the time in this section to express all my gratitude to all the people who have supported me and my work during these years, or whom I just shared some nice moments with, namely:

The reviewers of my manuscript, Mehdi Khamassi and Hannes Rakoczy, who have taken the time and made the effort to read my manuscript and to provide helpful critiques to my work.

The juries of my PhD thesis, Catherine Crockford, Jean Daunizeau, Valérie Dufour, Alice Gomez, Mehdi Khamassi and Hannes Rakoczy, who have accepted to attend to my PhD defense.

My PhD supervisors, Jean-Claude Dreher and Edmund Derrington, who have not only guided my work and shared with me their advice and experience, but with whom my team and I have also spent some quality time at their place. I hope that we will keep this nice and easy relationship in the future.

My collaborators, Nicolas Claidière and his colleagues, Sebastian Möller and Igor Kagan, Yansong Li and his colleagues, Chen Qu and her colleagues, Jean-Baptiste Van der Henst, Alireza Soltani and Jae Hyung Woo, without whom I would have never been able to work on such amazing research projects.

My colleagues, and now friends, from my team, my institute, and the research community in general, with whom I have discussed, learnt, and thought a lot. But our interactions were not limited to research- or work-related subjects, far from that. They also, and most importantly, included unforgettable friendly moments. The heated debates and funny discussions on any and every subject during breaks, the long and intense sessions of table football that concluded almost every lunch breaks, the regular evening parties at bars or at someone's place, the social, and often food-related, – to our stomachs' and palates' great pleasure – events at the institute, the other amazing social events outside of the institute that included escape games, restaurants, laser games, karting, BBQs, badminton sessions, martial arts trainings, etc...

During my last few years, I am glad that I have managed to maintain a balanced life, with plenty of hobbies and sports. I am grateful to have had the occasion to spend nice quality time together with lots of amazing people through different activities. This includes:

My friends of the "Zoo" with whom I have also spent wonderful times playing RPG, TCG card games (Long live Yugi & Magic !), video games (Smash Bros ;p), board games, etc.

The Mudo club, which is supervised by my Master Lee Yong Seon, and with whom I have exchanged kicks, kihaps and food. Special thoughts also to the Taekwondo enthusiasts that I have met at the preparation of my DIF, at the ENVOL special trainings, and to Feyzin Taekwondo Club with whom I have spent my internship to train and learn how to train.

Tricking Lyon, with whom I have shared sweat, laughs, and B-twists. I hope that we will keep training and improving together so that we can all master all variants of corks (we all counting on you, François 😉) !

My group of (Centrale)Supelec friends, with whom I have spent unforgettable years at our engineering school in Metz, and whom I can always crash by when I come to Paris :D

My childhood friends, with whom I could not meet very often, but hopefully we will find some time to remember our good old days in the near future.

My family (brother, sister, cousins, parents, aunts and uncles…), who have supported me since my birth and through the years. You have raised me and have actively participated in molding the man I am today at his very core.

You have all participated in my development and helped me improve myself. Thank you.

Publications and communications

A. Publications

T. Nong, S. Möller, R. Philippe, E. Derrington, I. Kagan, J.-C. Dreher. Differences in computational decision processes between human and macaque in a transparent Bach-or-Stravinsky Game. *In preparation*.

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Mentalising underly strategic coordination in Guinea baboons (Papio papio). *In preparation*.

T. Nong, Y. Li, C. Qu, J. H. Woo, Y. Wang, C. Miao, X. Liu, R. Philippe, J.-B. Van der Henst, E. Derrington, A. Soltani, J.-C. Dreher. Computational mechanisms underlying the emergence of theory-of-mind in children. *Submitted*. Psyarxiv: https://psyarxiv.com/y876r

B. Oral presentations

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Computational mechanisms used for coordination learning in free ranging baboons (Papio papio), *vSNE 2021, 19th Annual Meeting of the Society for NeuroEconomics*, virtual, September 29 – October 1, 2021.

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Computational model of strategic coordination learning in baboons is shaped by social ranks, *Scientific Day of the Doctoral School of Cognitive NeuroScience*, Lyon, France, June 27, 2022.

C. Poster presentations

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Computational mechanisms used for coordination learning in free ranging baboons (Papio papio), *Scientific Day of the Doctoral School of Cognitive NeuroScience*, virtual, January 29, 2021.

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Computational mechanisms used for coordination learning in free ranging baboons (Papio papio), *SBDM 2021, 10th Symposium on Biology of Decision-Making*, Paris, France, May 10-12, 2021.

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Computational model of strategic coordination learning in baboons is shaped by social ranks, FENS 2022*, Forum of European Neuroscience of the Federation of European Neuroscience Societies*, Paris, France July 9-13, 2022.

T. Nong, N. Claidière, J. Fagot, R. Philippe, E. Derrington, J.-C. Dreher. Computational model of strategic coordination learning in baboons is shaped by social ranks, *Workshop on Comparative Neurobiology of Higher Cognitive Functions*, Erice, Italy, September 9-14, 2022.

T. Nong, C. Qu, Y. Li, Y. Wang, C. Miao, X. Liu, R. Philippe, J.-B. Van der Henst, E. Derrington, J.-C. Dreher. Computational mechanisms underlying adaptation to fluctuating intentions of others in children, *Neurofrance*, Lyon, France, May 24-26, 2023.

T. Nong, C. Qu, Y. Li, J.H. Woo, Y. Wang, C. Miao, X. Liu, Q. Liu, J.-B. Van der Henst, E. Derrington, A. Soltani, J.-C. Dreher. Submitted. COMPUTATIONAL MODELING OF THE ABILITY TO READ CHANGES OF COOPERATIVE VS COMPETITIVE INTENTIONS: DEVELOPMENTAL CHANGES IN CHILDREN, *SBDM 2023, 11th Symposium on Biology of Decision-Making,* Paris, France, June 1-2, 2023.

T. Nong, C. Qu, Y. Li, J.H. Woo, Y. Wang, C. Miao, X. Liu, Q. Liu, J.-B. Van der Henst, E. Derrington, A. Soltani, J.-C. Dreher. Submitted. COMPUTATIONAL MODELING OF THE ABILITY TO READ CHANGES OF COOPERATIVE VS COMPETITIVE INTENTIONS: DEVELOPMENTAL CHANGES IN CHILDREN, *Scientific Day of the Doctoral School of Cognitive NeuroScience*, Lyon, France, June 21, 2023.

CHAPTER 1: General Introduction

"If we were all given by magic the power to read each other's thoughts, I suppose the first effect would be almost all friendships would be dissolved; the second effect, however, might be excellent, for a world without any friends would be felt to be intolerable, and we should learn to like each other without needing a veil of illusion to conceal from ourselves that we did not think each other absolutely perfect."

— Bertrand Russell, The Conquest of Happiness

Theory of Mind (ToM), also referred to as 'mentalizing', 'meta-representation' or 'mindreading', has been fascinating scientists and philosophers for centuries. Often considered as one of the most essential capacities for complex social interactions, understanding mental states (also called mental representations) is fundamental to navigate in a complex social world [1]. Strangely enough, it was not until the late 70s that scientists began to thoroughly address the question of ToM [2]. They first questioned the specificity of this ability to humans, which led to many more questions. From then on, Pandora's box was opened, and a myriad of studies about ToM, its origins and its development began.

This work aims at providing some elements to contribute to some of the ongoing debates. In particular, the processes and computations engaged to perform ToM, its presence in non-human primates, and its development in human children are hotly debated. Some of the specific questions concern: what are the computational processes that underly ToM? To what extent do they differ between human adults and non-human primates? Between adults and children? Answers to these questions might shed light on the origins (evolutionary and ontogenic) and on the mechanisms behind ToM.

To approach these questions, I will first introduce and briefly review the history (A.1) and state-of-the-art of ToM in neurotypical children (A.2) and in non-human primates (A.3). This will allow us to determine the challenges and limitations in ToM research. We will first see that defining precisely Theory-of-Mind is not such an easy task (B.1). Interpretations of ToM experiments can besides be a bit tricky (B.2). This has led to a validity and replicability crisis in some subfields of ToM research (B.3). This crisis might actually stem from interfering intuitions and biases (B.4) as well as from problems of ecological validity (B.5). This will lead us to consider appropriate approaches to study ToM. The first step to tackle these questions being to acknowledge the multi-faceted nature of ToM (C.1).

I have then proposed an approach to solve the issues commonly found in this field. The first step to tackle these questions being to acknowledge the multi-faceted nature of ToM (C.1). This will allow us to define better the scope and context in which I will study ToM (C.2) and then I describe the tools and methodology that I have used in my thesis to characterize specific aspects of ToM (C.3).

A. Theory of Mind

ToM is the ability to attribute mental states, preferences, beliefs and intentions to other individuals, and to act accordingly. In this section, I will first briefly remind some historical notions of ToM literature to the reader. I will then review briefly the literature concerning its possible existence in non-human animals (later referred to as animals in this work) and more specifically non-human primates. In the third subsection, I will sum-up current knowledge about ToM development in children.

1) The origins of ToM research

Historically, the concept of ToM was introduced by Premack and Woodruff in a study in which they asked to which extent Chimpanzees represent mental states of others[2]. This study paved the way to heated and rich debates among scientists and philosophers. Their original definition of ToM as the ability to attribute mental states, such as intentions, beliefs, knowledge or desires, to oneself and others, is still used today. They used the term "theory" to describe this concept because 1) the mental states of the "mind" are not directly observable, 2) this concept can be used to make predictions about the behavior of others.

Originally, they showed short movies of physical challenges faced by a human actor to a chimpanzee, named Sarah. Then, after each movie, they presented photographs to her. Among these photographs, there was a solution to the challenge. For example, one video clip showed the human actor locked in a cage and struggling to escape. They alternated between grasping the bars of the cage and the padlock on the door. A key was shown among the images proposed to Sarah, and she successfully selected it from among the other images. During the experiment she consistently chose the photograph that contained the solution to the challenge presented in the video clip. The soon-to-be controversial interpretation of Premack and Woodruff was that Sarah could recognize the problem shown in the video, understand the goal of the actor, and choose solutions adapted to the experimental settings.

However, other researchers objected that Sarah was possibly not representing the goals of the human, but simply able to predict a behavioral sequence that was likely to happen, based on some perceptual similarities between the clips and the solutions, or, on associative learning [3]. This controversy triggered a bloom of ToM studies in the 80s and 90s, not only in animals [4]–[8], but also in the developmental psychology of children [9]–[14] that I will shortly describe in sections A.2) and A.3) respectively. The main view shared by the majority at the end of the 90s was that there was, until then, no convincing proof that non-human species had ToM abilities [15].

A fundamental shift occurred at the beginning of the 2000s in primate studies: the methodology used in the 90s was deemed too artificial and ecologically invalid to reveal the true cognitive abilities of primates [16]–[18]. Tasks in which animals have to cooperate with a human to obtain food can indeed be considered out of context for species in which social life is regulated, in the majority, by competition with conspecifics [19]. Therefore, many studies started to focus on competitive tasks between conspecifics in more ecologically valid contexts. For example, Hare and colleagues [20], [21] tested whether a subordinate chimpanzee would prefer to grab a piece of food that was visible to a dominant conspecific or one out of view of the dominant conspecific. The hidden piece of food was preferentially selected, which led the

authors to the conclusion that chimpanzees could at least understand what others can and cannot see in the present [20] as well as in the recent past [21].

Although criticized, this study was one of the ones that renewed the approach of ToM studies in non-human primates. Since then, ToM has been studied under multiple lenses, such as in psychological and behavioral economics research [9], [22]–[24], with standard neuroimaging methods [24]–[26], or with computational models of behavior [27]–[33]. I will however not review all of these aspects in this thesis, but only those essential for the readers to understand my work and situate my research in the appropriate context. I will for example mainly focus on ToM studies in neurotypical children and primates, not so much on studies in non-typically developing humans, e.g. autistic children, or in non-primate animals such as dogs or corvids, unless such studies are particularly relevant.

2) Development of ToM in children

The controversy in the 80s concerning Premack and Woodruf's study also initiated a bloom in the study of ToM in developmental psychology. This led to the advent of a cluster of experiments that continue to constitute the canonical test for ToM in the opinion of many psychologists: False Beliefs (FB) tasks. These experiments were primarily designed to test ToM in children [9], [12]–[14]. In these studies, participants, (here children), witnessed a scene in which a first character hides a toy in a given location before leaving the scene. A second character, that was present the whole time, and that also apparently saw the first character hide the toy, then changes the toy's location while the first character is away and cannot see the change of location. Finally, the first character returns to the scene. The participants were asked where the first character will search for the toy. Children younger than 4 years old systematically failed to answer correctly, as they were apparently unable to take into account the false belief that the first character would presumably have about the location of the toy when she comes back. This location-based FB task was the first of its kind and remains the standard test to measure ToM abilities.

A comparatively vast number of different tasks have now been tested on children to evaluate their ToM abilities, which has led to the development of scales for ToM tasks [34], [35]. These scales indicate how children's understanding of different mental states generally develop with age, e.g., desires vs beliefs, ignorance vs false beliefs, etc. Usually, a stereotyped developmental progression is observed [34], [35]. A large literature suggests that ToM actually fully emerges at about 4 to 5 years old, during the "4 year revolution" [36]. At this age, children acquire new concepts that trigger the emergence of a fully-fledged meta-representational ToM. In particular, they start to succeed in a battery of tasks that all require meta-representation, including FB tasks. These FB tasks include standard location FB, but also unexpected content FB [37], aspectual FB [9], and appearance-reality FB [37]–[39] (see Figure 1 from [36] for a description of the tasks). All of these tasks measure if and how the child is able to ascribe to another agent a subjective misrepresentation that differs from and is incompatible with their own perspective. For all of these tasks, children only start to succeed from about 4 years old, with the same overall pattern of ToM mastery across cultures [40]–[43], although with some slight differences in the sequence of ToM steps that vary as a

function of, for example, the collectivist vs individualist nature of the culture [40], [41].

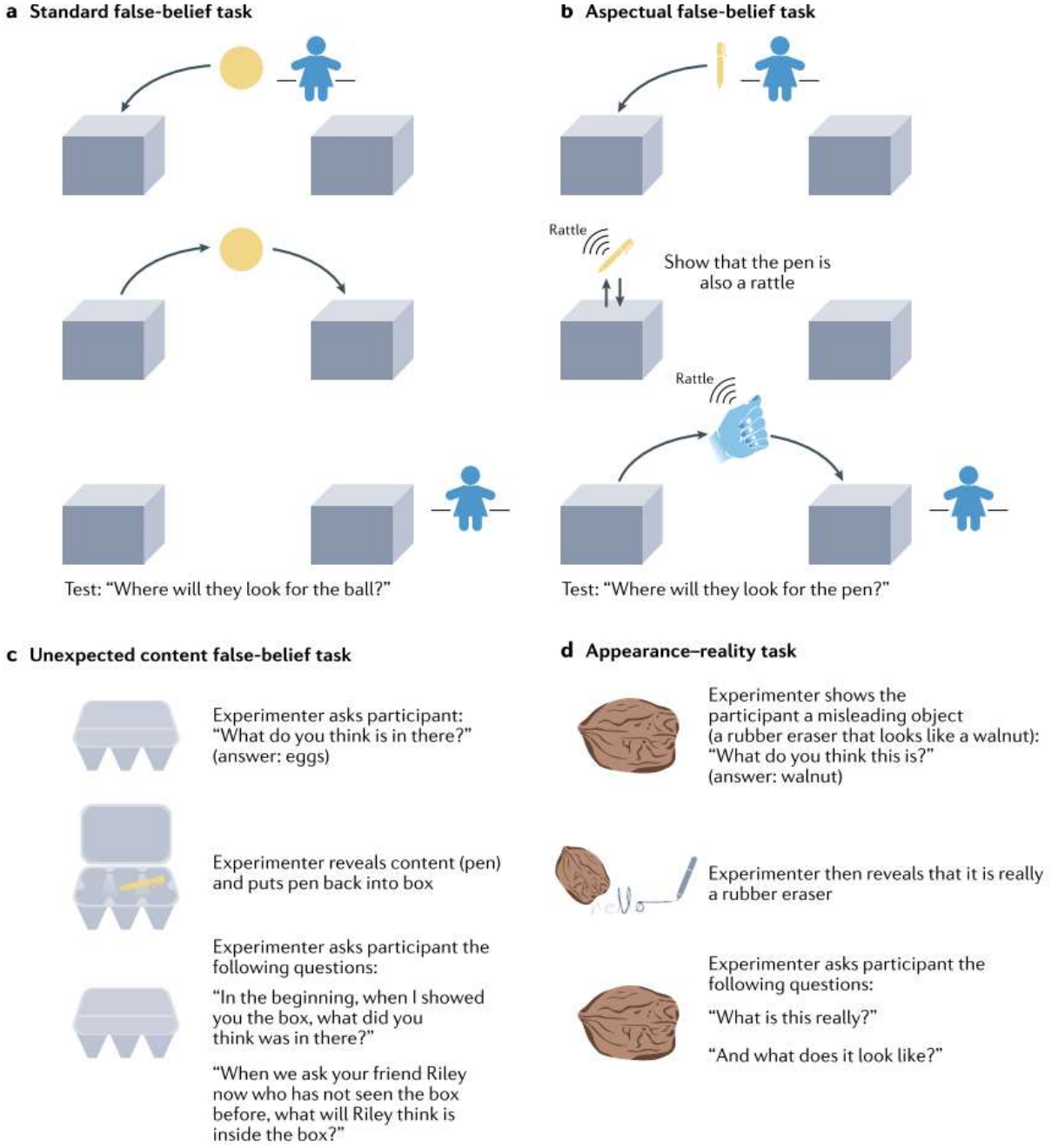


Figure 1 : ***False-belief and related theory of mind tasks.*** *a | The child (participant) sees another agent put an object into a box. The agent leaves and the object is transferred to another box. The child is then asked where the agent will look for the object. To answer correctly, the child needs to meta-represent how the agent represents the location of the object. b | The child watches an object with two identities (a pen that is also a rattle) be put into one box while shown as a pen and then transferred to another box while shaken as a rattle. The child therefore knows that the pen/rattle is now in the second box. Another agent also witnesses the initial placement of the object (as a pen) in the first box and its relocation (as a rattle) to the second box but is unaware of the dual identity of the object. The child is then asked where the agent will search for the pen. To answer correctly ('The first box'), the child needs to meta-represent how the agent represents the placement and relocation of the object. c | The child sees a familiar container (such as an egg carton), is asked what is inside (answers 'eggs'), and then learns about the unexpected content (a pen). The child is then asked what they initially thought was in the box, and what another naive agent would think is in the box. To answer correctly, the child has to meta-represent how the world (wrongly) appeared to them previously, or would appear to another person. d | The child initially sees a misleading object (such as a rubber eraser that looks like a walnut) and is asked what she thinks the object is (answers 'walnut'). The child then learns that the object is actually something else (a rubber eraser). The child is*

then asked what the object really is and what it looks like. To answer correctly, the child needs to meta-represent how the object appears, by contrast to what it really is. From Rakoczy (2022) [36].

ToM development does not stop at 4-5 years old. Understanding of mental states indeed gets enriched throughout childhood and even adolescence, with the experience of more complex social situations, the further development of language, and the improvement of executive functions that foster more sophisticated representations of mental states. Advanced Theory of Mind (AToM), as opposed to first-order ToM, that I described above, has been developed to describe the multiple aspects of an advanced understanding of the minds of others [44]–[47]. Higher-order false-belief understanding [11], [48], social understanding [49], perspective-taking abilities [50], and emotion- and mental-state recognition [51] are examples of AToM abilities. This set of capacities, apparently specific to humans, requires both first-order ToM abilities and improved general cognitive abilities. Distinct constructs and concepts such as reasoning in recursion, and the understanding of interpretations, which may differ from individual to individual, are needed for the development of AToM [52].

Research suggests that children begin to understand more complex forms of mental-state reasoning on starting school, i.e., around 6 years old [45], [53], and reach a milestone at 7 years old when children attain the core conceptual insight that mental states can be recursive [45]. For example, the realization that someone can hold a false belief about a belief of someone else, such as A thinks that B thinks that x… The acquisition of the concept of recursive reasoning might be the main factor that allows children to solve AToM tasks as proposed by the conceptual-development hypothesis [10]. Thus, the acquisition of different concepts, e.g., that of recursion, may be the principal driver of the development of AToM abilities in children. A second hypothesis, which is not necessarily mutually exclusive, is that improvement on AToM tasks with increasing age is simply a consequence of increased general information-processing capacities. This increase may be necessary to resolve the relative complexity of AToM tasks [54]. Evidence in favor of conceptual development has been obtained, at least for AToM abilities that involve reasoning [45].

At the opposite end of the spectrum, some premises of ToM are observed in infants as young as 9 months old. New capacities can be identified at this age that allow infants to "track others' perceptions of their surroundings, the goals they have, and how they pursue the latter in light of the former" [36], also known as perception-goal psychology [55]. From then on, children become able to represent other agents' perception [56], pursued goals and intentions [57], [58], and even to form expectations about rational actions that other agents will perform to achieve their goals, based on the agents' perceptual access [59]. In the second and third years of life, their understanding of how different agents can have different goals and preferences increases in flexibility and in sophistication [34], [60], as does the child's ability to track the possibility that different agents hold different perceptions. This pnenomenon is also called Level-I perspective taking [61]. Notably, children become able to differentiate between intentional and accidental events [58]. Although the ability to acknowledge the preferences and visual perspectives of different agents can be found as early as in 2-year-olds, this ability does not necessarily involve meta-representational ToM, but only basic pre-mentalizing ToM. Fully-fledged ToM requires the ability to reason about conflicting versions of reality.

Some findings go even further, and provide evidence that children as young as 2 years old actually have full ToM abilities, as stated in the review of Baillargeon et al [62]. Two-year-old children would be able to represent mental states but fail to explicitly answer FB tasks correctly, possibly as a result of conceptual, linguistic, executive-function, or other limitations.

Indeed, most standard ToM tasks test explicit ToM with elicited-response tasks that require the children to directly answer or select the correct option. But paradigms that were developed to test implicit ToM, through spontaneous-response tasks, suggest that the ability to attribute FB might be at 2 years old or even younger [63]–[68]. These paradigms exploit spontaneous behaviors of the infants, based on measures such as looking time or gaze direction, while the infant observes a scene unfold, to infer their understanding of an agent's FB. Different principles underlie such paradigms, such as violation-of-expectation (VOE) tasks [63], [64], [66], [67], anticipatory looking (AL) tasks [65], [69]–[72], intersection tasks [73], [74], or altercentric interference tasks [22], [75]. The failure to perform explicit elicited ToM tasks in children younger than 4 years old would be due to an overwhelming cognitive demand to execute false-belief representation, correct response selection (i.e., the FB location), and real response (i.e., the actual location) inhibition simultaneously. In contrast, success in implicit spontaneous ToM tasks only requires the execution of false-belief representation, which would be cognitively tractable for 2-year-olds [62].

These implicit ToM paradigms in infants have greatly inspired research in other fields, especially social cognition in non-human primates. They have notably bloomed in the second wave of further studies in primate ToM, that challenged the previous consensus of the 90s, that only humans were capable of ToM.

3) ToM in non-human primates

Despite Premack and Woodruf's study [15], the majority of the scientific community considered ToM to be unique and specific to human cognition. However, new approaches and tools, partly inspired by ToM studies in young children, reset the cards for ToM research in animals. Thus, more recent comparative studies have investigated whether ToM is specific to humans, or whether it is shared with different species of non-human primates [2], [76], [77]. Non-human primates (NHP) are classified into different families based on their evolutionary proximity with humans (from closest to furthest: Great Apes, Old world monkeys and New world monkeys). Old world monkeys and New world monkeys have generally failed to show ToM-based behaviors [21], [76], [78]–[82]. Nevertheless, this question remains controversial, because recent works show that some species of monkeys and apes appear capable of at least precursor forms of ToM, such as awareness relations, or attention/intention reading abilities [77], [82], [83]. Seyfarth & Cheney [84] studied rudimentary forms of affective ToM in wild female baboons (Papio hamadryas ursinus). They focused on the degree to which baboons could attend to other individuals' social interactions and empathize with them. They showed that female baboons were not only able to be reflexively attentive and sensitive to others' attention, emotions and intentions, but also that the more they displayed such behavior, the more evolutionary adaptive they proved to be. "Nicer" females, that were friendly to all females, irrelative of ranks, anticipated more the threat of potentially dangerous males that immigrated in the group, and that would eventually attack infants of the group. They also had more grooming partners compared to loner females, that were unaffected by close-relative's death. Their ability to share attention and to be sensitive to others' emotional states constitutes a cornerstone of the more sophisticated meta-representational ToM that is found in humans.

Another interesting approach from Devaine et al. [85] was to characterize computationally the degree of sophistication used by individuals in strategic decision-making tasks. In their essay, they compared seven different species of primates, from the lemurs' taxa to the great apes, in competitive and cooperative decision games, Hide-and-Seek and Matching-Pairs respectively, with a human caretaker. They found that the nature and sophistication of the primates' decision learning process mainly depended on the mean brain size of their species, rather than the mean social group size. The primates in the smaller brain group had their decision process best explained by a reinforcement learning model based on past actions and outcomes to learn the best course of action to maximize reward. Basically, their decision making process takes into account the reward feedback and a computed expected reward at each trial to calculate a prediction error, then updates the expected value of each possible action [86], [87]. In contrast, primates in the bigger brain group, were best fitted by a model that accounted for a rudimentary mentalizing term that represented the influence of one's own choice on the decision of the other, namely the Influence Learning model [28], [29]. This study, besides showing that certain species of primates are able to perform at least precursor forms of ToM reasoning in strategic decision games, also provides an interesting computational approach that formally defines sophistication levels of decision processes in mathematical terms. This allows the authors to compare models of different levels of sophistication against each other. This study partly inspired my own approach in this thesis work.

According to studies by some other groups, several ape species might be capable of full-blown ToM abilities. For instance, partly inspired by ToM studies in very young children, Krupenye et al. [88] tried to show implicit ToM in great apes (bonobos, chimpanzees and orangutans) by recording their gaze direction while they watched videos of socially engaging scenes that involved ToM (Figure 2). A video would for example show a human actor in conflict with a costumed gorilla. The actor would search for the gorilla, who had originally hidden in one of two locations while the actor was watching but later left the scene while the actor was away. This scenario constitutes an equivalent of a False Belief test adapted to apes. The apes anticipatingly looked at the location where the actor falsely believed that the gorilla would be, even though they knew the real location of the gorilla had changed. This constituted evidence that apes might not only be sensitive to others' knowledge or ignorance, but also to their beliefs, and might therefore be capable of implicit ToM.

Figure 2 : ***Eye-tracking setup used by Krupenye et al.*** [88]***.*** *An infrared eye-tracker non-invasively records an orangutan's gaze as she attends to pictures and videos presented on a monitor, just outside of her enclosure. A nearby laptop controls stimuli presentation. Subjects voluntarily approach the setup to view the stimuli and sip a slow stream of juice. Copyright: Christopher Krupenye and Leipzig Zoo. From Krupenye et al (2019)* [83]*. Reproduced with consent.*

Four years later, Hayashi and colleagues [89] went a step further by using eye-tracking technologies in combination with chemogenetics to show implicit ToM in monkeys (Japanese macaques). They used a similar anticipatory looking version of the location False Belief task by showing monkeys three different movies. In these movies, an agent observed where an object was hidden, between two possible locations. In the agent's absence, the object was removed, thereby generating a false belief in the agent. Two behavioral measures, the first location that was fixated by the monkey and the difference in time spent fixating the two locations, that the monkeys behaved in this task as if they expected the agent to approach the false-belief location. Furthermore, they tested the causal role of the mPFC in this process by inactivating it using specifically designed drugs (DREADDs technology [90]). They showed that the macaques' anticipatory looking toward the false-belief location was eliminated by disruption of the mPFC function. Control experiments enabled the authors to rule out that the inactivation of the mPFC altered attention or short-term memory. Nor did it modify the animals' ability to track moving stimuli or to guide their eye movements based on memory. These results not only suggest that Japanese macaques are able to use ToM, but also that the mPFC plays a pivotal causal role in ToM-based behavior by encoding actionable information about others' behaviors.

The debate concerning the presence of full-blown ToM in non-human primates (NHP) continues, mainly focused on challenges to methodology and interpretations of results. Many reviews have highlighted the difficulties and challenges of studying ToM in non-human primates [77], [82], [83], [91]–[94]. Some of these difficulties are shared with studying ToM in infants and children, as we will see in the following section.

B. Challenges and limitations in ToM research

Although comparative studies provide important information regarding whether different non-human primate (NHP) species or very young children attribute mental states to other agents, many issues remain. Full-blown ToM, understood to be a mature meta-representational ability to "represent that, and how, other agents represent the world from their own point of view" [36], also called the mentalist account, has not yet been proven in any species other than humans [77], [82], [83], [91]–[94]. Even in humans there remains debate concerning the replicability and validity of implicit ToM tasks in children [36], [95]. The current state of debate on ToM in monkeys, great apes and humans has been rapidly summarized in Figure 3. In this section, I will present the main challenges and limitations encountered in ToM research, especially in infants and NHPs. I will first describe the difficulty to clearly define ToM and tasks that can measure it. Consequently, the interpretations of the results of these tasks suffer from challenging limitations, that I will present in the second subsection. Adequate controls are absolutely necessary to counter intuitive biases, as we will see in the third subsection. I will then address the matter of ecological validity in ToM research. Finally, I will highlight some difficulties to replicate and/or validate some tasks and measures in the field. Altogether, this will allow us to determine how to proceed from the current "state of the art".

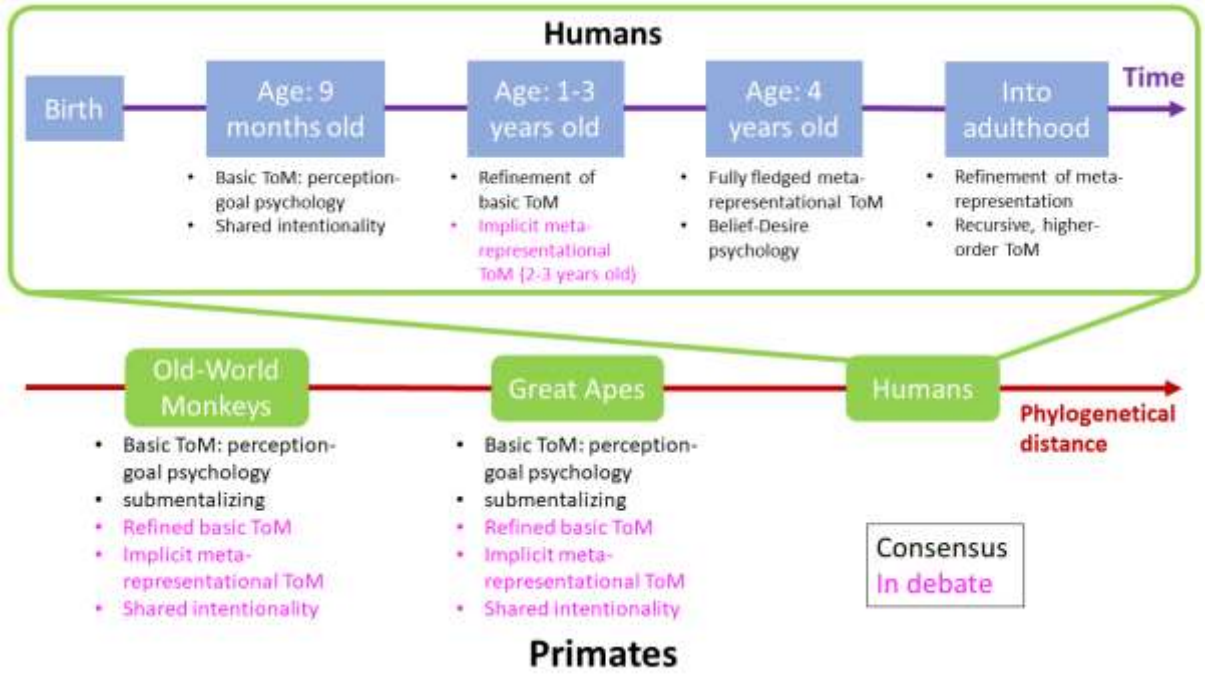


Figure 3 : ***Theory of mind (ToM) in Non-Human Primates (bottom) and milestones in its development in humans (top).*** *Top: The most basic form of theory of mind (ToM), perception–goal psychology, emerges from around 9 months of age. It allows agents to represent that others may have different perceptual perspectives on the world, and different goals, and act accordingly. Fully fledged meta-representation emerges later, from around 4 years of age, in the form of belief–desire psychology. It involves an appreciation that others subjectively represent the world in fine-grained aspectual ways that may be incompatible with one's own view and that may be false. The presence of implicit ToM abilities at ages younger than 4 years old is debated (in pink). Bottom: NHPs include, from closest to furthest phylogenetic distance to humans (Hominins), great apes (Hominoids), old-world monkeys (Catarrhines), new-world monkeys (Platyrrhines), and lemurs (Strepsirrhines). Here, I only included great apes and old-world monkeys to be concise but also because they are the most studied in the literature. The consensus is that they are capable of at least basic ToM, i.e. perception-goal psychology or submentalising* [94]*. Debate*

remains with respect to their ability to actually implicitly mentalise (in pink). For a complete review on ToM in NHPs, see Lewis & Krupenye (2022) [93]. *Adapted from Racokzy (2022)* [36].

1) A problem of definition

A difficulty that is too rarely mentioned in the field is the problem of clearly defining the terms used to describe Theory-of-Mind. Words such as representation, belief, knowledge, desires, intentions, goals, etc. can be interpreted in a variety of ways to suit different points of view and support different positions. Indeed, even philosophers, who have studied these problems for centuries, cannot agree upon the precise meanings of these terms or the general properties of these mental states. The definition that is most often used to describe ToM is "the ability to attribute mental states [such as desires and beliefs] to others, to make predictions", or several paraphrased version of it [2], [83], [92], [93]. This definition might sound clear at first, and suggests a single, unified, clear-cut cognitive process, which might be true to some extent, however, on closer inspection, it allows a wide range of interpretations, [96]. Some authors have therefore insisted on the importance of meta-representation as the core of ToM, to achieve higher clarity on what constitutes ToM [36]. Still, this definition does not answer the question of the definition of the mental states that are meta-represented. To my knowledge, terms such as "desires", "intentions", "knowledge", "preferences" or "emotions" are almost never clearly defined, and quite often, ToM tasks might tackle more than one of these mental states. For example, what would be the difference between ToM tasks that would involve "intentions" but not "goals"? The definition of these terms might also vary depending on the authors and context. Moreover, theoretically it would be possible that some species, or that children at a certain stage of their development, might be able to represent only some categories of mental states and not others. I will not define these terms in this dissertation, as this would go way beyond the scope of this work. In the following sections, I will instead use the terms that the authors themselves have used to describe their studies.

When defining ToM, yet another clarification is needed, that concerns the gradation in the depth, or level, of reasoning required for each ToM ability, also referred to as their "degree of sophistication" [96]–[99]. The difficulty is that ToM studies between non-human primates, non-primate animals, and human children can differ enormously with respect to the degree of ToM sophistication that they are considering. On the one hand, non-human animal studies often consider very simple forms of ToM, sometimes referred to as precursors, or sometimes not even ToM, but only abilities thought to be necessary to develop ToM. In contrast, developmental studies in children can deal with full-blown ToM abilities, sometimes including advanced degrees of sophistication of ToM. I think that it would therefore be more correct to talk about multiple ToM abilities, or components, rather than one unified ToM ability, and that each of these components might resolve more refined categories of mental states, at different degrees of sophistication.

In conclusion to this section, as Schaafsma et al. [96] wrote in their review: "some usages of ToM pertain to early cognitive development, whereas others pertain to adult social cognition; some refer to understanding of the self, whereas others refer to the perception of others; some refer to logical inferences, whereas others refer to emotional or empathic reactions. The term ToM is used interchangeably with mentalizing or mindreading, mind perception, and social intelligence, to name only a few. This diversity of terms used is probably

telling: different investigators have different concepts in mind." The big challenge will be to clearly define the concepts under study, especially the mental states, and to determine the minimal degree of ToM sophistication needed to solve a ToM task. In the following sections, I will mostly review belief-based cognitive meta-representation, as it is what many researchers mean when they talk about ToM.

2) Mentalistic vs non-mentalistic interpretations

One of the main difficulties in the field is that studying ToM in non-verbal populations requires scientists to infer mental states, whatever definition of "mental state" we use, based on behavior. It is indeed impossible to clearly and explicitly communicate with infants or animals. Implicit ToM tasks, from which inferences about the possible meta-representations occurring in the mind of the participant are drawn, are therefore unavoidable. Consequently, implicit inference implies the existence of more alternative explanations. Combined with the lack of clear definitions for ToM and mental states, as I discussed previously, it becomes unclear what properties ToM abilities would and would not imply, and what clearly separates it from non-mentalistic behaviors. This is actually the center of debates in the field: to what extent is meta-representation needed to explain a success in a given task? Many have expressed the absolute necessity to ensure that the proposed experiments discriminate a behavior reader from a mind reader, i.e., that they differentiate between non-mentalistic accounts, that consider that the task can be solved purely based on the processing of behavioral cues, and the use of heuristics, and mentalistic interpretations that assume that meta-representation is used to solve the task [77], [82], [83], [91], [92], [94].

Krupenye et al. [88], in their study of implicit ToM in great apes, were met with alternative non-mentalistic interpretations of their results. One of them was that apes might submentalize, i.e., predict behavior via domain-general psychological processes, including several low-level cueing effects [94]. To control for this alternative, Heyes recommended to use inanimate controls, i.e., stimuli matched for low-level perceptual cues but devoid of social or agentic features, and to check the level of attention of participants. The submentalizing alternative predicts that domain-general processes will elicit common patterns of anticipatory looking, even with inanimate controls, and therefore levels of attention comparable to those of animate agentic ones. Kano et al. [100] tested this hypothesis and actually found no evidence that submentalizing could explain the previous findings from the False Belief test [88]. This suggests that apes were actually able to predict the behavior of an agent with a false belief, at the very least, based on a rich understanding of social cues. Nevertheless, some non-mentalistic alternatives might still be plausible. Apes could have "simply" tracked belief-like states, what Butterfill and Apperly described as the minimalist account [101], [102]. Alternatively, they could have relied on sophisticated rules that they learned during their lives, e.g., that agents are more likely to search for things where they last saw them. This is the behavior rule account [103].

The existence of mentalistic vs non-mentalistic accounts has led to the emergence of two camps within the scientific community that studies animal social cognition: the romantics vs the killjoys [104]. The romantics are inclined to attribute humanlike qualities, such as ToM, to other animals, while killjoys prefer more behaviorist, cognitively simpler, explanations. The past 10-20 years of ToM research in animals consists of a ping pong game between these 2

camps. It usually goes follows this pattern: research studies find some evidence that look like ToM thanks to new paradigms and/or tools [88]; they face skeptical views that suggest non-mentalistic alternatives [94]; they answer to some of the main critiques [83], [100]; other research studies find some evidence that also look like ToM thanks to newer paradigms and/or techniques [105], [106]; then once again are faced with skeptical critiques [107]; to which they also answer [93], etc…

Finding a reliable way to dissociate between non-mentalistic and mentalistic accounts is therefore one of the biggest challenges, if not the biggest one, that animal ToM research has been facing up to this day [77], [91]–[93]. Some have proposed paradigms to do so, e.g., the goggles test proposed by Heyes [91], [108], or the seeing-as paradigms of Lurz et al. [109]. However, these paradigms require complicated setups, which makes the odds of successfully running such paradigms with any animal small [92]. Coupled with the fact that many scholars emphasize that no single experiment will unequivocally prove the existence of meta-representational abilities in animals [110]. The difficulty to study ToM has grown to the point that some research teams might have become discouraged. Hence the decrease in number of research groups that actually conduct experimental research on ToM in animals [77], [111]. Notably, in ToM research in young infants, such a strong opposition between mentalistic and non-mentalistic accounts does not exist. Researchers who use implicit ToM tasks in infants however face different kinds of challenges, that are also shared with ToM researchers in animals.

3) A replicability and validity crisis?

Research in ToM has benefitted for the past 10-20 years from major technological advances. For example, advances in eye-tracking technologies, that are nowadays based on high-performance cameras and optimized tracking algorithms inspired from Machine Learning, enabled precise measurements of gaze direction and duration. Advances in biotechnologies, e.g. chemo- or optogenetics, or in electrical engineering (TMS, tDCS…) allow inhibition or excitation of more precise types of cells or locations, especially with chemo- or optogenetics. These new technologies, combined with new experimental setups, were used to develop new methods and protocols to measure ToM in non-verbal populations [62], [67], [93].

However, the problem with novelty in research is that it has yet to pass the validity and reliability tests. Validity concerns the extent to which the results really measure what they are supposed to measure. This includes the construct validity, i.e., how well a test measures the concept it was designed to evaluate. Notably, this is at the center of debate on ToM experiments in NHPs that I described in the previous subsection. This is mainly because of the existence of both mentalistic vs non-mentalistic accounts for the interpretation of such experiments. Similarly, according to a part of the scientific community, ToM research in infants also suffers from construct validity limitations. Indeed, some experiments that aimed to remove problematic confounds in the original studies, and at adding more suitable controls, did not replicate the initial findings, shedding doubt on implicit ToM in infants [112], [113].

A second objection raised by implicit ToM skeptics concerns the lack of convergent validity. Ideally, two tests that measure the same construct, here implicit ToM, should be highly correlated. In explicit ToM tasks, for example, tasks that measure the same meta-

representational structures but appear very different in their surface features are systematically convergent [36]. The high inter-correlation of performance in these different explicit ToM tasks, combined with the ontogenetical emergence in tandem of proficiency in the different tasks, corroborates the convergent validity of explicit ToM measures [114], [115]. However, some researchers remain doubtful of the existence of any systematic correlation 1) between different types of implicit tasks, 2) within different tasks of the same type, that try to measure implicit ToM [113], [116]–[120]. At the opposite end, other researchers remain convinced that implicit ToM evidence in infants is solid and sufficiently convergent, and that other factors account for failures to replicate seminal findings [121], [122].

Reliability measures the extent to which results are reproducible and replicable under the exact same conditions. Reproducibility is achieved when applying the same research methods to an already existing set of data leads to the same results. As for replicability, it requires to reconduct the entire research process again to produce new data but with the same methods and to find the same results as in the original study. This is one of the main challenges that implicit ToM experiments in children are currently facing. Many studies, published as well as unpublished, have failed to replicate the original results of implicit ToM tasks, especially those with Anticipatory Looking (AL) measures, but also violation-of-expectation (VoE), interaction, etc. [36]. Even though some of the studies have procedural differences that could explain the difference in findings [122], part of the research community remains unconvinced and calls for caution [95].

On the other hand, the application of implicit ToM paradigms to NHPs is still very recent (10 years or so), thus there has not been enough time yet to test for replicability. There are also inherent difficulties to replicate studies in NHP populations because of, e.g., few research groups, limited access to subjects, small sample sizes, etc. [123]. Therefore, many teams call for a high degree of caution regarding results of implicit ToM in NHPs [107], [123]. Notably, some researchers claim replicability in their experiments [83], [93], but there will need to be some time to check the replicability of implicit ToM tasks for NHPs.

4) Interfering intuitions and biases

One challenge that all researchers have to face in their experiments, more specifically in the interpretation of their observations, concerns their biases. Especially in ToM research, our own intuitions might be biased in favor of an interpretation with positive findings [92], [104], [124]–[126]. This bias might exist for many reasons. One of them is that media prefer stories of how animals are unexpectedly similar to humans, or of how infants are unexpectedly cleverer and adult-like than we might think. This creates an incentive for researchers to interpret their data in favor of ToM [126]. Furthermore, researchers are also humans, who also read and watch popular press and media. This might consequently influence and bias them towards this popular interpretation.

Another reason that incentivizes positive findings in ToM is that such findings might be considered as a corroboration of evolutionary theory [125]–[127]. Indeed, the presence of ToM in NHPs would constitute a convincing evidence that the cognitive complexity of humans could be the result of gradual natural selection. According to this point of view, NHPs with ToM would be a living illustration of an evolutionary stage through which the human mind might have gone million years ago. As for infants, the existence of an implicit ToM mechanism in

early life would also be evidence, arguably, of an evolutionary selection of abilities that strongly improve social cognition and would corroborate why humans developed such complex social abilities [121]. Such speculations are enticing and therefore quite popular with media, and hence may bias investigators interpretations of data.

Human ToM might itself bias researchers' reasoning. It is indeed designed to infer mental states—and later even mental states about mental states— from observable behavior [128]. Therefore, we might automatically think in terms of mental states when we observe behavior of other agents, because our own ToM, rather than objectively evaluating results concerning ToM in non-verbal populations, favors the interpretation that they possess similar ToM faculties [92]. They lead to biased interpretations that imply mechanisms that would be more cognitively demanding than necessary.

Such biases can be conscious or unconscious. Although researchers studying comparative ToM do their best to exclude as many alternative hypotheses as possible, relevant control conditions may also be overlooked [92]. For example, with respect to Krupenye et al. [88] and the Anticipatory-Looking False-Belief paradigm in great apes, Heyes [94] posited that apes might be submentalizing rather than mentalizing. They would use domain-general psychological processes, including several low-level cueing effects to predict behavior, rather than meta-representation. This would require a non-social, inanimate, control condition. Although their own follow-up experiments showed this hypothesis to be false—patterns of anticipatory looking were elicited in animate conditions but not in inanimate controls, which suggests that apes deeply understand social cues to predict the behavior of an agent with a false belief [100]—it was not initially discussed as a possible alternative explanation. This indicates that many other results may also come from insufficiently controlled experiments, as skeptics claim [82], [107], [125], [126], [129]. However, to be fair, controlling for all alternative explanations, especially in ToM research, is a hard task, whereas objecting because there were insufficient controls, after an experiment was run, might be useful for scientific advancement, but requires less effort, resources, or time.

Together, the criticisms listed above could contribute to the tendency to lean towards an interpretation that would rather fit the narrative of ToM in non-verbal populations [92] and justify the development of novel approaches to address these issues.

5) Ecological validity

A focus on the ecological validity of ToM experiments in animals emerged and gained traction at the beginning of the 2000s. As we saw in subsection A.3), such reservations participated in the questioning of the consensus at the time and in re-popularising this field of research. Ecological validity measures how generalizable experimental findings are to the real world. In the case of animal research, it also has to take into account the necessity to present to the animals an experimental setup that would be "understandable" to them, in the sense that, if a species is capable of ToM, whatever type of ToM or its level of sophistication, it evolved in a specific environment. Therefore, to investigate the limits of a species cognitive abilities, we have to frame the problems to be solved in such a way that the species might be familiar with tasks which should be plausible renditions of naturally occurring problems they might have to solve in their natural environment. In this view, experiments that would

undoubtedly need ToM abilities to be solved become even harder to develop, as those that have been proposed previously often suffer from low ecological validity [16], [130]–[132].

Another difficulty is that there is also a need for control over the experiment and its conditions, to limit all of the interfering intuitions and biases described in section B.4). Thus, a balanced compromise between ecological validity and experimental control must be found. Notably, too much ecological validity in an experiment might be tricky to interpret because a high ecological validity increases the odds that subjects are actually equipped with a built-in response that was favored by natural selection that does not require any cognitive abilities such as ToM [133], [134].

The use of existing experimental tasks is appealing and important because of their proven track record of validating a ToM ability in one or more species. This also allows, *en passant,* a direct empirical comparison between populations. We still have to keep in mind that they were designed with this specific group, a species of NHPs or an age group in human children, in mind. Consequently, such tasks, without relevant adaptation to the group under study, might prejudice the results, particularly if the context of the initial task happen to be less motivating or intuitive [83]. For instance, competitive contexts are often relevant for, and therefore applied to studies in chimpanzees, but they might be more stressing for bonobos and motivate them less [135]. A similar reasoning can be applied to infants and children. What motivates and appears more intuitive to a 1-year-old might not be so for a 3-year-old or a 5-year-old. Also, if success might be sufficient to prove some understanding of mental states, failure is actually not very informative about the ability to reason about mental states [109], [110]. This applies to any population, NHPs, human infants, children, or even adults. To better picture this fact, let's imagine any neurotypical human adult that would not consistently pass a ToM task designed for non-verbal populations. Would the conclusion be that they are not able to meta-represent mental states? Although this has never been tested, the answer to this question is obviously not.

To sum up, research in ToM is still facing many challenges, especially in non-verbal populations. To determine whether a population is able to meta-represent at least some mental states, and at which degree of sophistication they do so, it is essential to keep in mind all of the challenges mentioned in this section. Some of these challenges cannot be addressed easily with one unique study. Replicability will for example take a rather long time and quite some effort by several teams to reach a consensus. However, to design an experiment that tackles any aspect(s) of ToM, it is impossible to ignore the many challenges that must guide the design of new ToM experiments. At the very least the experiment must engage mentalistic cognitive processes, account for possible intuitive biases, be sufficiently controlled and also be adapted to the tested population.

C. How to study ToM?

The question now is to determine how to study meta-representation in any kind of population: verbal and non-verbal, human and non-human. To answer this interrogation, I will continue to review some literature on ToM. This time, I will especially focus on proposed solutions and recommendations to address the challenges described previously in section B. These will guide us to develop our own approach, that I will explain in more details in the second chapter of this dissertation.

In this section, I will first check the attempts to better define ToM: the diverse nature of the mental states, the degree of sophistication of meta-representational abilities, and the different cognitive processes possibly involved. In a second part, I will then discuss different levels of task immersion to better understand how the involvement and motivation of the participants can shape their behavior. Next, after introducing Social Decision-Making (SDM), I will see how the formal and theoretical framework of Game Theory can help us to study and understand ToM. Finally, in the last part of this section, I will explore how computational modeling can be applied within the SDM framework.

1) A multi-faceted ability

Before going any further, it is essential to take the time to highlight the multi-facetted aspects that ToM abilities can take. There is indeed a dire need to draw a coherent picture of what constitutes ToM before studying how humans and other species engage in it. This will require to define the different mental states that can be involved in ToM processes. As mentioned in subsection B.1), meta-representational abilities can also have multiple layers of sophistication. We will see in the second subsection how the depth of sophistication of ToM can be defined. My last subsection will be dedicated to the cognitive processes underlying ToM, which depend on both the nature of the mental states under study and the depth of sophistication involved in the task.

a- Diverse mental states

Originally, a conceptual distinction between affective ToM and cognitive ToM has been made to dissociate different mental states and social cognitive processes [2], [98], [136]–[139]. ToM for motivation, i.e. affective ToM, includes understanding of emotions, inferring a con-specific's emotional reactions based on typical situations that elicit certain emotions (Affective knowledge understanding [140]), or understanding that an individual may have different feelings than one's self about an event (Affective perspective taking [141]).

Understanding of desires is also part of affective ToM. Understanding that different individuals may have discrepant desires (Discrepant desires/Yummy-yucky task [60]), or that other individuals' emotions and actions are influenced by their preferences and desires (Desires task [142], [143]). In animal studies, affective ToM is often studied hand-in-hand with empathy [84], [144], but it is especially difficult to study in animals and young infants due to many challenges, as we saw previously. The impossibility to directly communicate with subjects, and therefore for them to communicate their emotional state, and personal biases such as anthropomorphism are some of these difficulties, among others [92].

ToM for knowledge, i.e. cognitive ToM, includes explicit perspective-taking and strategic reasoning about another's beliefs. For example, understanding that access to perceptual information (i.e., by looking, hearing, etc.) is a condition for access to knowledge (See-know task [145], [146]). Therefore, as we have already seen in previous sections, a false belief can be held by another individual (i) when a familiar container has an unexpected content (Content false belief paradigm [12], [13]), (ii) when a change of location or of a course of action that was not witnessed by the individual has occurred (location false belief) (Change-in-location paradigm/Sally-Ann task [9], [14]), or (iii) when something smells, sounds or looks like something else (Appearance-reality test [38]). Understanding that actions of another individual can be predicted based on their intentions is also an example of cognitive ToM (Attention to intention [147]). It is often cognitive ToM that is studied in animals: "Experimental studies of nonhuman primates often use observational and perspective-taking tasks under asymmetric environmental uncertainty" [98]. We therefore need to distinguish the different nature of ToM abilities that can fall in at least either one of two categories, affective or cognitive.

This has led to attempts to conduct thorough and detailed investigations about constitutive sub-processes of affective and cognitive ToM, especially in children, for example in Beaudoin et al.'s review [97]. In their work, they attempted to provide an exhaustive categorization of ToM abilities based on a systematic review of studies that measured ToM abilities in children. They divided these abilities into 7 categories, each of them being also subdivided, for a total of 39 sub-abilities (Figure 4). Their proposed AToMS (Abilities in Theory of Mind Space) framework classifies ToM tasks on the basis of what mental states they measure. However, it does not precisely define the mental states that constitute such categories. Instead the authors followed the description of the studies that they reviewed, and also placed each reviewed task under only one category that they judged to best reflect its measurement scope, although the study may have tapped more than one. Despite these weaknesses, their taxonomy reflects current literature's understanding of ToM and its components.

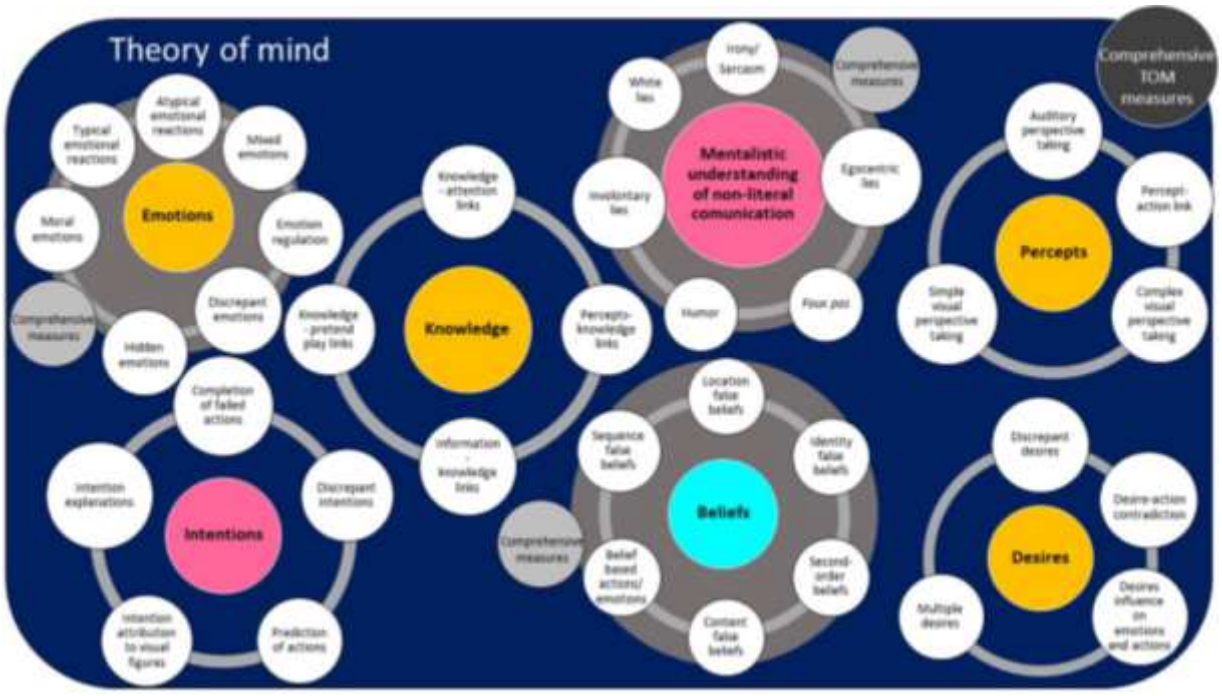


Figure 4 : ***AToMS framework****. The AToMS framework (Abilities in Theory of Mind Space) is a visual representation of the ToM categories and sub-abilities that emerge from the systematic review of ToM measures for young children. Theory of mind space is represented as a large area that includes seven ToM categories of mental states and social situations understanding (colored circles): Intentions, Desires, Emotions, Knowledge, Percepts, Beliefs, and mentalistic understanding of non-literal communication. Thirty-nine specific ToM sub-abilities (white circles) gravitate around the ToM category to which they pertain. Comprehensive measures that measure sets of abilities (multiple sub-abilities) for any one ToM categories, when they exist, are represented as gray circles. An eighth overall category "Comprehensive ToM measures" includes measures that encompass multiple ToM categories and is represented as a black circle. ToM categories (colored circles) are further represented using three different colors according to the proportion of reviewed studies that measured these types of ToM abilities: the pink circles represent ToM categories measured in <5% of studies, yellow circles represent ToM categories measured in 5–25% of studies, and the blue circle represent the only ToM category (Beliefs) measured in more than 25% of studies. From Beaudoin et al. (2020)* [97].

b- Levels of ToM sophistication

The need to define different levels of sophistication for meta-representational abilities arises very quickly when studying populations that differ greatly in cognitive abilities (for example infants vs primary schoolers vs adolescents) or living conditions (for example New World Monkeys vs Great Apes vs Humans). Some mental states might be easier to represent, some slightly more difficult, gradually up to the most difficult ones. For example, if we refer to the classification of ToM capacities and precursors that Lewis et al. (2022) [93] used in their review, others' goals might be the easiest to understand, others' perception slightly more difficult to represent, others' knowledge and ignorance moderately more so, etc., up to (false) beliefs which would be the most difficult to meta-represent in NHPs or infants. Furthermore, if the existence of implicit ToM abilities ever reached a consensus, arguably this could indicate a gradation in the acquisition and construction of meta-representational abilities, in terms of evolution as well as human development.

In humans, even after the stage that children pass FB tasks, their meta-representational abilities continue to develop and get richer with life experience up to adulthood [36], [148]. Therefore, the concept of second-order (and beyond) ToM, or

Advanced Theory of Mind (AToM), has been developed to describe the multiple aspects of an advanced understanding of the minds of others [44]–[47]. Higher-order false-belief understanding [11], [48], social understanding [49], perspective-taking abilities [50], and emotion- and mental-state recognition [51] are examples of AToM abilities. This set of capacities, not only requires first-order ToM abilities, in addition, improved general cognitive abilities, distinct constructs and concepts in their own right are needed for the development of AToM [52]. Such constructs and concepts include reasoning in recursion [148], [149], understanding of interpretation [150] (the same reality can have different interpretations depending on the individuals), or understanding of nonliteral speech such as irony, or jokes [35], [151]–[153].

Many researchers advocate the study of more ToM abilities under the lens of sophistication [92], [98], [99]. This is especially the case when studying Social Decision-Making (SDM) and recursive strategic thinking, which I will present in more details in later sections. For example, Barnby et al. (2023) [99] formalized social representations of how agents model beliefs with computational models (Figure 5). Their Depth-of-Mentalization (DoM) models have different levels of sophistication. For example, the sophisticated ones capture recursive, nested inferences of first, second, or k-th order, e.g. the self's model of the other's model of the self's model of the other. Importantly, they also provide a model for a shallow depth of mentalization that models the magnitude and uncertainty of a self's beliefs about the intentions of the general population. Such shallow DoM models could prove to be useful in characterizing non-verbal populations' mentalizing abilities.

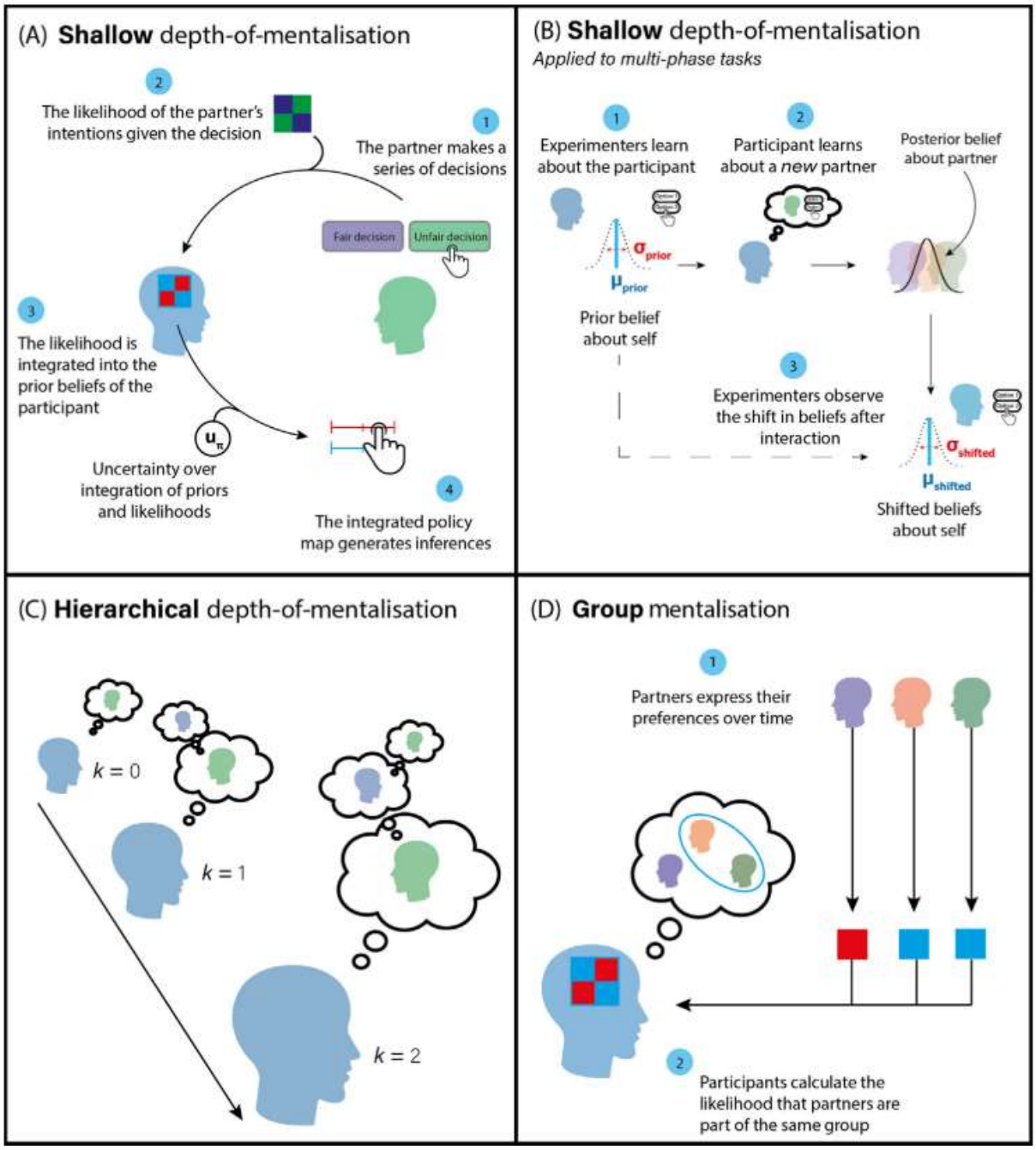


Figure 5 : ***The depth-of-mentalization (DoM) models with different recursive depths k****. (A) Schematic of shallow DoM models with recursive depth (k ≤ 0). In this example, the experimenters first learn about the magnitude and uncertainty of a self's beliefs about the intentions of the general population. The self then approximates the likelihood about the intentions of the other given the other's decisions. (B) Shallow DoM models can be used in inter-temporal tasks to assess changes to the representation of the self and other social agents. The experimenter's beliefs over the preferences of the self and the self's beliefs about the other can provide a basic scientific tool to test how the self and other are statistically represented, and what may perturb these representations. In this example, the experimenters first learn about the magnitude and uncertainty of a self's beliefs. The experimenters can then use the self's predictions about the other to build a possibly approximate model of the self's beliefs about the preferences of the other. (C) Hierarchical DoM models (k > 0) allow selves to make recursive inferences in their model of an other, and the model an other may hold about the self, up to level k (in one flavor, assuming that the other is level k-1; in another flavor, that the other is level -1…k-1 or 0…k-1, with a prior probability for each such level). (D) Group mentalization models are affiliated with the inferences and*

social behaviors of the group. This may be non-socially regarding (e.g., group classification based on an other's private preferences) or socially regarding (e.g., the probability of defection in a Public Goods Game based on the prior history of the group). From Barnby et al. (2023) [99].

c- Different cognitive processes underlying ToM abilities

From precursors to adult-level, ToM abilities have been described in both NHPs and humans. However, as we saw in the previous subsections C.1) a- & b-, there lacks a clear taxonomy of ToM abilities. Such a taxonomy would greatly benefit from a systematic decomposition of the processes responsible for these ToM abilities, because this would help clarify and dissociate the mental states at stake, as well as the level of sophistication needed. Identifying and separating distinct components of ToM in different behavioral studies is a challenge that would require us to choose on which criteria to base the generation of a list of more basic processes [96]. These components could be organized in a hierarchical scheme where the most sophisticated ToM constructs would be related to intermediate level components, which may themselves be further decomposed into more and more simple elemental ones (Figure 6). Examples of such elements would include (but not be limited to): perceptual discrimination and categorization of the socially relevant stimuli, as well as interoceptive signals elicited by those stimuli, semantic or conceptual knowledge, executive processes and motivational processes [96].

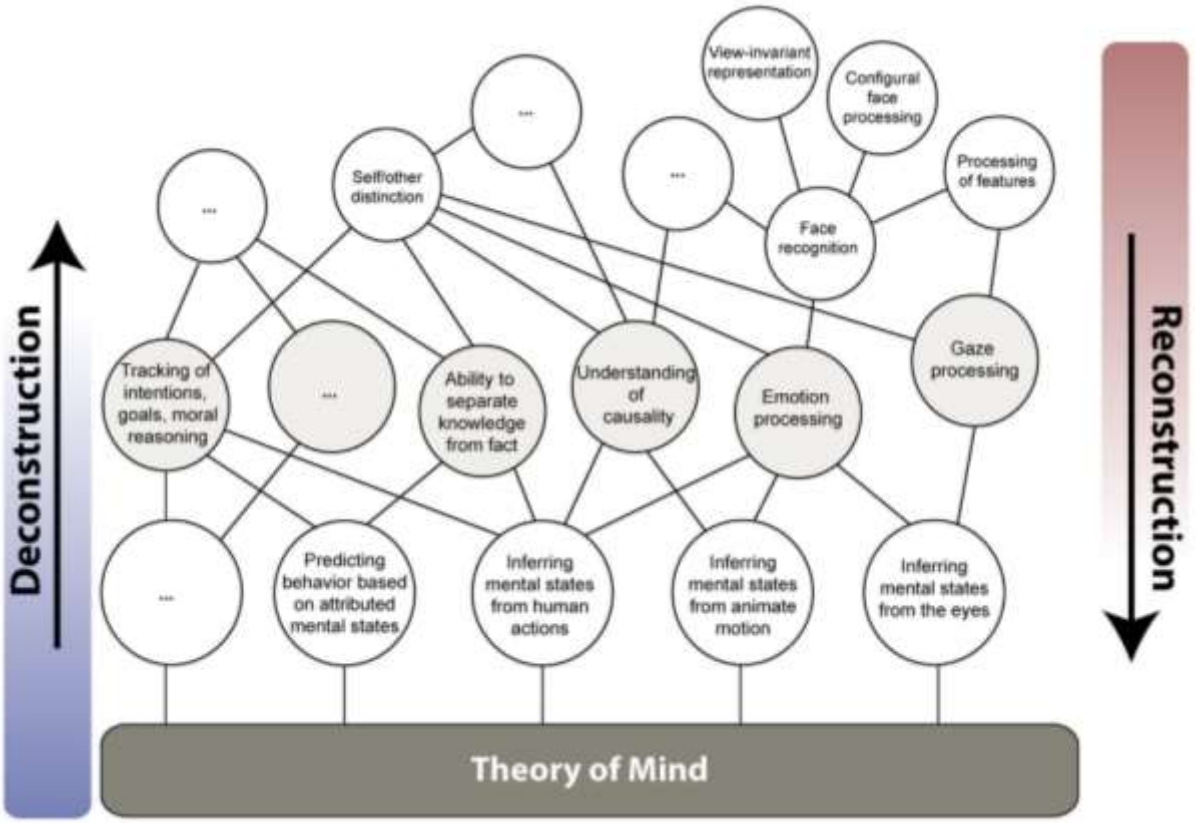


Figure 6 : ***Cognitive processes underlying ToM.*** *An illustrative example of the reformulation of ToM by deconstruction into a comprehensive set of basic component processes on the one hand; and a complementary reconstruction on the other hand, with the aim to construct a richer and scientifically tractable concept of ToM. From Schaafsma et al. (2015)* [96].

Another approach to clarify the elemental components and distinguish the core processes that are constitutive of meta-representation is to pool, through quantitative meta-analyses, the brain regions implicated in ToM tasks. Although imaging data are often purely correlational, and therefore need the support of causality-related data such as TMS and brain lesion studies [24], they can still provide a good overview of the common brain network underlying ToM processes [24], [154]. This network includes the temporo-parietal junction (TPJ), the medial prefrontal cortex (mPFC), the posterior superior temporal sulcus (pSTS), and parts of precuneus and posterior cingulate cortex. Interestingly, Schurz et al. (2014) [24] found that the brain areas that usually activate for ToM tasks vary in activity, both in strength and localization, depending on the ToM abilities required for the task. For example, "more dorsal/posterior areas in the TPJ (TPJp[…]) showed stronger activation for false belief vs. photo, and tendentially also for trait judgments, compared to other task. More anterior/ventral areas (TPJa and pMTG) showed stronger activations for social animations, mind in the eyes and rational actions compared to other tasks." This indicates a potential gradual functional specialization of brain areas involved in ToM, and more generally also for any cognitive process [155]. This is the "overarching view": each subregion mediates a different aspect of the more global cognitive ToM ability supported by the broader region. This corroborates the view that ToM abilities are multiple and differ between each other.

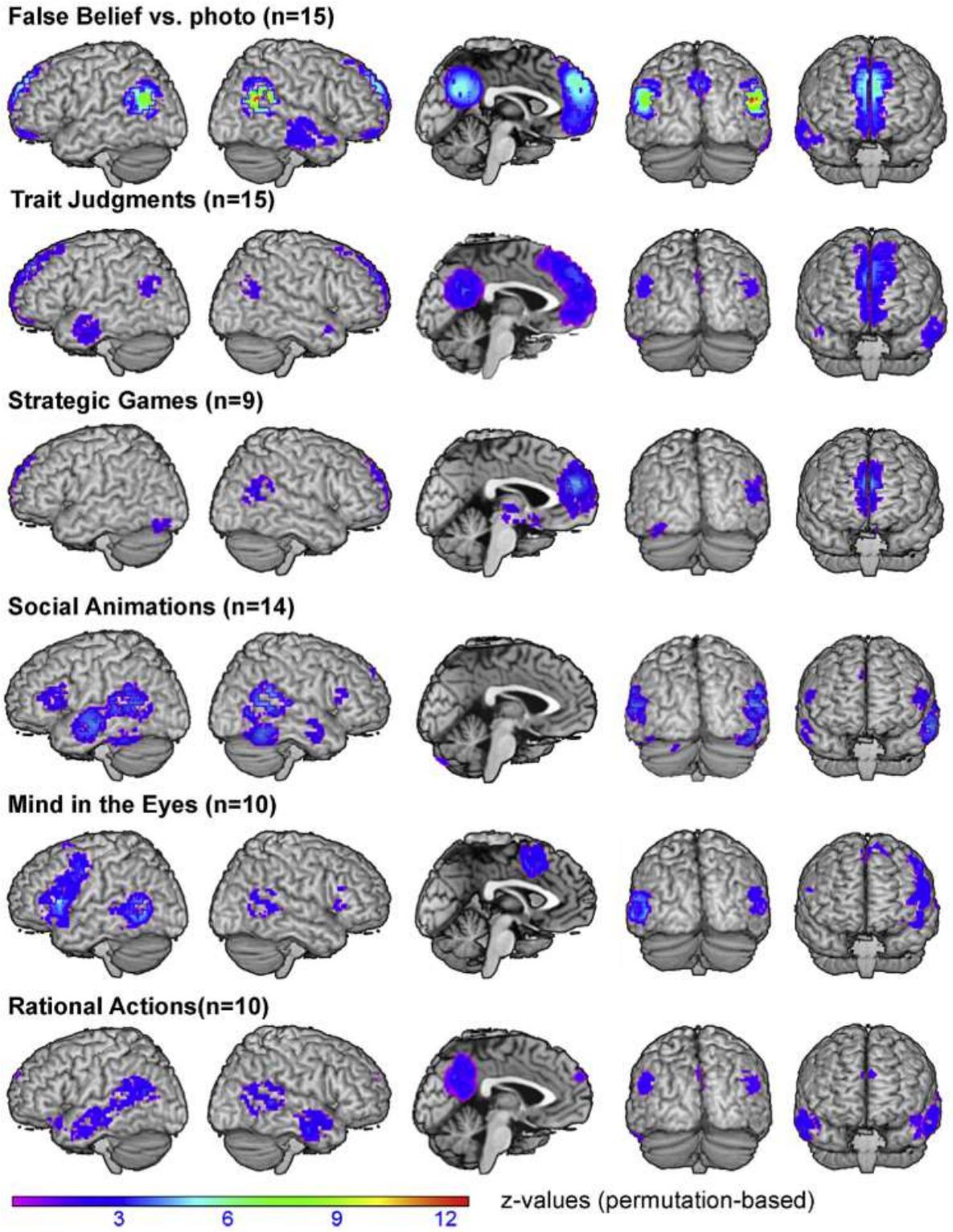


Figure 7: ***Results of separate meta-analyses for task groups.*** *Colors represent probability values from statistical permutation testing (z-values). Maps were thresholded at voxel-wise threshold of p< .005 uncorrected and a cluster extent threshold of 10 voxels. From Schurz et al. (2014)* [24].

In the end, clearly defining the diverse nature of ToM abilities and mental states, their sophistication level, and the cognitive processes underlying them will require a lot of time and effort from the broad neuroscience community. We are in dire need of theoretical models that provide clear definitions, assumptions and predictions with regard to the taxonomy of ToM abilities. Meta-analyses of the field that provide a wider view of the literature about ToM

abilities, sophistication, and cognitive processes will help to propose theoretical models, but in the meantime, single research teams that intend to study ToM in their experiments should be as accurate as they can with regards to the nature, sophistication level, and cognitive processes that they think might be at play in their ToM experiment to facilitate future meta-analyses on the subject.

2) Immersion and engagement in ToM tasks

To study ToM, an important factor to take into consideration is the motivation on the task, as well as its intuitiveness for the population under study. The studies that report positive findings of ToM abilities in NHPs allegedly, "capitalized on novel methodological tools in combination with highly engaging third-party social stimuli" [93]. The use of socially dynamic stimuli may be fundamental in order to elicit high engagement and reproducible results, even though this still needs to be directly tested [156].

In particular, for less studied species, taking the time and effort to adapt the task to each species is essential to prevent any experimental disadvantage for this species, for example less intuitive or motivating contexts. This will contribute to the ecological validity of the task. For instance, bonobos have received vastly less empirical attention with regard to ToM research than chimpanzees, although they are also one of the closest taxa to humans [93]. To capture the subtle differences between bonobos and chimpanzees, a combination of validated tasks that have provided positive evidence of chimpanzee ToM as well as novel tasks that are specifically designed to maximize motivation and ecological validity for bonobos will be required (Krupenye, MacLean, et al., 2017).

An idea to encourage the engagement and motivation of NHPs, but also of human infants, children, or even adults, would be to promote social rewards rather than only material (or financial) ones. For example, social information about dramatic interactions as third-party observers can serve as ecologically valid motivation. In animals living in groups, social events that can have major consequences in the group, such as agonistic or mating interactions, constitute highly motivating opportunities to incorporate ToM tasks [83]. In infants and children, praise and encouragement from an adult can motivate participants highly, and social stimuli of familiar contexts can help participants to engage in the task.

Crucially, direct social interactions that involve two or more individuals who act upon each other's state through their decisions might, under the right conditions, be better tailed for mental states meta-representation. Indeed, passively observing others' actions without any requirement for any response or judgement, nor any consequences for the observer, not only requires less engagement in the task. It also demands less ToM abilities and less self-referential processing than personally and directly interacting with another agent, with gains and losses dependent upon the decisions made by both [98]. Therefore, to study ToM abilities in NHPs or in humans, it is worth capitalizing on fields that exploit direct interactions between the participant and others rather than leaving the participants as simple observers of social scenes.

3) Social decision making and ToM

Social Decision-Making (SDM) fits the conditions described above very well. In SDM research a clear theoretical framework is defined to make predictions about interactions between rational social agents. It sets interesting hypotheses and properties that can be used to study ToM. In this subsection, I will introduce the reader to this field, by first defining the theoretical framework and assumptions often made in SDM. This will help us to then explain how ToM can be studied with Game Theory.

a- Theoretical framework & assumptions

Decision-making and its underlying process is studied in several fields such as psychology, cognitive neuroscience, behavioral economics, or informatics, in social as well as in non-social contexts. This term generally refers to the selection process for an option among other options that have different expected outcomes. Decision-making can be studied from the point of view of an individual or a group. It can be applied to unicellular living organisms as well as machines, and this field of study can be used to guide political decisions, inspire economic strategies, or influence individual preferences.

In the common British language, *Decision-making* is the process of reaching decisions, especially in a large organization or in government (Collins). It is the process of making choices, especially important choices (Cambridge). In psychology, decision-making is "the cognitive process of choosing between two or more alternatives, ranging from the relatively clear cut (e.g., ordering a meal at a restaurant) to the complex (e.g., selecting a mate). Psychologists have adopted two converging strategies to understand decision making: (a) statistical analysis of multiple decisions involving complex tasks and (b) experimental manipulation of simple decisions, looking at the elements that recur within these decisions." (American Psychological Association).

Decision-making requires a close examination of the probabilities and utilities that are derived from the different choices and that integrate various and multisensorial information sources (tactile, visual, auditory…), autonomous and emotional responses (e.g. heart rate, hunger, fear…), past outcomes, and aims for the future. Uncertainty, temporal constraints, and cost-benefits analysis must also be taken into account. Decision-makers often have to be fast, but flexible so that they can adapt to various situations e.g., social or non-social ones. This complex process can be divided into subprocesses (e.g., evaluation process, learning process, integration process…). Here I will make here a first assumption, that may be debatable, but necessary to start making inferences on observed data. This is similar to the process in mathematics in which axioms then lead to theorems. Decision-making becomes social, i.e., SDM, when two or more agents act upon each other's outcome through their decisions.

Just like in many fields, SDM can be studied under two complementary approaches. The (first) theoretical approach works under a formal framework and precise assumptions. For example, in Game Theory, decision-makers' choices are assumed to be guided by the maximization of a utility function. In practice, however, real behaviors often differ from theory, and the (second) experimental approach is therefore necessary to confirm or not the relevance of the theoretical approach. This will serve to gradually improve the theoretical

framework, that needs to be tested in a setup as close to reality as possible. In this chapter, I will describe the theory behind decision making.

Our first hypotheses are known as the von Neumann and Morgenstern axioms [157]. They consider that each individual is an economic agent, sometimes referred in literature as Homo Economicus, that is 1) perfectly informed, 2) infinitely sensitive, 3) rational.

1) Perfectly informed: The agent not only knows the whole set of actions available to them, but also their consequences. In some specific cases, we will not assume perfect knowledge of the consequences of the actions. This will indeed allow a learning process based on the past history of actions and outcomes.
2) Infinitely sensitive: The agent is able to discriminate between two alternative options that can be infinitely close. Their discriminative power over options is therefore a continuous function, that can be infinitely dividable. This hypothesis is necessary to have continuous and differentiable preference functions for all agents. In practice, an infinitely sensitive agent would be able to discriminate between all stimuli.
3) Rational: The agent is able to order by desirability the states in which they want to be, and will choose options based on the maximization of their criterion of choice (e.g. their well-being, personal reward…). A rational agent, when facing 2 options A and B, is able to attribute a respective desirability a and b to A and B, and to then distinguish the state in which they prefer to be. If a third option (C;c) is offered to them, the transitivity rule will apply, which means that if a > b and b > c, then a > c. Furthermore, a major principle of the hypothesis of the rational agent is that they will try to maximize a function, that is called "utility". This utility function, or expected utility function when there are risks or uncertainties associated with options, may depend on many criteria, or factors, among which option desirability.

b- Game Theory

Game theory studies mathematical models applied to strategic interactions among rational agents [158]. It has applications in a variety of fields, such as in social sciences, as well as in logic, systems science and computer science. Here, "game" is used as a broad concept that includes all situations in which rewards (e.g. money, food, sexual favors…) can be earned through relevant strategic choices. In Game Theory, a game is defined by 3 elements: the set of players of the game, the set of all choices and information available to the players, and the set of each players' utility functions that define the payoffs of each player for each outcome [159]. Players are supposed to be independent (if more than 1) and acting in their own interest. They know that their rewards will depend on their actions and the actions of the other players.

A strategy is a set of possible actions to choose during the game. A strategy can be pure, in the sense that a fixed action is associated to each state of the game and will be chosen deterministically if the player is in this state of the game. A strategy is called mixed if the strategy associates probabilistically a set of actions to perform to each state of the game. The utility function of each player associates each state with a value that represents the preference of the player to be in that state. Each strategy can then be associated with a computed (expected) utility by applying the corresponding utility function to the strategy. The output of the utility function is an objective, quantitative value.

i. Optimality and equilibria

Many studies have shown optimality for decision-making in humans during simple choices, e.g. during perceptual decisions with different options in a non-social context [160]. The optimum is here defined as the ratio $\frac{\text{Precision}}{\text{Reaction Time}}$. This definition of optimality makes sense in terms of natural selection where there is an evolutive advantage for and a selective pressure on individuals to develop abilities that would help to maximize this $\frac{\text{Precision}}{\text{Reaction Time}}$ ratio. But this is not the only way to define an optimal. For example, other studies that support the idea of a "Bayesian Brain" propose that humans make optimal or almost-optimal inferences. In that definition, optimality is viewed as the minimization of the frequentist measure of long-term losses. It is indeed possible to prove that a decision rule based on Bayes rule minimizes the sum of the errors on the long term.

These different interpretations of optimality are not incompatible; however, it is important to clearly define in which sense we consider optimality. In a lot of cases, we can observe behaviors that deviate from rationality or from any definition of optimality that we considered previously. This might be due to impulsivity, incorrect estimations of risks, or misunderstandings/misconceptions of probabilities. These cognitive biases are revealed through experimental approaches that show a deviation from the theoretical predictions. In a social context, predictions can become even more complex. But it is also in social interactions that the most obvious limits of the model of the Homo Economicus appear.

One of the most common lab studies that show how theoretical predictions differ from reality uses the ultimatum game. The principle of the game and of the experiment is quite simple and will involve one, and only one, interaction between two individuals. The first player receives money that they have to split between themselves and the other player. The other player can then either accept or refuse the money allocation. If they accept, the two players earn the amount of money allocated as decided by the first player. If the second player refuses, neither of them receives anything. In this setup, we regularly observe a refusal from the second player if they judge the offer unfair. This behavior is irrational from the perspective of the Homo Economicus model that would always accept, as it is the best and only choice to maximize their personal reward. They would indeed earn more by accepting any (non-zero) offer, as unfair as it would be, than by refusing. This prediction would not hold anymore if there were more than one, and only one, interaction between the two individuals.

Despite a potential divergence between theory and experimentation, the formal approach of Game Theory enables a classification of strategies, and sometimes even allows a comparison and ordering between them. In a social game, an economic agent from Game Theory tries to maximize their utility function by choosing rationally among possible strategies depending on the actions of the other players. As each player individually tries to do so, the optimal strategy for each of them depends on the strategy of the other players. Some strategies can therefore be better than other strategies, and it is sometimes possible to find a subset of strategies that are the best in certain given situations, and that we call solution.

Pareto-optimality

From the perspective of an external observer that is out of the game, each player adopts a strategy. This forms a strategy profile X1. If any player changes strategies, this forms

a new strategy profile X2. The strategy profiles can be ordered by the Pareto dominance relation. With this ordering relation, in the mathematical sense, we can define the notions of domination and non-domination of a profile X1 over another profile X2. Let's define the utility functions U: X -> U_i(X) as the utility functions of the multiple objectives constituting the game. The term objective can designate a specific player that has their own "objective", i.e. utility function. But it can more generally refer to any variable of interest to optimize over.

A strategy profile X1 dominates another strategy profile X2 if:

- the utilities of X1, i.e. U_i(X1), are no worse than those of X2 in all objectives i, so $U_i(\mathrm{X2}) \leq U_i(\mathrm{X1})$ for all i.
- at least one of the utilities of X1 is strictly better than one of those of X2, i.e. there exists j such that $U_j(\mathrm{X2}) < U_j(\mathrm{X1})$.

Strategy profiles X1 and X2 are non-dominating with respect to each other if:

- at least one of the utilities of X1 is strictly better than one of those of X2, i.e. there exists j such that $U_j(\mathrm{X2}) < U_j(\mathrm{X1})$.
- at least one of the utilities of X2 is strictly better than one of those of X1, i.e. there exists k such that $U_k(\mathrm{X1}) < U_k(\mathrm{X2})$.

The Pareto-optimality is reached for the set of strategy profiles that are either dominating or non-dominating with respect to any other strategy profile. Said differently, any strategy profile that is dominated by at least one other strategy profile is not Pareto-optimal.

Nash-Equilibrium

If we take the perspective of each player within the game, a Nash-equilibrium is reached when there is no gain for any player to deviate from their strategy, assuming that all the other ones do not change theirs. If the Nash-equilibrium is maintained through pure strategies, then the Nash-equilibrium is pure. If the Nash-equilibrium is maintained through mixed strategies, then the Nash-equilibrium is mixed.

Note that a Nash equilibrium is not necessarily Pareto-optimal, and vice-versa. Pareto-optimality can be seen as the overall most efficient situations that could be observable in a game from the outside. For the Nash equilibrium, it can only be reached by adopting the point of view of each player and trying to selfishly maximize one's utility without changing the strategies of the others. Let's take the example of the Prisoners Dilemma: 2 suspects are caught on the scene of a crime that they conjointly committed. However, there is no sufficient proof to condemn either of them, yet. However, there were already smaller charges from other previous cases against both of them. If both prisoners remain silent about their crime, they each get 1 year of sentence because of their provable previous felonies. If both confess and denounce the other, they both get 2 years. In the case where only one of them confesses against the other, he is set free while the other suspect gets 3 years of prison. In this dilemma, the Nash Equilibrium is the situation where both confess and get 2 years, but the Pareto optimal equilibrium will be the one where both don't confess both get 1-year sentences.

Example of application: the evolution of biological interactions

Another application of Game Theory is for example its use by evolutionary ecologists to better understand complex behavioral relationships between organisms in interaction to gain resources [161]. As we will see in more detail in the following section, repeated interactions can range in a spectrum from competitive to cooperative. Usually, interactions, whatever their nature, are costly, i.e. energy, or cost, has to be invested by each organism in interaction to derive benefits from it. The relative cost to benefits ratio of an interaction determines the payoff of this interaction. It can be a net gain if its value is positive, or a net loss if it is negative.

The most successful organisms maximize their payoff and increase their ability to reproduce by applying successful strategies over the successive trials. Evolutionary ecologists sometimes refer to strategies as phenotypes, and the best phenotype has the highest fitness. Natural selection will favor the fittest phenotype, i.e. the optimum strategy. Evolutionary consequences of iterative interactions can then be studied under the framework of Game Theory. Stable payoff equilibriums were found to explain cooperation and altruism (i.e., reducing one's own fitness to benefit the fitness of another organism or group of organisms), which otherwise might seem in contradiction with the more widespread, individualistic, competitive survival of the fittest in evolution [162], [163].

ii. Intentions

May it be in a dyadic interaction or interactions with an entire group, or even with a larger social network as we will see later, an individual has desires, goals and intentions behind their actions. I will focus more specifically on intentions during dyadic interactions in this subsection. More precisely, intentions in a strategic social game can be ranged in a continuum between pure cooperation and pure competition. By pure cooperation, I mean that the parties involved in a purely cooperative interaction share the same goals and interests. Coordination is therefore part of cooperation.

In purely competitive interactions, the goals of the players interacting are opposed or incompatible. As we will see later, game theory will allow us to adapt the rewards to create intermediate states between purely competitive and purely cooperative games. The mode of interaction in a game can therefore vary dynamically. The hypothesis is that feedback from a social partner – indicated by her moves in the game – is spontaneously used to infer her intentions, even if participants are not explicitly told to meta-represent these intentions [164].

Competitive intentions:

The American Psychology Association (APA) defines competition as: « any performance situation structured in such a way that success depends on performing better than others. Interpersonal competition involves individuals striving to outperform each other. Intergroup competition involves groups competing against other groups, with such groups including both small, face-to-face gatherings and very large groups, such as nations. Intragroup competition involves individuals within a group trying to best each other. Because competing individuals sometimes increase their chances of success by actively undermining others' performances, such goal structures can lead to conflict ».

Competition appears when resources are limited and more than one individual wants access to this resource. According to the APA, « in ecology, [competition for resources is] the use of the same resource by individuals of the same species (intraspecific competition) or of different species (interspecific competition) when the supply of the resource is insufficient for the combined needs of all individuals. It is a major factor in natural selection. Also called struggle for existence ». In nature, resources can be primary resources that are essential for survival and reproduction such as food, reproduction opportunities, shelter or territory. In social species, the resources can be secondary, for example prestige, money or power.

If resources are not shared consensually, individuals have to compete physically and/or socially for a part of these resources. Competition is therefore often generated from a conflict of interest in which victory of one side means defeat for the other. Defeat can also result for both sides, but I will not focus in detail on the cases of mutual defeat in this work. To achieve victory against an intelligent opponent, mentalizing their mental states can provide an advantage to predict their opponent's future actions and counter them.

Decision-making in the presence of competitive intelligent agents is fundamental for social and economic behavior [165], [166]. Competition is therefore often studied in conjunction with other social mechanisms that are closely tied to it such as ToM, social hierarchy, or cooperation. Competition can indeed also emerge from a comparison between one's own resources and another's, which might represent a difference in social status. This comparison with others can increase risky choice behavior even when there is no direct competition for the same resources [167].

Cooperative intentions:

At the opposite end of the spectrum of the existing modes of interactions between entities is cooperation, defined by the APA as a "process whereby two or more individuals work together toward the attainment of a mutual goal or complementary goals. This contrasts with competition, in which an individual's actions in working toward a goal lessen the likelihood of others achieving the same goal. In game theory, cooperation is regarded as the strategy that maximizes the rewards and minimizes the costs for all participants in the game; this is sometimes posited as an explanation for altruism. Often cooperation leads to outcomes, such as increased food, predator avoidance, or survival of kin, that make it adaptive (see adaptation), but the benefit to each individual is not always obvious."

In animals, it is more specifically a "behavior in which two or more nonhuman animals act together in a way that leads to mutual benefit. Examples include the cooperative nest building, food finding, and care of young in social insects, the mobbing behavior of some animals toward a predator, and the production of specific signals to indicate sources of food or shelter to others. It is not known whether cases of animal cooperation require that the animals understand the need to cooperate". Although competition is found in all living species due to their evolutionary need to survive and reproduce, cooperation is less frequent, especially cooperation based on mutual altruism between two strangers. Studies of NHPs often suggest cooperation, but the extent to which animals understand that individuals must act together to reach a common solution is still unclear.

A key idea is that players can achieve superior outcomes in the long run by working together rather than working against each other. A way to study cooperation formally, as well

as competition, is to use Game Theory to define the tools and concepts needed to characterize cooperative and competitive behaviors. Games are often represented as a payoff matrix (Figure 8). From this payoff matrix, strategies, i.e., successive choices, sometimes also called policies, can be derived. As we saw previously in subsection C.3)b-i., optimal strategies (policies) can be determined formally. In a dyadic social game, a cooperative dilemma is defined when the 2 players are more rewarded when they both choose to cooperate rather than compete (or defect), i.e., P > S (see Figure 8), but still have an incentive to compete (defect), e.g., R > P and/or S > Q and/or R > Q. In the Prisoner's Dilemma described previously, the payoff matrix satisfies the inequalities R > P > S > Q from player A's perspective (Q > P > S > Q from player B's).

(A)

Payoff matrix		Player B Strategy X1 (Cooperate)	Player B Strategy X2 (Compete)
Player A	Strategy X1 (Cooperate)	P (-1/-1)	Q (-3/0)
	Strategy X2 (Compete)	R (0/-3)	S (-2/-2)

(B)

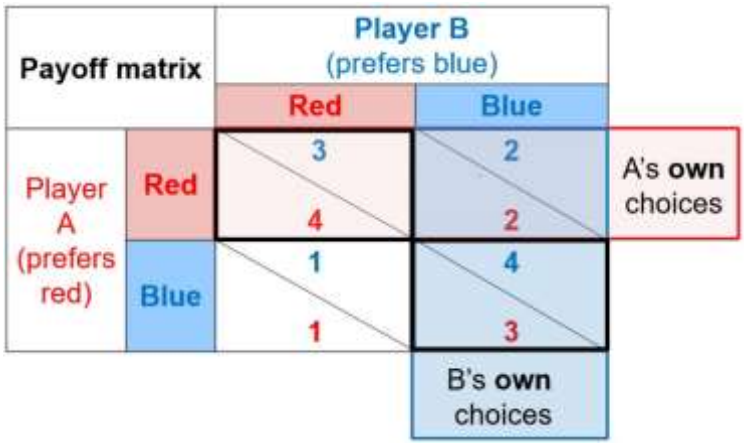


Figure 8: ***Examples of payoff matrices****. (A) General form of a payoff matrix. In parenthesis are the example values of the Prisoner's Dilemma described previously in section C.3) b-i. Both agents get 1 year of prison if they cooperate between each other by staying quiet. They both get 2 years if they both compete by defecting and accusing the other player. If only one of them stays quiet while the other on defects, the defector is set free while the other one gets 3 years of sentence. (B) Payoff matrix of the game "Battle of the Sexes". In this game, each agent has an individually "preferred" option, in the sense that they are more rewarded if they choose their "preferred" option. However, coordination on the same option by both agents is also rewarded with a bonus reward. The bonus is distributed equally to both players so that any coordinated choice is rationally better than no coordination for both agents. Still, one coordinated choice is better for the first agent and the other better for the second agent, which creates an inherent conflict about who profits the most. Here, if player A chooses "Red" and player B chooses "Blue", i.e. their own "preferred" choice, both get a payoff of 2. If player A chooses "Red" and player B chooses "Red", then they manage to coordinate and therefore get a bonus of 2. So, player A gets a payoff of 4 and player B gets a payoff of 3.*

4) Computational modeling: a powerful tool to characterize ToM

With a clear theoretical and mathematical framework to study ToM, it is now possible to make assumptions, predictions and simulations that we can compare to observed behaviors. Computational modeling constitutes the perfect tool to leverage calculation limitations and provide mechanistic explanations to mentalizing processes in the scope of strategic SDM. In this section, I will describe the main steps that I will apply in my projects to study strategic SDM and ToM by computational modeling. I will try to give the gist of the reasoning behind my approach to the reader without entering too much into technicalities (or at least as rarely as possible). Interested readers can refer to the review of Wilson & Collins (2019) [168] if they wish to obtain more details.

a- Definitions

An important first step is to define computational models. Computational models are, in their most abstract form, mathematical equations. The general goal of computational models is to provide explanatory hypotheses about how behavioral data were generated in the brain. In my own work, presented in this dissertation, these data will be behavioral data of final decisions made by groups of individuals. However, such data can also include reaction times, eye movements, or even recordings of brain activity. Computational models take these experimentally observable variables as inputs, usually past or current stimuli, outcomes or actions, to then produce outputs that usually correspond to predicted future actions or behavior [168].

We can distinguish at least two types of computational modeling approaches: correlational and algorithmic. Correlational computational models, such as linear or logistic regressions, provide broad qualitative descriptions of the data while algorithmic computational models, such as Q-learning [86], [169] or Bayesian Sequence Learning [170] models, make quantitative predictions and constitute algorithmic hypotheses about how data are generated (Breiman, 2019). Computational models can be used in different manners to answer different questions [172]. They include: generative simulation, parameter estimation, model comparison (to identify the “best” model), and latent variable inference [168].

Generative simulation consists of generating a set of data artificially by running the model with particular parameter settings. These simulated data can then be analyzed as if they were actual behavioral data. Qualitative and quantitative measures can then be used to make precise and falsifiable theoretical predictions about expected patterns in the real data. These simulations can also be used to validate the use of certain models or approaches, to ensure that the conclusions drawn are not wrong or misleading [173], [174], which is sometimes referred to as “sanity checks”.

Parameter estimation is used to find the best set of parameter values for a given model that reproduce the behavioral data the most closely. Also called “model fitting”, this account of a given data set can be summed up as a model and its best set of parameters. This summary can prove to be useful to investigate individual or group differences, or to quantify the effects of interventions, e.g., drugs, lesions or variations in experimental conditions.

Model comparison requires to define, fit and compute a set of possible models that could describe the behavioral dataset. The aim is then to determine which one among the

tested models best accounts for the data. The mechanisms that underlie the winning computational model are more likely to also underlie the actual process that produced the behavioral data. Similarly to the parameter estimation, model comparison can be used to investigate individual or group differences, or to quantify effects of interventions.

Latent variable inference involves computing for a given model the values of hidden variables that cannot be observed directly in a set of data. These latent hidden variables can then be used to draw correlations based on other sets of data. For example, expected rewards for different actions, which are hidden latent variables that are indirectly inferred, can be computed by using a Q-learning model on a dataset containing choice behaviors, which are directly observable behavioral data. These expected rewards can then be correlated with other sets of data, such as neural activity measured by functional magnetic resonance imaging (fMRI) [27]–[29], [175], electroencephalography (EEG), or electrophysiology and pupillometry, among many other data sources [176]–[180].

b- How to study behavior with computational modeling

Computational modeling can be applied to a variety of domains. In this dissertation, I want to apply it to the study of ToM abilities involved in strategic SDM. To do so, I will follow the general recommendations and process proposed by Wilson & Collins (2019) [168] in their review (Figure 9), and adapt it to my studies. More precisely, I will proceed following this general structure of reasoning:

1) Build the experimental design hand-in-hand with the computational models
2) Implement the experiment and collect the data
3) Calculate the general statistics and compute correlational models on the data
4) Design, test and select algorithmic computational models that are adapted to the experiment and the previous results
5) Explain behavior based on the selected model(s), the estimated parameters of the model(s), and/or inferred latent variables of the model(s)

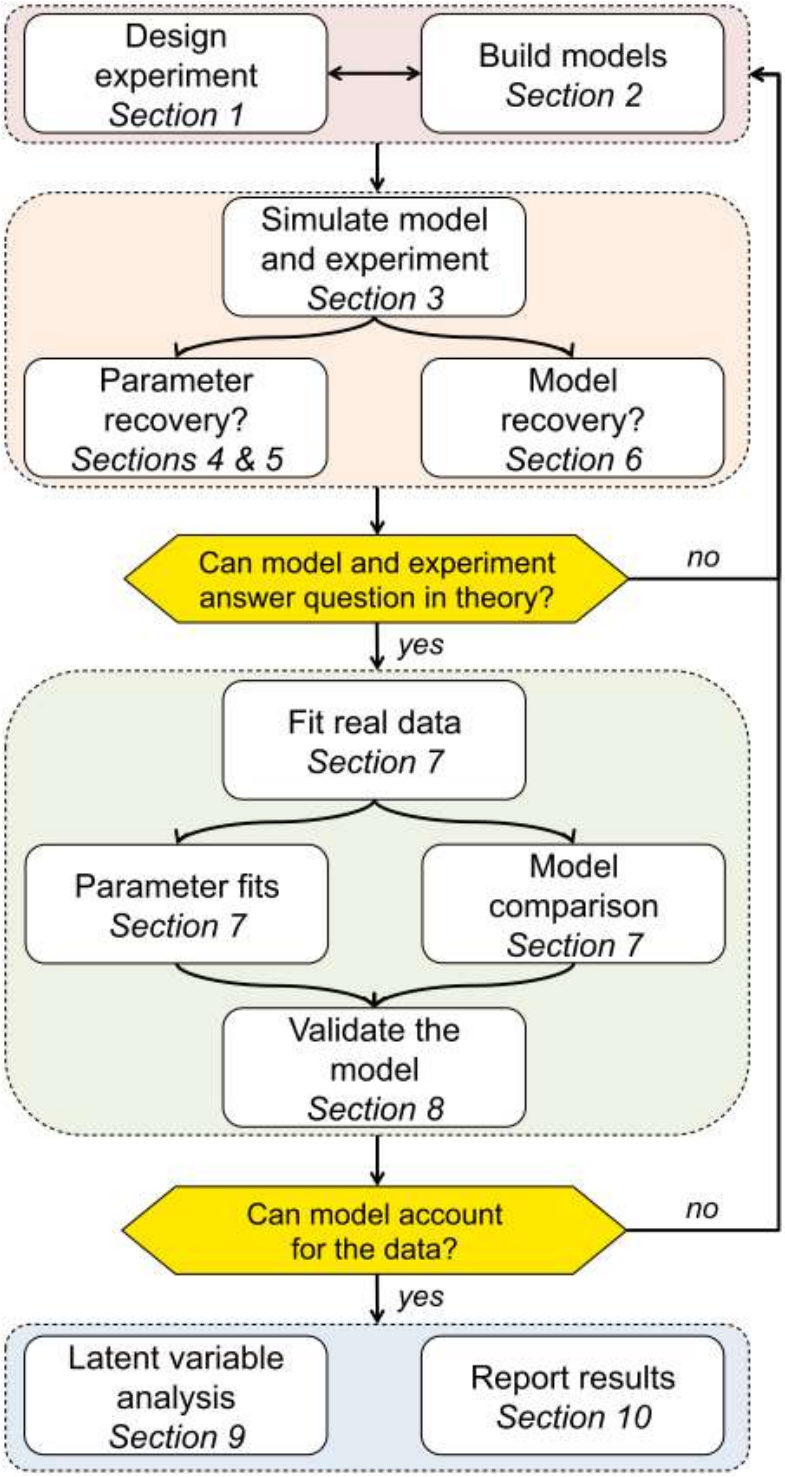


Figure 9: ***Schematic of the 10 rules and how they translate into a process for using computational modeling to better understand behavior****. From Wilson & Collins (2019)* [168].

i. The importance of a good experimental design

Computational modeling aims at capturing how information is used by agents behind the scenes, i.e., in their minds, for them to act in a certain way, i.e., their behavior. Consequently, the behavioral data fundamentally set limitations to the computational models, and the limitations of the data itself are crucially determined by the experimental design. This is why

a good experimental design is essential to fully unleash the potential of computational modeling approaches. To optimize the experimental design for computational modeling approaches, critical questions must always stay in mind.

What is the scientific question? This basic but primordial question must underlie and guide the experimental protocol to make sure that the goals are not lost from sight. What are the targeted cognitive processes? Mental states? What aspect of behavior should be captured? What are the hypotheses to be distinguished? Are there alternative hypotheses and can they also be distinguished? Also, does the experiment engage the targeted process? This difficult question requires expertise and/or piloting to be answered. If these questions are not answered beforehand, they will have to be answered during the review process. So, it is better to answer them (or at least be prepared to answer them) as soon as possible.

The next step is to actually implement the experiment and collect the data. With the many advances in technology, it is now possible to record a lot of diverse and precise data, sometimes remotely and/or automatically. Coupled with the great improvement of computational resources, this creates the perfect conditions for the advent of Machine Learning (ML) techniques to analyze a great amount of data. New ways to record data that are more practical, faster to collect, and that require less human intervention can nowadays be more easily deployed, however, such tempting technological tools are no excuse to rush the experimental design. Finding how these technologies can help answer immediate questions posed by the scientific community, in ways that were not feasible previously, will require discussing with researchers from various backgrounds to combine their respective expertise artfully.

In particular, experimentalists are of great help in designing an experimental protocol that can unveil specific behaviors. Specialists who consider any improvement possible to test their models also have an experimental thinking that grounds their ideas in reality and that help to generalize the application of their models to different situations [168]. Reciprocally, experimentalists that do not have any experience in computational modeling might seek the expertise of computational modelers to check whether the experimental protocol is properly designed to collect data to which computational approaches can be applied. Co-designing concrete experiment is therefore the best way to ensure that the experimental protocol is implementable, and that data collection and analysis will unfold smoothly.

ii. General statistics and correlational analyses

Ideally, behavioral signatures of the cognitive processes under study should be detectable from simple statistics or classical correlational analyses of the data. Indeed, experiments that clearly engage the intended processes also make them identifiable, to a certain extent, with basic, classical, behavioral analyses [174]. Detecting early signs of the computations under study from simple analyses of the behavioral data reinforces the relevance of the (algorithmic) computational modeling approach that aims at providing deeper insights that would remain unachievable with simpler approaches.

Simpler approaches include statistical approaches that can simply describe and summarize the data with descriptive statistics. This is also called exploratory data analysis (EDA) [181], and is useful to summarize the main characteristics of a data set and visualize it. It does not necessarily require any model. Statistical graphics and data visualization methods are at the

core of this approach. Plotting the maxima, minima, medians and quartiles, or any other relevant way to represent the data is usually the first step. Graphical techniques such as box plots, histograms, scatter plots, heat maps, bar charts, etc… are therefore crucial to show the data efficiently, and may help to discover new hypotheses about the causes that influenced data generation and guide the appropriate selection of the following statistical tools and techniques. The definition of model-independent measures that capture key aspects of the process(es) to be modeled is essential for further analyses. These constitute crucial qualitative signatures of the models.

Then, in order to determine more complex relationships between variables of interest, linear models in all of their forms (ANOVAs, ANCOVAs, linear and logistic regressions, generalized linear models…) provide ways to measure the correlational effects of regressors on a measure of interest. They allow us to explore more precise hypotheses and evaluate estimates such as inferential statistics, either with a frequentist or a Bayesian approach. Besides, they provide insights for (algorithmic) model specification, i.e., the process of building a model with an appropriate functional form and the relevant variables to be included.

Linear (correlational) models provide insights in the linear dependency between a predicted variable and predictors (see equation [1]). Depending on the nature of the predictors and the predicted variable, linear models can take different names. Assuming that we have a continuous predicted variable: when we have categorical predictors (also called independent variables), we use ANOVA (Analysis Of Variance), when we have continuous predictors (also called regressors), we use linear regressions, when we have a mix of categorical and continuous variables, we use ANCOVA (ANalysis of COVAriance), etc… (see Figure 10).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e \qquad [1]$$

Where:

- Y is the predicted (or dependent) variable.
- X_i are the predictors.
- β_i are the regression coefficients (β_0 is the intercept).
- $e \sim N(0, \sigma^2)$ is the deviation of the observations from the slope that follows a Gaussian distribution with 0 mean and variance σ^2.

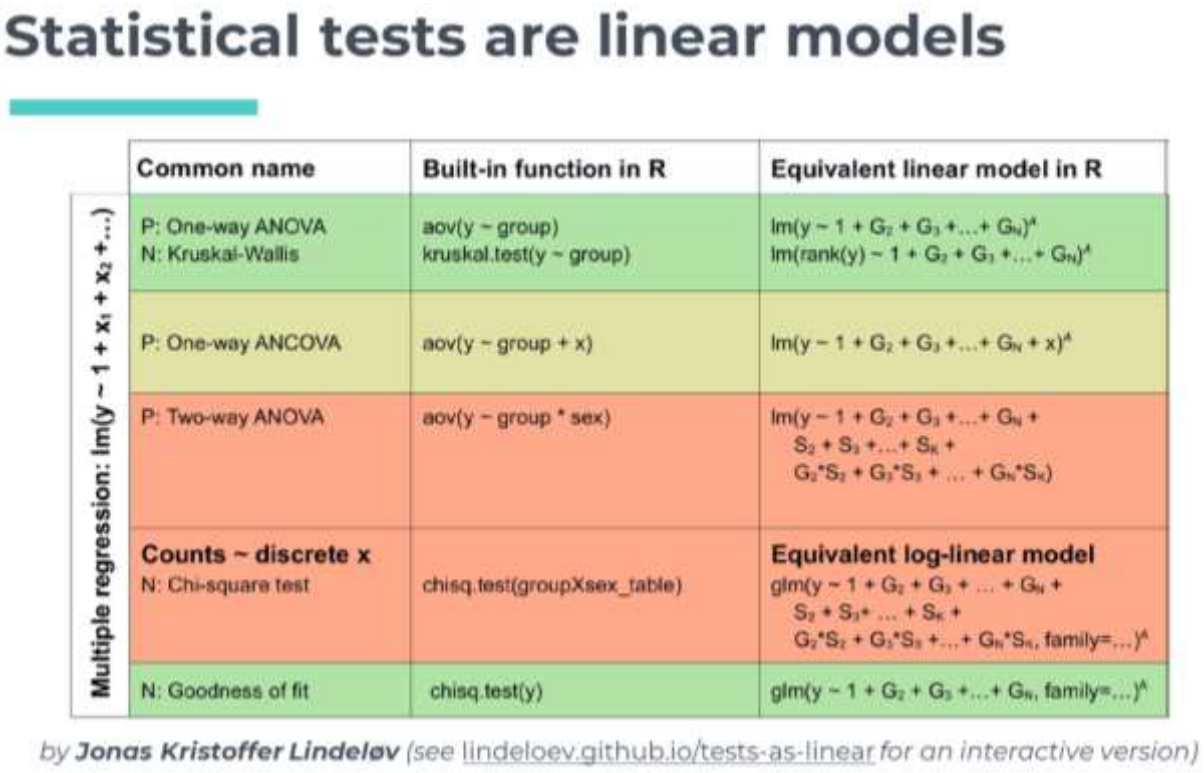


Figure 10: ***Statistical tests are linear models, part 1****. From O. Abdoun's online stat courses.*

Most statistical tests (e.g., t-tests, Pearson and Spearman correlation, Mann-Whittney U-test, etc… see Figure 11) are actually linear models. If the predicted variable is not continuous, we talk about generalized linear models. Logistic regressions are usually applied to binomial data, multinomial logistic regressions can be applied to categorical data, ordinal regressions are applied to ordered data, Poisson regressions are applied to count data (e.g., spikes), etc… All of these linear models have to be defined, then estimated to minimize the residuals *e* with model fitting. This is usually done with Ordinary Least Squares (OLS), which implies some assumptions about the data that must be checked. In order of importance: linearity between predictors and the dependent variable, independence and identical distribution (iid) of data points, homoskedasticity, no collinearity between predictors, normality of residuals. For more details about the consequences of a violation of these assumptions and corrections to perform on the data, see this course on statistics [182].

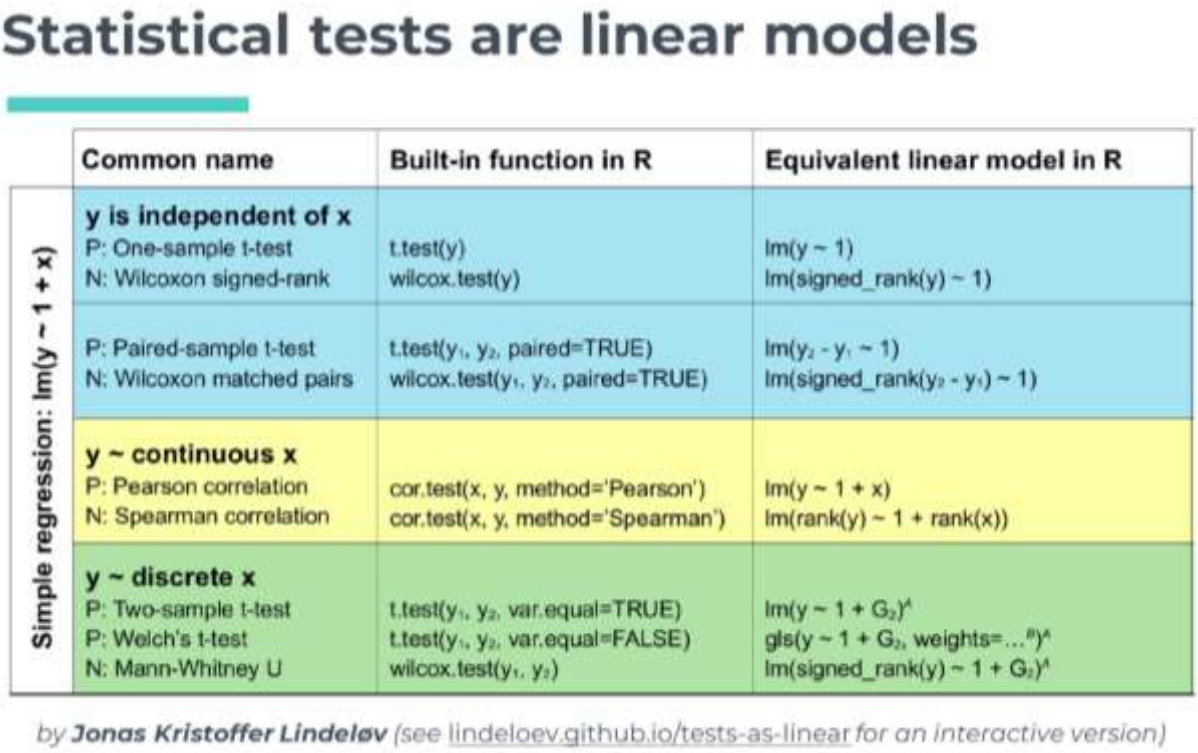


Figure 11: ***Statistical tests are linear models, part 2****. From O. Abdoun's online stat courses.*

Often times, data are not independent, in particular in repeated measures designs. The iid assumption is strongly violated in this case, because clusters of data points from each participant/animal/cell are sampled from a different distribution. To take into account the non-independence of data within clusters (participant, animal, etc.), linear mixed(-effects) models (LMM, or LME models) estimate intercepts and slopes for each cluster (i.e. subject), see Figure 12. They achieve this by taking advantage of the fact that subjects can be treated as a random sample drawn from a more general population. That means that it is sufficient to estimate the parameters of the underlying population (rather than of every single participant), and that it can be done using the entire dataset.

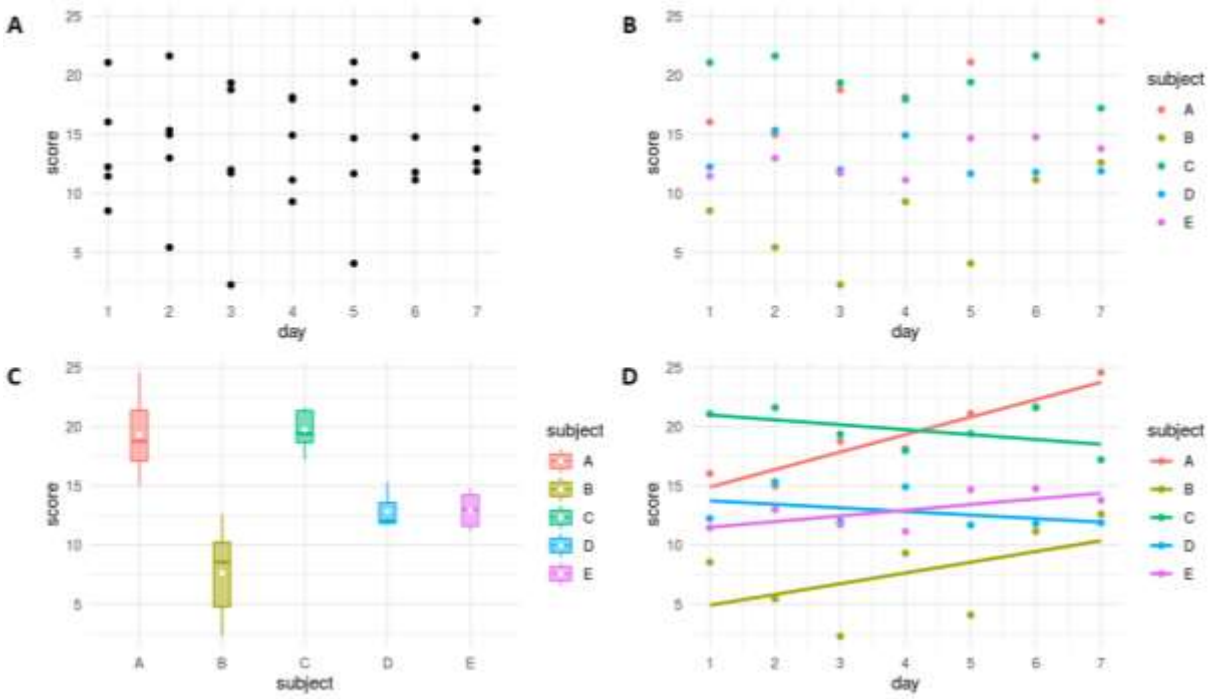


Figure 12: ***Illustration of Linear Mixed Effects Models.*** *(A) Data points represented without clustering them by subjects. A linear regression on these data would consider this scatter plot as the dataset. (B) Data points represented by clustering them by subjects. (C) Distributions for each subject. There are great variations in scores across Subjects. (D) Linear Mixed Effect Model applied to subjects as clusters with varying intercept and slope. From Abdoun's online stat course* [182].

iii. Designing good models

Based on the general statistics and correlational model analyses, it becomes possible to restrain further the set of candidates among all potential algorithmic computational models that could explain how data were produced. Indeed, correlational model analyses allow us to determine factors of influence that covary with the variables of interest. New models adapted to the current experimental setup and inspired by previous ones in the literature are easier to build in the light of the behavioral signatures detected by simpler approaches. Plausible computational mechanisms should indeed integrate and/or predict significant factors that were identified in previous step.

A clear understanding of the scale and end goal of the model is fundamental to designing it correctly. A mechanistic model that simulates high-level cognitive processes and that will be correlated to anatomical brain activity will differ in essence from one that simulates basic perceptual processes and be correlated with electrophysiological spike activity in single neurons. The wide variety of goals for computational models has led to a myriad of approaches that have been successfully applied in different contexts [172]. In this dissertation, I will mainly focus on learning models applied to decision making.

One of the simplest algorithmic computational models is the heuristic. This approach produces a response from the data based on basic behavioral rules. The Win-Stay/Lose-Switch (WSLS) heuristic [183] is an example of how a simple algorithmic rule can link behavioral information to the observed data. The reinforcement learning approach [172] inspired from Rescorla-Wagner's update rule [184] is another approach that has met with success in linking

human and animal behavior to brain functions/activity [87], [185]–[188]. Another perspective is that the brain performs Bayesian inference based on observations using probabilistic representations of the world, and utilizes the results of Bayesian inference to choose optimal actions [189]–[193]. Humans and non-human animals could therefore learn, and decide how to take action, based on Bayesian learning strategies.

The general approaches previously described can guide computational modelers who want to study ToM. Algorithmic computational models that employ meta-representational processes to solve strategic SDM problems can be inspired from the field of Reinforcement Learning (RL) [28], [29], [194], [195] as well as Bayesian learning [196]–[199]. They have frequently been applied to human studies, but, surprisingly, not so often to (non-human) animals, at least to my knowledge. One of the reasons why they have yet to be applied to animal studies is the lack of a convincing experimental setup that can fully exploit all the advantages of Game Theory tasks that might recruit ToM. Moreover, adequate controls that can limit alternative explanations would be needed and are difficult to conceptualize.

In any case, the interpretability of a computational model should never be traded off for other criteria, for example "goodness-of-fit" that we will see in more details later. Model components that are uninterpretable as a sensible processing of information but still improve the model fits indicate important missing pieces from the model, or a misunderstanding of the cognitive process at play altogether [168]. For example, a negative depth of reasoning makes no sense in the framework of recursive thinking in Game Theory.

iv. Explaining behavioral mechanisms with algorithmic models

After designing and selecting a first round of candidate models to test, these will have to be fitted to the data. Indeed, most models have internal parameters that modulate the predictions of the model, independently from the input data. The best set of parameters that produce an output "as close as possible to the real data" is calculated for each model through an optimization method. Multiple methods exist that take different approaches [200]–[208], with their assumptions, advantages, but also limits [209], [210]. Besides, the criterion (or criteria) that measures the "as close as possible to the real data" also needs to be defined. I will review some relevant information criteria in the next subsection (see section C.4) c-i.).

At this point, sanity checks will be required before proceeding any further. Model sanity checks ensure that any result given by a computational modeling approach are trustworthy [173], [174]. Each approach employed to test a hypothesis usually comes with assumptions that have to be verified and/or sanity checks. For example, model comparison calls for a model confusion analysis. Indeed, before selecting a "best" model, ensuring that the models are discriminable between each other is unavoidable. Similarly, a parameter recovery will be needed if any interpretation must be made from model parameters. To check absolute goodness-of-fit of a model, entire sets of data can be first generated by the model with the fit parameter values. Then, the generated data is analyzed with the same as previously described general statistics and correlational analyses. We will see later in a dedicated section the main principles behind these sanity checks (see section C.4) d-).

Assuming that these steps have been cleared, the fitted models that passed the model confusion analysis can now be compared between each other, to select the one(s) that best explains the data. In the model comparison, all of the models should embody reasonable

potential explanations of the behavior that capture different, graded, competing hypotheses. Different criteria exist to discriminate models between each other, each with their assumptions, advantages and limitations, too. Also, the model selection procedure itself can vary between researchers.

Usually, at this stage, computational modelers want to interpret the best selected model. What is the meaning of the computational mechanism behind? of the free parameters of the model? of the inferred hidden variables? Depending on the specific context and question of the study, the approach and interpretation will vary. Crucially, remember that "all models are wrong. But some are useful" [211]. Consequently, it is highly unlikely that the set of models that were investigated actually contains the "real" model that generated the behavioral data. Instead, it is worth considering that some mechanisms that intrinsically constitute the selected model are also most likely shared with the "real" model that produced the data. We will see examples of computational approaches applied to the study of ToM processes in a later section (see section C.4) e-).

c- Information Theory, Parameter Optimization and Model Selection

The use of computational models not only requires technical knowledge of the models themselves or their mechanistic implementation. A rigorous and thorough computational approach also deals with information theory, parameter optimization (aka model fitting), and comparison and selection methods. I want in this section to dive a little deeper into some technical aspects that I deem relevant to understand, so the reader may fully grasp the scope, but also limits, of computational approaches in general.

i. Information criteria

Any rigorous model optimization and/or selection method clearly states the evaluation criterion that measures the performance of the model. Many of them stem from the fields of Statistics and Information Theory. Among these criteria, some account for the complexity of a model by penalizing those that are unnecessarily complicated. Indeed, the more free parameters a model possesses, the easier it is to learn the data and potentially "overfit". Overfitting, in computational modeling, corresponds to learning the noise within a set of data rather than its generalizable properties. To avoid this, a correctly balanced compromise must be found between the complexity of the model, which includes the number of its parameters, and its goodness-of-fit to the observed data. Here is a non-exhaustive list of criteria with a brief description:

- Criteria that do not penalize model complexity:
  - Accuracy: the simplest and most basic criterion represents the percentage of correct estimations that the model made with regard to the observed data. It represents the number of times that the model's and the actual participant's choices were identical at a given time point. This is how most naïve people would evaluate the quality of a model, but it has some weaknesses that make it rather unfit for many cases.
  - Balanced accuracy: accuracy is known to be a poor indicator of performance when datasets are not perfectly balanced. Indeed, a bias within the dataset towards a

specific choice will favor models with an intrinsic bias in favor of the majority class. This will result in an overoptimistic accuracy for such models, which motivates the use of an alternative performance measure that removes this bias from performance evaluation. The balanced accuracy does so by independently calculating for each class of choice the accuracy of the model (Brodersen et al., 2013). For example, if only 2 choices are possible, red and blue, and if, in the data, a participant happens to play red 80% of the time, a "model" that plays red 100% of the time would get an 80% accuracy. However, if rather than measuring the accuracy of the "model" we calculate the balanced accuracy, it would get a 100% of the red choices, but 0% of the blue choices, resulting in an overall 50% balanced accuracy.

- R2: also called coefficient of determination, corresponds to the proportion of the variation in the dependent variable (the behavior that we want to predict) that is predictable from the independent variable(s) (the input data that are used by the model to predict said behavior). This statistic measures how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model (see equation [2]). It usually ranges from 0 to 1, where 1 is the best-case scenario. 0 corresponds to a model that would always predict the mean value of the observed data. A negative value means that the model is a worse predictor than always predicting the mean value of the data. The main drawback of this criterion is that it promotes the complexity of models because, by construction, it never decreases as variables are added and is very likely to increase due to chance alone. Therefore, modelers either need to adjust the R2, or simply use different criteria.
- (Log-)Likelihood: simply put, this represents the amount of information brought by the observed data about the model under study. It is calculated by the product of the probabilities at each time point that the model correctly estimates the actual behavior observed in the data (see equations [3], [4*J*]). The logarithm of the likelihood is often used for practical purposes, as it is more convenient to estimate its maximum with the log-likelihood. Indeed, because logarithms are strictly increasing functions, maximizing the likelihood is equivalent to maximizing the log-likelihood. Maximizing the (log-)likelihood is particularly useful to estimate parameters of probability distributions or models given some observed data. Indeed, if we assume that the form of the model as given and the set of observed data as fixed, the set of parameters that maximize the (log-)likelihood (for the fixed set of observed data) are also the set of parameters that give the best chance of producing the observed data with the given model. This is, on a side note, partly why this criterion is called "likelihood".

➢ Criteria that penalize model complexity:

- AIC: the Akaike information criterion (AIC) estimates the error of prediction of statistical models, and thereby the relative quality of models for a given set of data [212], [213]. For a given set of models fitted on the same given set of data, AIC estimates the quality of each model, relative to each of the other models. It therefore provides a means for model selection, as we will see later. AIC deals with

the trade-off between the goodness of fit of the model and the simplicity of the model by penalizing the maximized value of the log-likelihood with a term proportional to the number of estimated parameters in the model, which is a proxy for its complexity (see equation [5*J*]). The Akaike information criterion is named after the Japanese statistician Hirotsugu Akaike, who formulated it (Akaike, 1974).

- BIC: the Bayesian information criterion (BIC) (or Schwarz information criterion), developed by Gideon E. Schwarz [214], is also a criterion for model selection among a finite collection of models. Just like the AIC, it is based, in part, on the maximized log-likelihood, and, in the other part, on a term that penalizes model complexity (see equation [6]). The penalty term is usually larger in BIC than in AIC, especially for large sample sizes.
- Free Energy: free energy can refer both to a theoretical framework, the free energy principle, and to a quantitative measure in Information Theory [215]. Formulated by Friston and Stephan as a general principle that underlies interactions between the brain's internal model and its perception of the external world at first [216], this principle might generalize to interactions between any biological system and its environment [215]. It suggests that the brain reduces surprise or uncertainty by making predictions based on internal models and updating them using sensory input to enhance its prediction accuracy of the external world. The quantitative measure of free energy is also used in (Bayesian) model selection, as we will see later. Models with minimum free energy provide an accurate explanation of data under complexity constraints. Just like with AIC and BIC, the accuracy is derived from the log-likelihood of the model, and complexity is penalized with a term that, in this case, measures the effective degrees of freedom used to explain the data, called Kullback-Leibler divergence (see equation [7]).

ii. Parameter optimization

A key aspect of computational modeling is estimating the values of a model's parameters that best describe the behavioral data. Model parameter optimization, also called (model) parameter estimation (or fitting), essentially consists of finding a compromise between the accuracy of the estimation of the optimal parameters, and the time of computation. To do so, many approaches exist, that will depend on the chosen criteria to optimize (cf. previous subsection). The most intuitive approach is to optimize a criterion of choice by generating data from the model to optimize with multiple and diverse test sets of parameters, then to choose the set that obtained the best approximation of the data based on the criterion of choice. This can however be complex and fastidious to compute, especially if the model is probabilistic rather than deterministic.

If we instead take a Bayesian approach, our aim becomes to estimate either the probability that, given the model M and the data y, a set of parameters for this model θ_M produced the data, or the probability, given the model M and its parameters θ_M, to observe the set of data y. The quantities $p(\theta_M|y, M)$ and $p(y|\theta_M, M)$ can be linked to the log-likelihood (see equations [8][9]) which must then be maximized with regards to θ_M. All the challenge is how to approximate the set of parameters that maximize it, because an exact computation is

not usually tractable. To do so, different methods exist, each with different strengths and weaknesses.

This can be done by using sampling methods that simulate behavioral data based on the given model and parameters sampled from a distribution (e.g., a gaussian), such as Markov Chain Monte Carlo (MCMC) methods [217]. Otherwise, we can estimate the set of parameters that maximize either the log-likelihood with maximum likelihood estimate (MLE) methods [168], or the log-posterior probability distribution with maximum a posteriori (MAP) estimation methods [218]. Alternatively, variational Bayes approaches (VBA) [204], [219]–[221] make approximations of $\mathrm{p}(\theta_M|y,\mathrm{M})$ and $\mathrm{p}(y|\mathrm{M},\theta_M)$ indirectly by making key simplifying assumptions.

MCMC methods are a subset of Monte Carlo sampling methods that, as stated in the name, assume that the problem to be solved has a structure of a Markov Chain (also called Markovian Process). But what are Monte Carlo methods and Markov Chains? Monte Carlo methods are a technique that can be used to solve a mathematical or statistical problem by simulating repeated sampling to obtain the statistical properties of some phenomenon or behavior. They tend to follow a particular pattern: 1) Define a domain of possible inputs, 2) Generate inputs randomly from a probability distribution over the domain, 3) Perform a deterministic computation on the inputs, 4) Aggregate the results. As for Markov Chains, it can be informally summed up as "What happens next depends only on the state of affairs now." More formally, it is a stochastic model that describes a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

MCMC methods are a combination of the two (Markov Chains and Monte Carlo Samplings), in the sense that they first create samples from a continuous random variable, with probability density proportional to a known function. In practice, these samples are simulations of random walks in Markov Chains. This means that each sample starts from a set of points arbitrarily chosen and sufficiently distant from each other, and the following points of the sample are simulated with stochastic processes of "walkers" which move around randomly according to an algorithm that looks for states in the chain with higher probabilities. Then, these samples can be used to evaluate quantities, such as integrals over variables, e.g. expected value or variance. Note that this method is very versatile, so that it can actually be used for solving subproblems of the MLE, MAP or VBA approaches.

MLE methods are usually performed with variants of gradient descent, which is a method to find a minimum (or maximum, that is derived by taking the opposite of the function) of a mathematical function (here the likelihood). Its general principle is to derive, i.e. compute the gradient of, the function to optimize at an arbitrary starting point, then to iteratively move in direction of the lowest value of the gradient to find a local minimum. By testing many different starting points, a global minimum can be identified in the interval of study. Theoretically, a brute force grid search that consists of computing the value of the function to optimize at any point of the interval of study, with a small enough distance between each point, would also work. But this method is generally too costly computationally when the step is very small and the interval of study is very wide.

MAP methods improve parameter estimates by including prior information about parameter values. When combined with the (log-)likelihood, these priors can be used to compute the posterior, which in turn can be used to find the maximum *a posteriori* (MAP)

parameter values. With good priors, parameters estimated with MAP methods can be more accurate than MLE parameters (those estimated with maximum likelihood approaches) [222], [223], although this method suffers from its own problems when the priors are bad [224].

The downside of MLE and MAP models is that interesting information about uncertainty over the parameter distribution is lost. This is not the case with sampling approaches such as MCMC methods [168]. Furthermore, more complex assumptions can also be investigated with MCMC, because it allows rigorous data combination from different subjects, and provides a good approximation of the posterior distribution over each parameter value that can be reached with enough samples, despite small data sets [168]. However, this method is complex and slow to implement, while MAP, MLE, and also VBA as we will see, are much faster to compute and give results close to the MCMC methods for large enough data sets [168].

The last technique that I will briefly describe in this part is the variational Bayes approach (VBA). This method aims to optimize the evidence and posterior density indirectly by decomposing the log model evidence (equation [11]) and maximizing the free energy F(q) (equation [7]). This maximization is performed under two simplifying assumptions: the mean-field approximation and Laplace's approximation (see [219], [225] for more details). This renders q(theta) an approximate posterior density over the model's parameters and F(q) an approximate log model evidence (actually, a lower bound). Notably, MAP and MLE are particular cases of variational Bayes, also known as approximate inference by variational free-energy minimization, or ensemble learning [216]. Interestingly, this method can be applied for both Bayesian parameter optimization and Bayesian model selection [219]–[221], [225]–[227].

iii. Comparison and selection method

To select the best model from among various candidates based on a performance criterion, we proceed to Model Selection. Model selection can have two main goals. The first is to understand the underlying data-generating mechanism and interpret the nature of the data. The second is to be able to predict future or unseen observations. Usually, computational modelers want to do both. Our goal is therefore to figure out which model of the set of the tested models is most likely to have generated the data.

Naively, we could choose the model that performs best based on a given criterion by choosing the model that maximized said criterion. In the case of accuracy, balanced accuracy and R2, this would favor overly complex models, which would tend to overfit. Indeed, more complex models can adapt their shape to fit the data more easily, while the additional complexity, measured through the number of added parameters, may not represent anything meaningful. This is called overfitting. Hence, a good model selection technique will balance goodness-of-fit with simplicity. Selection based on criteria such as the AIC, BIC or free energy is therefore more relevant for our goal.

However, blindly choosing the model that minimized the AIC, BIC, or free energy will lead to more biased and statistically less accurate results than performing Bayesian Model Selection (BMS) to identify which model is best at the group (or population) level [220], [226], [227]. Assuming that the experimental data is the same, BMS also allows one to easily compare any set of models, including nested models (i.e., models that are included in other models).

Importantly, BMS can account for random effects in a population, which means that it considers that models might differ between subjects and that they actually have a fixed, unknown distribution in the population (i.e., it treats models as random effects). The challenge then becomes to estimate the frequency with which any model prevails in the population (see Figure 13.A). The notion of Exceedance Probability (EP) therefore becomes particularly crucial, as it measures how likely it is that any given model is more frequent than all other models in the comparison set: $EP_i = P(r_i > r_j|y)$, where j≠I [226]. This measure is more sophisticated and less biased than a simple ratio of the number of subjects that were best fit by each model [218].

An even more important advantage of BMS is that it makes it possible to compare the same group of subjects under different conditions, or different groups of subjects. It is particularly adapted for model comparisons within and between populations. Indeed, assessing between-group model comparison in terms of random effects is equivalent to asking whether model frequencies are the same or different between groups. Two hypotheses must thus be compared: either subjects' data y come from the same population, which means that model frequencies are the same for all subgroups ($H_=$, see Figure 13.B), or subjects' data y come from different population, which means that model frequencies are distinct between subgroups ($H_≠$, see Figure 13.C). By computing the (log-)evidences $p(y|H_=)$ and $p(y|H_≠)$, we can deduce the posterior probability that the two groups have the same model frequencies: $P(H_=|y) = \frac{1}{1+e^{\log p(y|H=) - \log p(y|H≠)}}$.

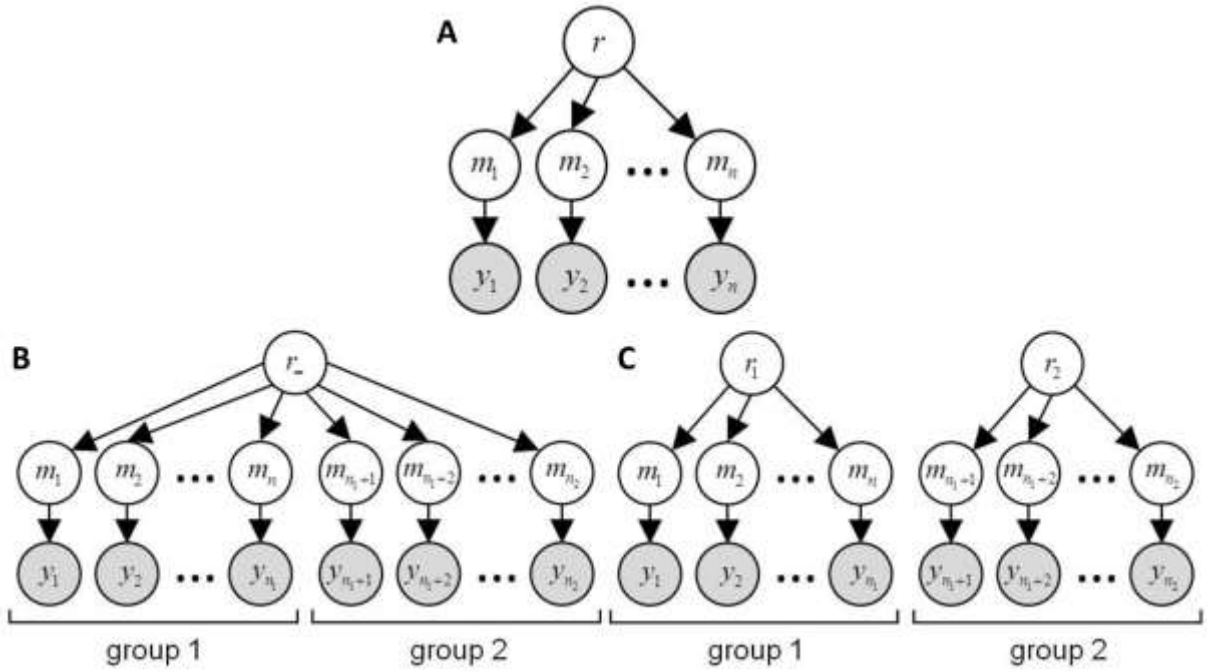


Figure 13: ***Random Effects Bayesian Model Selection (RFX-BMS).*** *(A) RFX generative model, where r is the population frequency profile, m is the subject-specific model label (it assigns each subject to a given model) and y are (within-subject) experimentally measured datasets. (B) Between-group model comparison with random effects, hypothesis $H_=$. (C) Between-group model comparison with random effects, hypothesis $H_≠$. Adapted from Rigoux et al (2014) and Stephan et al (2009)* [226], [227].

iv. Formula

This following subsection sets the formal definitions and equations discussed in the previous subsections.

$$R2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad [2]$$

Where:
- $SS_{res} = \sum_i (y_i - f_i)^2$ is the sum of squares of residuals (also called the residual sum of squares),
- $SS_{tot} = \sum_i (y_i - \bar{y})^2$ is the total sum of squares (proportional to the variance of the data),
- y_i are the observed values of the behavioral data set,
- $\bar{y} = \frac{1}{n}\sum_i y_i$ is the mean of the observed data.

$$L(\theta|x) \stackrel{\text{def}}{=} P(X = x|\theta) = P_\theta(X = x) \qquad [3]$$

$$L(\theta|x) \stackrel{\text{def}}{=} \theta \to f(\theta|x) \qquad [4]$$

Where:
- X is a discrete random variable with probability mass function p depending on the (possibly multivariate) parameter θ in the discrete case (equation [3]).
- $x \to f(x|\theta)$ is the probability density or mass function and x is the realization of the random variable X in the continuous case (equation [*4]*]).

$$AIC = 2k - 2\,ln(\hat{L}) \qquad [5]$$

Where:
- k is the number of estimated parameters in the model.
- $\hat{L} = P_{\hat{\theta}}(x|M)$ is the maximized value of the likelihood for the model M, on the set of observed data x, with $\hat{\theta}$ the parameter values that maximize the likelihood $L(\theta|x)$.

$$BIC = k\,ln(n) - 2\,ln(\hat{L}) \qquad [6]$$

Where:
- n is the number of data points in x, i.e. the sample size or number of observations

$$F(q) = E_q\left[ln\left(p(y|\Theta_M, M)\right)\right] - D_{KL}(q(\Theta_M); p(\Theta_M|M)) \qquad [7]$$

Where:
- $q(\theta)$ is any density over the model parameters,
- $D_{KL}(Q; P)$ is the Kullback-Leibler divergence (also called relative entropy), a type of statistical distance that measures how one probability distribution Q differs from a reference probability distribution P. It can be interpreted as the excess surprise (or Shannon information) caused by using distribution Q instead of P.
- $E_q[\,] = \begin{cases} \sum_i x_i q(x) \\ \int_{-\infty}^{+\infty} x\, q(x)\, dx \end{cases}$ is the expectation taken under q for discrete (top) or continuous (bottom) random variables.
- M is the model under study.
- Θ_M is the (possibly multivariate) parameter of model M.

$$p(\theta_M|y,M) = \frac{p(y|M,\theta_M)p(\theta_M|M)}{p(y|M)} \quad \text{(Bayes rule)} \quad [8]$$

$$log\, p(\theta_M|y,M) = log\, p(\theta_M|y,M) + log\, p(\theta_M|y,M) - log\, p(\theta_M|y,M) \text{ (log of [8])} \quad [9]$$

$$log\, p(\theta_M|y,M) = \sum_{t=1}^{T} log\, p(c_t|M,\theta_M,s_t,y_{1:t-1}) \quad [10]$$

Where:
- c_t is the choice at time t.
- s_t is the stimulus at time t.
- $y_{1:t-1}$ is the data.

$$log\, p(y|M) = F(q) + D_{KL}(q(\theta_M); p(\theta_M|y,M)) \quad [11]$$

d- Sanity checks

From a statistical perspective, given experimental data, model-based data analysis usually amounts to estimating unknown model parameters and/or comparing candidate models. However, such analyses are not always accurate. Therefore, in the same way that any experiment needs controls to make sure that the phenomenon under study is correctly characterized and measured, any computational analysis requires model sanity checks to make sure that its results are robust and consistent. To achieve this, numerical analyses can be performed to address the question of whether parameter estimation and/or model comparisons are indeed reliable, under a given set of experimental constraints.

i. Validation of the models and experimental protocol

A crucial step after designing the experiment and comparing the set of computational models is to validate them. To do so, we have to simulate the behavior of the tested models by creating fake artificial data, for example with MCMC methods, and to analyze them with the same methods as with the real data. This means the same EDA (exploratory data analysis) and correlational model analysis. The results should be very similar to those actually observed in the best model. A solid and rigorous way to validate and demonstrate the superiority of one model is indeed if that model can explain qualitative patterns in the data that are not captured by other models [168], [174]. Therefore, the model-independent measures that were previously defined to capture key aspects of the processes to be modeled should here be accounted for.

Simulating the models (especially the winning one) across a range of parameter values, to visualize the influence of the parameters on the behavior of the model, will help interpreting changes or differences between subjects or groups that are fitted by the same model, despite possibly significant differences in model parameters. The same is true for the simulation of the behavior of different models. Qualitative differences in the behaviors of the models allow them to be distinguished on the basis of qualitative patterns in the data. This however is not always possible [174], [228]. A quantitative method to ensure the distinguishability between different models is model confusion (or recovery) analysis. Similarly, for the estimation of model parameters, a parameter recovery quantitively measures the robustness and

consistency of the parameter estimation. I will briefly describe these two methods in the following subsections.

ii. Model confusion analysis

Can we arbitrate robustly between different models? This is the question that a model confusion (recovery) analysis tries to answer. To do so, all models can be simulated over a wide range of parameter values that covers the actual range that was observed in the observed data. The artificially generated data can then be analyzed in the same way as the actual data, i.e., fitted to all the models to find the model that best describes the artificially generated data. A robust and consistent model recovery would find that fake data generated from model A is best fit by model A as opposed to any other model. This process is generally summed up in a confusion matrix that quantifies the probability that each model fits the data generated from the other models best (see **Erreur ! Source du renvoi introuvable.** for an example). In a perfect world, the confusion matrix is the identity matrix. However, in practice, this is not always the case.

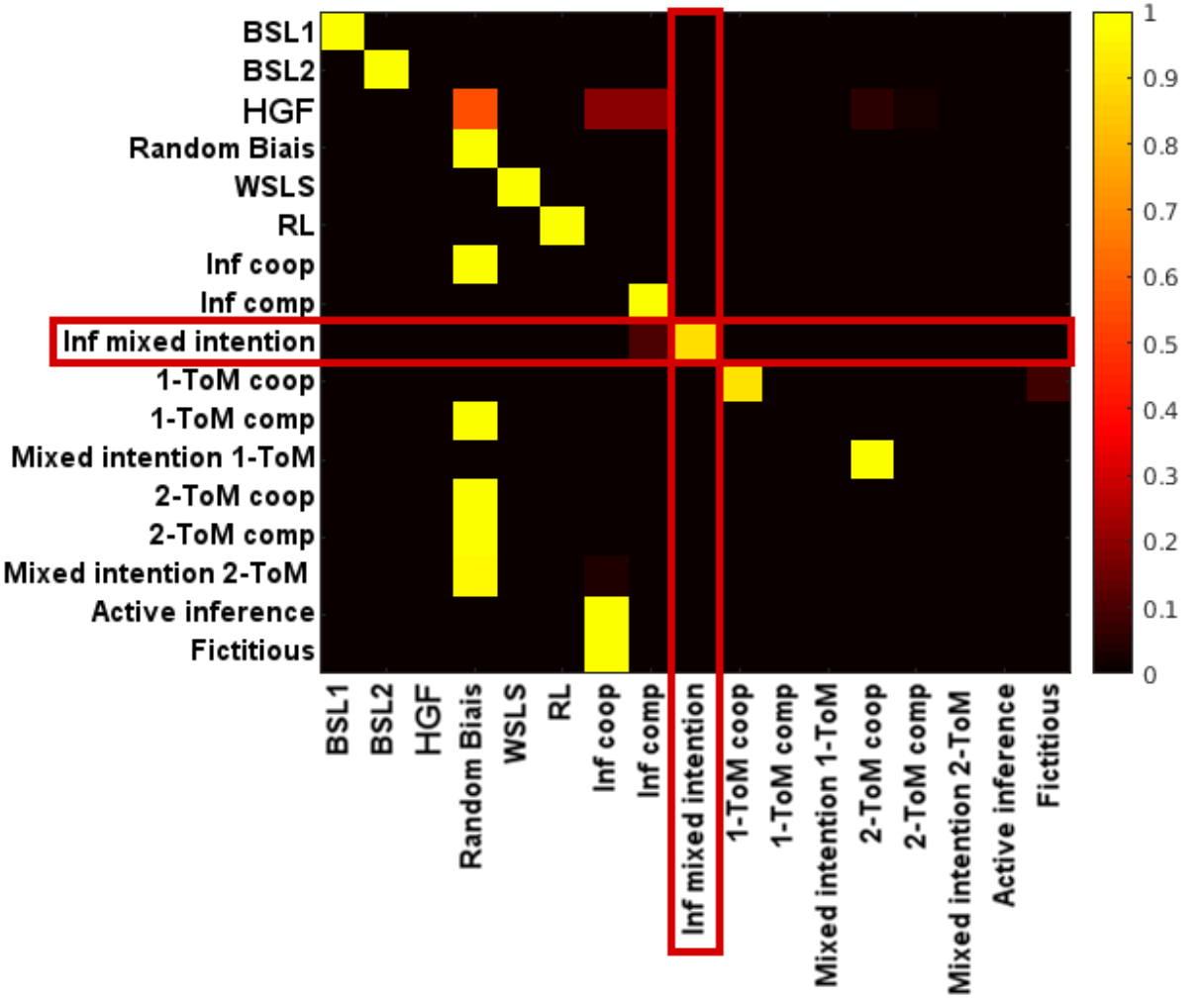


Figure 14: ***Example of a confusion matrix.*** *On the x-axis, we have the models that generated data. On the y-axis, the estimated model. The heatmap represents the Exceedance Probability (EP), i.e. the probability that a model explains the majority of the data compared to any other one. As one can see, the presented matrix is not diagonal. This means that some models are confused with other models. For example, the Random Bias model captures many other models, such as the Inf coop or 2-ToM models.*

Model confusion analysis can be performed by following the steps described hereafter:

```
1) for i=1:N (Monte-Carlo simulations)
    for sm=1:M [loop over simulated models]
      simulate data under model "sm"
      for cm=1:M [loop over candidate models]
        invert, i.e. fit, model "cm" on simulated data
      end
      perform bayesian model selection
    end
  end
```

2) confusion matrix = frequency with which each candidate model is selected (for each simulated model)

iii. Parameters recovery

Parameters recovery is absolutely necessary before interpreting any value of fitted parameters. It is indeed crucial to check whether the fitting procedure gives meaningful parameter values. To make sure that this is the case, we can fit data that were artificially generated by the model we want to test by parameter recovery and for which we know the "true" parameter values [229]. After fitting the model to the artificial data, a comparison between the "true" parameter values and the "artificial" parameter values will indicate to which extent they are correlated. Ideally, the simulated and recovered parameters should be tightly correlated, without any bias.

Critically, choosing a relevant range of values on which to perform model recovery is not always evident. If the data have already been fitted by the model, then matching the range of the parameters simulations to the range of values obtained by the "real" fit is better. A wider range is also possible. The next issue concerns how best to present the results. Plotting the correlation coefficient between simulated and recovered parameter values can summarize parameter recovery. It can then be depicted in a regression matrix, which would, in a perfect world, be diagonal. However, directly plotting the simulated against the recovered parameters makes the correlation clearer. It also reveals the range over which the correlation holds or not, as well as any existing bias, e.g., a tendency to recover higher or lower values on average. Notably, we might become interested in parameter differences between different populations or conditions. In such cases, parameter recovery can be informative to estimate statistical power. An interesting approach is to simulate and recover parameters for the groups or conditions and then perform statistical tests to detect group differences in this simulated data set. For example, the power of the effect size of the average difference in one parameter, between groups or conditions, is then the ratio at which the statistical tests detect an effect or not, knowing that a significant effect was actually found between groups or conditions in the data.

One way to perform a parameter recovery, sometimes also called simulation-recovery analysis, is to follow these steps:

1) for i=1:N (Monte-Carlo simulations)

sample model parameters under the prior distribution
simulate data given simulated parameters
invert model on simulated data and store estimated parameters
end

2) regress the estimated parameters on simulated parameters

e- Examples of computational modeling approaches to study ToM

i. Observation learning, ToM and computational modeling

Learning by observation might be one of the main mechanisms that could lead to the emergence of the ability to mentalize the mental states of others, i.e. the development of ToM. The behavior of other individuals is not viewed as simple movements anymore, but as intentional actions, planned and goal-directed, motivated by desires and guided by beliefs, that might sometimes be false. Some experiments, for example that of Castelli et al. (2000) [230] show that even in a non-social context, by observing dynamic schematics, humans can attribute beliefs and intentions to moving arrows. In their paper, Baker et al. (2017) [199] also use arrows that they described to the participants as representing actual individuals. In their study, they managed to simulate the participants' predictions concerning the arrows beliefs and desires. It is therefore possible to infer others' mental states (beliefs, desires, intentions...) by observing them (the other participants).

With that said, this raises the question about the choice between two learning strategies by observation: action imitation or new choice emulation by inferring their goals and intentions. Charpentier et al. (2020) [27] tried to answer this question with a bandit machine task in which participants alternated between observation and action. Volatility and uncertainty about the outcome of choice varied in their experiment. Uncertainty was linked to the probability that a token of a given color would come out of the machine. Volatility referred to the fact that the color of the token that would result in a reward could change, at a high frequency under high volatility and at a low frequency under low volatility. The participant did not know the color of the most rewarding token, but knew that the individual they observed did know it. Hence, the hypothesis was that under high volatility, i.e., when the color of the most valuable token changes frequently, the observed individual should be very volatile in their choice of token given that the observed individuals are fully informed.

Therefore, because the observer does not know the most valuable token, imitation is a safer option compared to emulating their own choice. On the contrary, if the behavior of the observed individual is more stable, but the bandit machines are more unpredictable, then the participant should emulate more their own choice. This was indeed the result that Charpentier et al. (2020) found with their arbitration model that simulated both the imitation and emulation processes simultaneously and then arbitrated between the two [27]. They not only found that the participants' behavior was well fitted by this model, but also that the brain regions that correlated in activation with the emulation process were also ones that have previously been linked to mentalizing tasks, whereas the imitation process was implemented by a network implicated in tasks involving mirror neuron systems.

ii. Identifying the nature of the computations behind strategic SDM (Social Decision-Making)

Traditionally, ToM is only viewed as an ability used to predict others' thoughts and actions by inferring their perceptions, beliefs, and desires. This would be equivalent to saying that "theory of cooking is useful mostly for guessing what is coming out of the kitchen" [231]. However, meta-representation is useful for more than passively predicting the future. It is also a pre-requisite for actively planning to change it. Thus, ToM can also be used for planning to change others' thoughts, feelings and actions by influencing their perceptions, beliefs, and desires [231]. If we come back to the analogy with cooking theory, this would equate to using this theory for planning and cooking a meal, rather than only guessing from the ingredients what is going to be cooked.

As stated previously, ToM in iterative social interactions is more engaging, in the sense that it implies direct concrete consequences for oneself, and the possibility to intervene on another's mental states with one's own actions. Consequently, planning one's actions carefully based on the context and the inferred mental state of others will engage ToM mechanisms. Of particular interest for computational modeling, "planning requires that Theory of Mind consists of abstract structured causal representations and supports efficient search and selection from innumerable possible actions" [231]. This will contrast with less cognitively demanding alternative computation mechanisms, such as simpler statistical

predictive models of other people's actions, or model-free reinforcement of actions according to their effects on other people (see Figure 15).

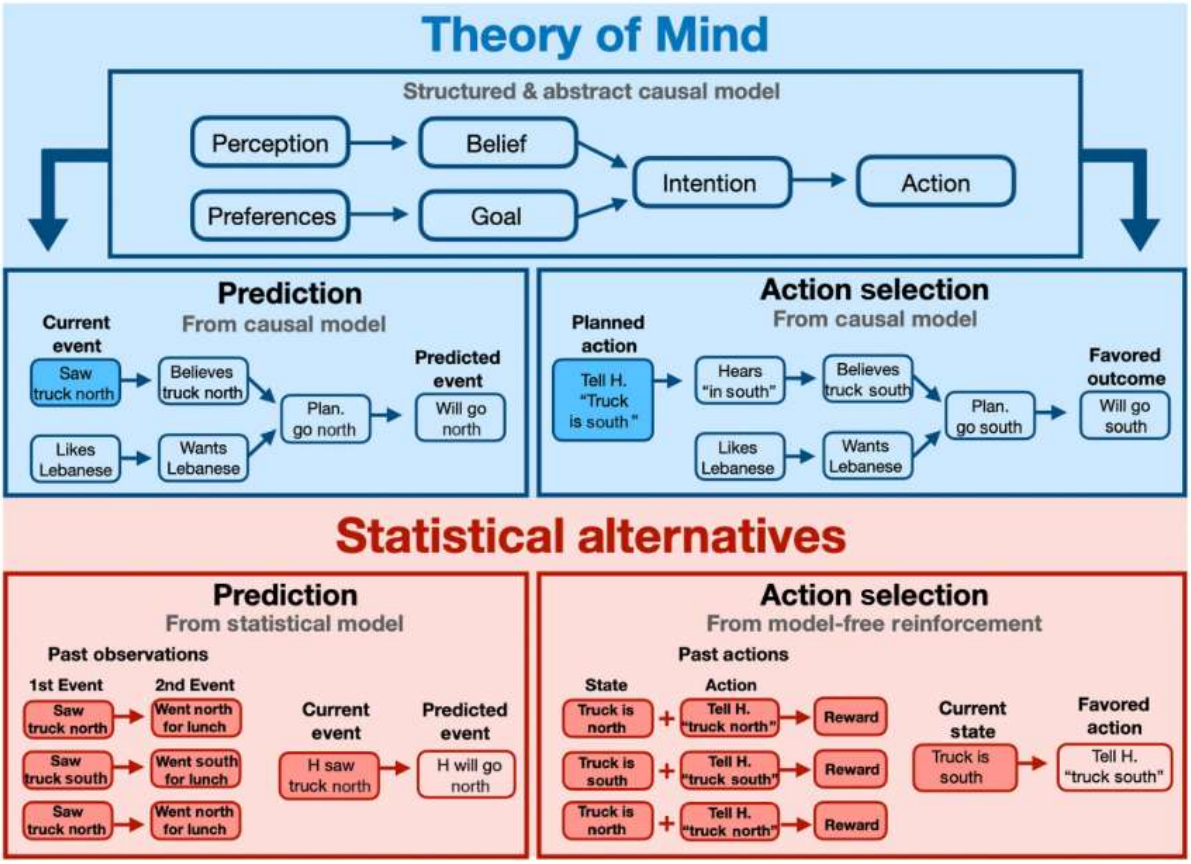


Figure 15: ***Contrasting Theory of Mind with statistical alternatives****. Context: "The classic problems (and psychological tasks) used to study Theory of Mind require an observer to predict or explain another person's actions. For example, consider Harold, who typically orders lunch from a Lebanese food truck. Earlier this morning, Harold saw the food truck in the north lot, but since then it has moved to the south lot. Where will Harold go at lunch time? In this variant of the standard false belief task, Harold's friend Grace can predict that Harold will go to the north lot (and will be disappointed when he gets there). Using Theory of Mind, Grace could predict Harold's actions by invoking how Harold's perception and inference (he saw the Lebanese truck in the north lot, he infers it has not moved) cause his belief (the Lebanese truck is in the north lot), which combines with his desire (to get Lebanese food for lunch) to create a plan (go to the north lot)"* [231]*. Top: Theory of Mind is an abstract causal model, specifying how mental states like perceptions, beliefs, and desires combine to cause actions and feelings and so can be used for both prediction (e.g., given the target's inferred beliefs and desires, predicting their actions) and action selection via planning (e.g., given a desired action, selecting the best intervention on beliefs and desires). Here we illustrate how Theory of Mind could be used (left) to predict a person's actions, if they have a false belief (here, that a food truck is in the north), and (right) to intervene to cause the true belief (the truck is in the south). Below: statistical models generalize prediction and action selection from prior experience of similar states or sequences, without building a causal model. A non-causal predictive model cannot be used for action selection and a model-free action selection mechanism cannot be used for predicting events. From Ho et al. (2022)* [231].

For the reasons discussed above, the framework of SDM is particularly well suited to study and characterize ToM abilities . Based on this line of reasoning, Philippe et al. (in press) recently characterized computationally the ability to adapt to unsignaled changes of cooperative vs competitive intentions in healthy human adults in an SDM task. The game consisted of choosing one card from two (of different colors), at the same time as another

player, whose current choice was not visible, and being rewarded if the colors were the same. Feedback was given after each trial.

The manipulation was that they played the game believing that they were interacting with another human player, but were actually playing an artificial agent (AA) that secretly and regularly switched in blocks between two modes: a competitive mode vs a cooperative mode. In the cooperative mode, the AA was trying to maximize its probability to coordinate with the player. In the competitive mode, its aim was to avoid coordination as much as possible. The unsignaled changes of intentions of the AA required the participant to adapts to them if they want to coordinate and maximize their reward.

They tested computational models of different natures in the sense that the models could be classified in families based on the kind of mechanistic computations that underlie them. The models could belong to either non-Bayesian vs Bayesian families, and to non-mentalizing vs mentalizing families. To test for a dynamic tracking of implicit intention, different subfamilies of models were tested. Bayesian mentalizing models included k-ToM models [85], [197], [232]. Non-Bayesian (reinforcement learning-based) mentalizing models included influence models [28], [29]. Bayesian non-mentalizing models included Bayesian Sequence Learning (BSL) models [170]. And finally, non-Bayesian non-mentalizing models included a Q-learning model as well as a Win-Stay/Lose-Switch (WSLS) heuristic. For a more detailed description of each of these classes of model, see subsection A.4)e-iv.

Importantly, they constructed an adaptation for each of the mentalizing models (influence and k-ToM) in which an arbitrator weights the cooperative and competitive versions according to their reliability before making the decision. This construction was partly inspired by the imitation/emulation model architecture of Charpentier et al. [27]. This led to the implementation of the Mixed-Intentions Influence Learning (MIIL) model which arbitrates between two influence learning models with different beliefs as to the intentions (competitive vs cooperative) of the other player. Crucially, this model captured fluctuations between cooperation and competition during social interactions.

They then compared the explanatory power of the different models on the data through a BMS based on the free energy as criterion. They found that the model that best described their data on their task was the MIIL model. Interestingly, they also correlated the BOLD signals with the hidden states of the winning model. The ventromedial parieto-frontal cortex (vmPFC) and striatum likely encoded the reliability difference of the controller of the MIIL model whereas the dorsolateral parieto-frontal cortex (dlPFC) and tempo-parietal junction (TPJ) likely encoded prediction errors (PE).

Their work thus characterized specific aspects of ToM. The mentalizing term of the influence learning model that constitutes the MIIL model takes into account the influence of one's own strategy on the strategy of the other player, and the arbitration process between strategies, depending on the reliability of each expert, constituted a meta-representation of the intentions of others.

iii. ToM sophistication and computational modeling for groups or conditions

Determining the nature of the ToM computations underlying decisions in strategic SDM tasks is only one aspect that computational modeling approaches enable. Such an

approach can also be used to explore the difference in ToM sophistication and/or in the nature of the computational process that lays behind decisions of different groups of subjects, or, of the same group in different conditions. For example, Devaine et al. (2014) [197] examined the difference in the computational processes of healthy human adults that played repeated games against artificial Bayesian agents. These AA differed in their sophistication. Crucially, participants were tested in a social vs non-social framing. In the social framing, they were made to believe that they were playing against each other. In contrast, they were also made to believe that they were gambling against a one-armed bandit-machine, like in a casino, in the non-social framing.

By comparing different models that incorporated, or not, meta-representation at different depths of sophistication, Devaine et al. showed an added-value of meta-bayesian ToM models in the context of repeated social interactions. Indeed, participants' choice sequences were best explained by sophisticated mentalizing Bayesian learning models (k-ToM models), but only in the social framing (see Figure 16). Interestingly, they found statistical evidence that ToM sophistication varied across people in this social condition, and that it was likely upper-bounded by a maximum recursive depth of meta-representation of 2. In contrast, in the non-social framing condition, they were better fitted by domain-general learning heuristics that did not incorporate any meta-representational component. The authors also observed a strong social framing effect at the behavioral level. Participants were frequently able to defeat sophisticated (artificial) mentalizing agents in the social framing condition, whereas they most frequently failed against the same artificial agents in the non-social framing condition. The ability of participants to predict the behavior of artificial mentalizing agents efficiently was consequently highly dependent on the social framing condition. This corroborates the difference in nature of the computational models that best modelled the participants behavior in the two conditions. Simple heuristics without any meta-representation are indeed disadvantaged against artificial mentalizing agents, while k-ToM models can engage in recursive inference. This also demonstrates the importance to be

incentivized *a priori* to attribute mental states to others in order to be able to decipher intentional behavior.

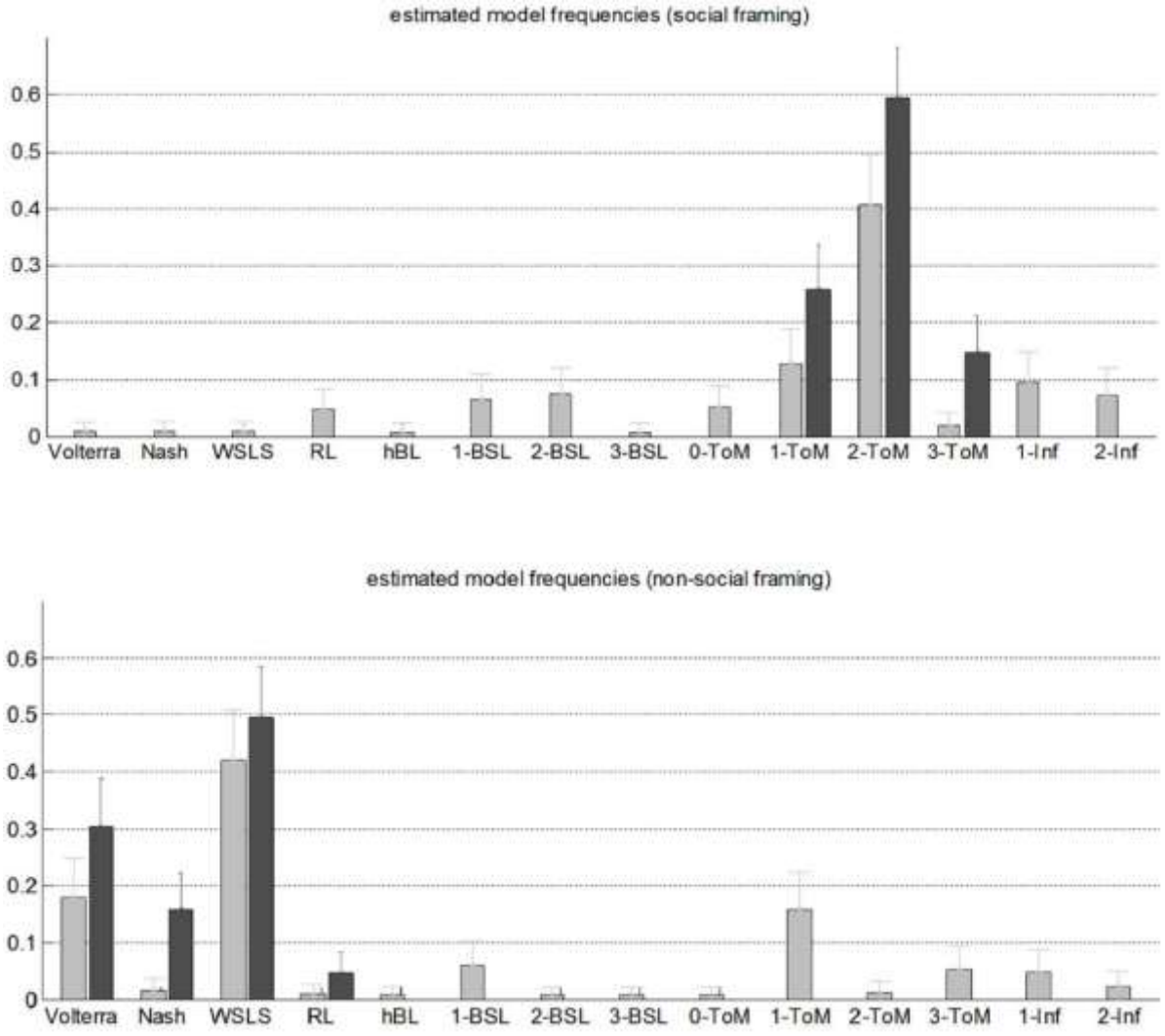


Figure 16: ***Distribution of ToM sophistication in Humans****. Top: Estimated model frequencies in the social framing (dark grey: having restricted the models to the winning Bayesian mentalizing family). Error bars depict one posterior standard error. Bottom: Estimated model frequencies in the non-social framing (dark grey: having restricted the models to the winning non-Bayesian non-mentalizing family). From Devaine et al. (2014)* [197].

The use of SDM and computational modeling approaches are particularly adapted to assess developmental aspects of ToM abilities. It is possible to quantify the drift in ToM sophistication that occurs with age, in early life as well as in older adults [148]. It is also relevant for performing ethological inter-species comparisons of ToM sophistication (e.g. monkeys, great apes and humans). This latter question motivated Devaine et al (2017) [85] to test ToM sophistication across different primate species in either competitive or cooperative conditions. They found a difference in learning models between large-brained vs small-brained primate species (n=7): species with larger brains were more frequently best fitted by mentalizing influence models while species with smaller brains were more frequently best fitted by non-mentalizing models (see Figure 17).

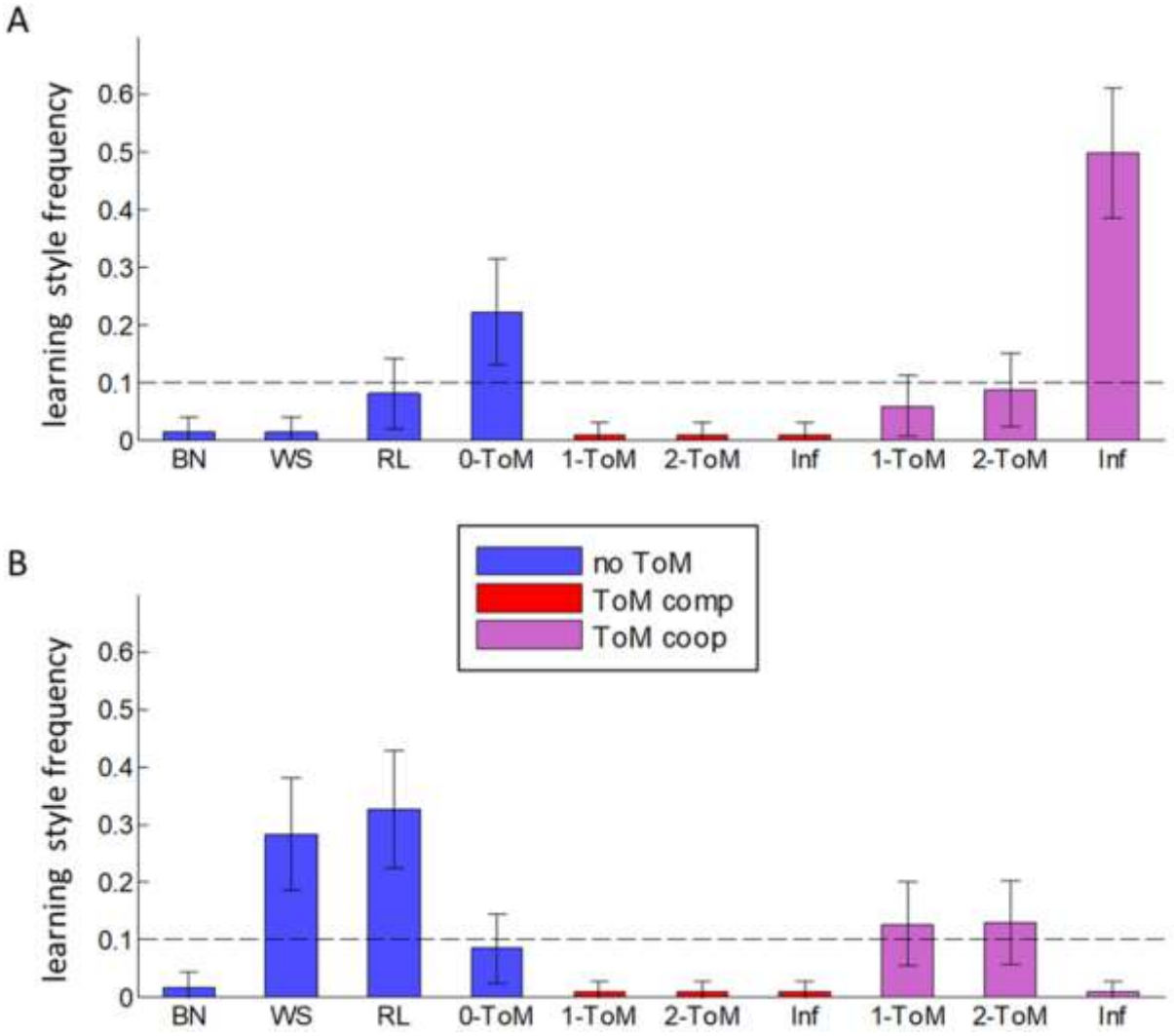


Figure 17: ***Estimated frequencies of learning styles in NHPs****. The posterior mean of model frequency (y-axis) is shown for each learning style (x-axis), among species with large brains (A) and small brains (B). Note that the median-split on ECV actually separates apes from prosimians and monkeys, which is consistent with primates' phylogenic relationships (see S1 Text). The colour code indicates the type of learning style (blue: no-ToM, red: competitive ToM, violet: cooperative ToM). Error bars indicate posterior standard deviations. For comparison purposes, chance level (10%) is indicated using black dotted lines. From Devaine et al. (2017)* [85].

iv. Description of some common computational models

Bayesian Sequence Learning:

The BSL model tracks the (log) odds of P(ot=1|ot-K), where o is the partner's move (binary outcome). This variable is updated according to a Laplace-Kalman filter, yielding 2 sufficient statistics (mean and variance) per combination of past outcomes. BSL can learn sequences of arbitrary depth K. For example:

- if K=1, then BSL tracks 2 probabilities, namely: P(ut=1|ut-1=1) and P(ut=1|ut-1=0). In this case, BSL needs to know about the opponent's previous move ut-1.

- if K=2, then BSL tracks 4 probabilities, namely: P(ut=1|ut-1=1, ut-2=1), P(ut=1|ut-1=0,ut-2=1), P(ut=1|ut-1=1,ut-2=0) and P(ut=1|ut-1=0,ut-2=0). In this case, BSL needs to know about the opponent's two previous moves ut-1 and ut-2. etc.

More generally, the BSL model tracks 2K probabilities.

In this scheme, the only evolution parameter is BSL's prior volatility about the (log) odds.
The choice of the BSL agent (a=1 or a=0) is then based upon its prediction of its partner's next move (o=1 or o=0), given the game payoff table. The evolution function takes 1 parameter which is the BSL's prior volatility about the environment and the observation function takes 2 parameters which are the agent's temperature, which represents the agent's tendency to explore vs exploit, and the agent's bias towards one choice.

Q-learning:

The constituents of this model in its simplest form are:

- a set of (action/item) value states. In two-armed bandit problems, there are two of these (n in a n-armed). In general, there will be as many values as there are available actions. Some behavioral biases can be captured by the initial conditions on these hidden states.

- a learning rate. This parameter controls the impact of reward prediction error on the value update. Note that one may want to ask whether the learning rate depends upon experimental factors (pathological condition, gain/loss domains, etc…)

- a behavioral temperature and a bias. These parameters control the exploitation vs exploration ratio of the agent, and capture choice randomness.

Q-learning models simply assume that subjects update the value of possible actions. In its simplest form, the Q-learning algorithm expresses the change in value Q(t+1)-Q(t) from trial t to trial t+1 as being linearly proportional to the prediction error. This yields the following learning rule:

$$Q_{t+1} = Q_t + \alpha * (r_{t+1} - Q_t)$$

Where rt is the reward delivered to the subject at trial t, and α is the (unknown) "learning rate" of the subject.

The Q-learning evolution function thus takes the agent's previous action and the feedback received for the previous action as data inputs. We complement with a softmax decision rule in the Q-learning observation function. As a reminder, this is an equation that expresses the probability Pt(ai) of the subject to choose action ai at trial t:

$$P_t(a_i) = \frac{\exp(\beta * Q_t(a_i))}{\sum_j \beta * Q_t(a_j)}$$

Where β is the (unknown) temperature.

Fitting the above Q-learning model to behavioral data means finding estimates of the learning rate α, the inverse temperature β, and the initial values Q0 that best explains the observed choices.

Win-Stay Lose-Switch (WSLS) model

This heuristic [183] keeps the same target as the previous one if previous reward was positive, and changes targets otherwise.

Influence model

This mentalizing model was described and adapted from Hampton et al, who also found neural correlates with this model in humans [28]. The influence model relies on a Taylor expended reinforcement learning. This model describes computations underlying the capacity to

mentalize (in the context of a strategic game). Let $p_n = P_n(o = 1)$ be the agent's prediction of the other's next move, i.e., the probability that the other will pick the first alternative option. The "influence learning" rule can be written as follows:

$$p_{n+1} = p_n + \eta * (o - p_n) - \lambda * k_1 * p_n * (1 - p_n) * (a - \frac{\beta * x_n - k_2}{k_1})$$

where o is the other's last move, a is the agent's last move, η is the weight of the agent's prediction error, λ is the weight of the other's prediction error, k_1 & k_2 are derived from the game's payoff table, x_n is the tracked log-odds of p_n, β is the other's temperature, i.e. the tendency of the other to explore vs exploit information (the closer it is to 0, the more it explores by choosing randomly without using previous information, and the closer it is 1, the more it exploits previous information and chooses deterministically).

k-ToM models

I will give here only the intuition behind k-ToM models, for detailed mathematical description of the model, refer to Devaine et al. [85], [197]. k-ToM models use recursive Bayesian inferences of depth k to predict the future choice of the other player. In brief, the repeated observation of the other's actions, noted a_{other}, gives the agent the opportunity to learn the other's behavioral tendency p_{other}.

At the lowest level of sophistication, i.e. k=0, the agent does not attribute any mental state to the other, but rather tracks her behavioral tendency p_{other} without mentalizing. They simply assume that their opponents choose their actions with a probability that varies in time with a certain volatility. They keep track of this probability by updating their prior estimation with Bayes' rule at each trial. Note that this would equate to a BSL0 agent.

ToM comes into play for higher recursion depth k>0, i.e. higher levels of sophistication. When k-ToM agents consider that the other is also a Bayesian agent whose decision policy p_{other} is driven by her hidden beliefs and desires. Distinct recursion levels will induce differences in the way agents update their subjective prediction p_{other}. For example, a 1-ToM agent considers that she is facing a 0-ToM agent, and will therefore predict 0-ToM's next move, based on her inference on the 0-ToM agent's beliefs and the choices' payoffs of both players. Eventually, a 1-ToM agent will learn how a 0-ToM agent learns about herself, and act accordingly. Similarly, a n-ToM agent will assume that they are facing a (n-1)-ToM agent and adapt accordingly.

D. Aim of dissertation

To conclude this general introduction, I want to highlight the broadness of the possible applications of the computational approaches previously described. I only focus on ToM abilities in this dissertation. Of particular interest, the following sets of open research questions arise:

(A) To what extent are the meta-representational abilities of non-human primates, especially monkeys, sufficiently sophisticated to allow ToM? [77], [83], [93],

(B) What is the developmental trajectory of ToM abilities in normal human children?

(C) How can a computational approach help answer these questions?

The challenge to study ToM abilities in NHP or young children is their inability to clearly communicate with and understand experimenters. As discussed in part C., an interesting

theoretical framework is strategic SDM in combination with computational modeling. However, in this dissertation, I only tackle a very specific domain of strategic SDM: ToM applied to coordination learning games. To fully unleash the potential of our approach, we will need to establish a robust experimental protocol to test such abilities. Therefore, I will explore, in our projects, novel experimental designs, developed by ourselves, and exploiting newly available technologies and setups, e.g., tactile screens.

I will study different types of populations: human adults and children of different ages (from 3.5 to 9 years old), baboons (Papio papio), and rhesus macaques. For each project, I wanted to study different aspects of ToM and of coordination learning with a computational modeling approach. The ToM subcomponents targeted in each of our projects are: 1) interactive online imitation/emulation learning (project 1, in macaques vs humans), 2) the ability to strategically mentalize the consequence of one's action on the other's outcome (project 2, in baboons), 3) Adaptability to non-signaled changes of competitive vs cooperative intentions (project 3, in children from age 3.5 to 9).

As for computational modeling, I will take the following approach in each of our projects:

1) First, display the behavior with relevant data plots (EDA step). It will usually be the action selection ratio and/or the tendency to persist with a previous action, averaged on subjects and/or time.
2) Second, to get a better idea of the set of algorithmic computational models to select, I will proceed to study the correlational dependencies between predictors and the dependent variable, i.e., a relevant behavioral signature. To do so, I will use GLME (generalized linear mixed-effects) models because of their versatility and interesting properties in describing repeated decisions for different subjects in complex experimental designs. I will optimize the GLME models with a step-by-step drop-term procedure based on the AIC. This will help us discard useless predictors. This will be done in R with well-known libraries and tools (e.g., lme4).
3) Based on the previous results, I will select algorithmic computational models already existing in the literature and design new ones that are relevant to our experiment and inspired by existing ones. These will be a first set of candidate models to explain our behavioral data.
4) The next step will be to fit each candidate model to the data. This will be performed with a Variational Bayes Approach on Matlab with the VBA toolbox. This will indeed ensure a good balance between the computational resources needed to optimize model parameters and the accuracy of the optimization.
5) Confusion analysis will be performed for all candidate models by simulation on the range of values that was estimated in the previous step. The Monte-Carlo simulations and model inversions will also be performed in Matlab with the VBA-toolbox. Models will be discarded until the confusion matrix becomes (mostly) diagonal.
6) From there, I can perform a BMS relatively confidently. The BMS will be based on the free energy as the model (log-)evidence because group-BMS was found to perform significantly worse with AIC and/or BIC indices [226]. From there, I will be able to make some relatively robust interpretations.

7) If I need to interpret or use the estimated model parameters of one model or more, for example to compare different groups or conditions, I will proceed to parameter recovery.
8) At this stage, all other supplementary analyses will be project specific.

E. References

[1] A. M. Leslie, O. Friedman, and T. P. German, 'Core mechanisms in "theory of mind"', *Trends Cogn Sci*, vol. 8, no. 12, pp. 528–533, Dec. 2004, doi: 10.1016/j.tics.2004.10.001.

[2] D. Premack and G. Woodruff, 'Does the chimpanzee have a theory of mind?', *Behavioral and Brain Sciences*, vol. 1, no. 4, pp. 515–526, 1978, doi: 10.1017/S0140525X00076512.

[3] E. S. Savage-Rumbaugh, D. M. Rumbaugh, D. Premack, and G. Woodruff, 'Chimpanzee problem comprehension: Insufficient evidence', *Science (1979)*, vol. 206, no. 4423, pp. 1201–1202, 1979, doi: 10.1126/science.505009.

[4] A. Whiten and R. W. Byrne, 'Tactical deception in primates', *Behavioral and Brain Sciences*, vol. 11, no. 2, pp. 233–244, 1988, doi: 10.1017/S0140525X00049682.

[5] D. J. Povinelli, K. E. Nelson, and S. T. Boysen, 'Inferences about guessing and knowing by chimpanzees (Pan troglodytes).', *J Comp Psychol*, vol. 104, no. 3, pp. 203–210, 1990, doi: 10.1037/0735-7036.104.3.203.

[6] D. J. Povinelli and T. J. Eddy, 'Chimpanzees: Joint Visual Attention', *Psychol Sci*, vol. 7, no. 3, pp. 129–135, May 1996, doi: 10.1111/j.1467-9280.1996.tb00345.x.

[7] M. Tomasello and J. Call, 'Social cognition of monkeys and apes', *Am J Phys Anthropol*, vol. 37, no. 19 S, pp. 273–305, Jan. 1994, doi: 10.1002/ajpa.1330370610.

[8] R. W. Byrne and A. Whiten, 'Cognitive Evolution in Primates: Evidence from Tactical Deception', *Man*, vol. 27, no. 3, p. 609, Sep. 1992, doi: 10.2307/2803931.

[9] H. Wimmer and J. Perner, 'Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception', *Cognition*, vol. 13, no. 1, pp. 103–128, 1983, doi: 10.1016/0010-0277(83)90004-5.

[10] J. Perner, 'Higher-order beliefs and intentions in children's understanding of social interaction', in *Developing Theories of Mind*, 1988, pp. 271–294. Accessed: Aug. 22, 2022. [Online]. Available: https://books.google.com/books?hl=fr&lr=&id=R5c4AAAAIAAJ&oi=fnd&pg=PA271&dq=+Perner,+J.+(1988).+Higher-order+beliefs+and+intentions+in+children's+understanding+of+social+interaction.+In+J.+W.+Astington,+P.+L.+Harris,+%26+D.+R.+Olson+(Eds.),+Developing+the

[11] J. Perner and H. Wimmer, '"John thinks that Mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children', *J Exp Child Psychol*, vol. 39, no. 3, pp. 437–471, Jun. 1985, doi: 10.1016/0022-0965(85)90051-7.

[12] J. Perner, S. R. Leekam, and H. Wimmer, 'Three-year-olds' difficulty with false belief: The case for a conceptual deficit', *British Journal of Developmental Psychology*, vol. 5, no. 2, pp. 125–137, Jun. 1987, doi: 10.1111/j.2044-835x.1987.tb01048.x.

[13] G.-J. Hogrefe, H. Wimmer, and J. Perner, 'Ignorance versus False Belief: A Developmental Lag in Attribution of Epistemic States', *Child Dev*, vol. 57, no. 3, p. 567, Jun. 1986, doi: 10.2307/1130337.

[14] S. Baron-Cohen, A. M. Leslie, and U. Frith, 'Does the autistic child have a "theory of mind" ?', *Cognition*, vol. 21, no. 1, pp. 37–46, Oct. 1985, doi: 10.1016/0010-0277(85)90022-8.

[15] A. Seed and Michael. Tomasello, 'Primate cognition', *Top Cogn Sci*, vol. 2, no. 3, pp. 407–419, 2010, doi: 10.1111/j.1756-8765.2010.01099.x.

[16] B. Hare, 'Can competitive paradigms increase the validity of experiments on primate social cognition?', *Animal Cognition*, vol. 4, no. 3–4. Springer, pp. 269–280, Jun. 02, 2001. doi: 10.1007/s100710100084.

[17] B. Hare and M. Tomasello, 'Chimpanzees are more skilful in competitive than in cooperative cognitive tasks', *Anim Behav*, vol. 68, no. 3, pp. 571–581, Sep. 2004, doi: 10.1016/j.anbehav.2003.11.011.

[18] C. M. Johnson and M. R. Karin-D'Arcy, 'Social Attention in Nonhuman Primates: A Behavioral Review', *Aquat Mamm*, vol. 32, no. 4, pp. 423–442, Mar. 2007, doi: 10.1578/am.32.4.2006.423.

[19] R. W. Byrne and A. Whiten, Eds., *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans.* New York, NY, US: Clarendon Press/Oxford University Press, 1988.

[20] B. Hare, J. Call, B. Agnetta, and M. Tomasello, 'Chimpanzees know what conspecifics do and do not see', *Anim Behav*, vol. 59, no. 4, pp. 771–785, Apr. 2000, doi: 10.1006/anbe.1999.1377.

[21] B. Hare, J. Call, and M. Tomasello, 'Do chimpanzees know what conspecifics know?', *Anim Behav*, vol. 61, no. 1, pp. 139–151, 2001, doi: 10.1006/anbe.2000.1518.

[22] Á. M. Kovács, E. Téglás, and A. D. Endress, 'The social sense: Susceptibility to others' beliefs in human infants and adults', *Science (1979)*, vol. 330, no. 6012, pp. 1830–1834, Dec. 2010, doi: 10.1126/science.1190792.

[23] D. Houser and K. McCabe, 'Experimental Economics and Experimental Game Theory', in *Neuroeconomics: Decision Making and the Brain: Second Edition*, Academic Press, 2013, pp. 19–34. doi: 10.1016/B978-0-12-416008-8.00002-4.

[24] M. Schurz, J. Radua, M. Aichhorn, F. Richlan, and J. Perner, 'Fractionating theory of mind: A meta-analysis of functional brain imaging studies', *Neuroscience and Biobehavioral Reviews*, vol. 42. Elsevier Ltd, pp. 9–34, 2014. doi: 10.1016/j.neubiorev.2014.01.009.

[25] H. L. Gallagher and C. D. Frith, 'Functional imaging of "theory of mind"', *Trends in Cognitive Sciences*, vol. 7, no. 2. Elsevier, pp. 77–83, Feb. 01, 2003. doi: 10.1016/S1364-6613(02)00025-6.

[26] M. Siegal and R. Varley, 'Neural systems involved in "theory of mind"', *Nat Rev Neurosci*, vol. 3, no. 6, pp. 463–471, 2002, doi: 10.1038/nrn844.

[27] C. J. Charpentier, K. Iigaya, and J. P. O'Doherty, 'A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning', *Neuron*, vol. 106, no. 4, pp. 687-699.e7, May 2020, doi: 10.1016/j.neuron.2020.02.028.

[28] A. N. Hampton, P. Bossaerts, and J. P. O'Doherty, 'Neural correlates of mentalizing-related computations during strategic interactions in humans', *Proc Natl Acad Sci U S A*, vol. 105, no. 18, pp. 6741–6746, 2008, doi: 10.1073/pnas.0711099105.

[29] C. A. Hill, S. Suzuki, R. Polania, M. Moisa, J. P. O'doherty, and C. C. Ruff, 'A causal account of the brain network computations underlying strategic social behavior', *Nat Neurosci*, vol. 20, no. 8, pp. 1142–1149, 2017, doi: 10.1038/nn.4602.

[30] W. Yoshida, B. Seymour, K. J. Friston, and R. J. Dolan, 'Neural mechanisms of belief inference during cooperative games', *Journal of Neuroscience*, vol. 30, no. 32, pp. 10744–10751, Aug. 2010, doi: 10.1523/JNEUROSCI.5895-09.2010.

[31] Y. Xiao, F. Geng, T. Riggins, G. Chen, and E. Redcay, 'Neural correlates of developing theory of mind competence in early childhood', *Neuroimage*, vol. 184, pp. 707–716, Jan. 2019, doi: 10.1016/j.neuroimage.2018.09.079.

[32] H. De Weerd, R. Verbrugge, and B. Verheij, 'How much does it help to know what she knows you know? An agent-based simulation study', *Artif Intell*, vol. 199–200, pp. 67–92, Jun. 2013, doi: 10.1016/j.artint.2013.05.004.

[33] I. Sher, M. Koenig, and A. Rustichini, 'Children's strategic theory of mind', *Proc Natl Acad Sci U S A*, vol. 111, no. 37, pp. 13307–13312, Sep. 2014, doi: 10.1073/pnas.1403283111.

[34] H. M. Wellman and D. Liu, 'Scaling of theory-of-mind tasks', *Child Development*, vol. 75, no. 2. Child Dev, pp. 523–541, Mar. 2004. doi: 10.1111/j.1467-8624.2004.00691.x.

[35] H. M. Wellman, F. Fang, and C. C. Peterson, 'Sequential Progressions in a Theory-of-Mind Scale: Longitudinal Perspectives', *Child Dev*, vol. 82, no. 3, pp. 780–792, 2011, doi: 10.1111/j.1467-8624.2011.01583.x.

[36] H. Rakoczy, 'Foundations of theory of mind and its development in early childhood', *Nature Reviews Psychology*, vol. 1, no. 4, pp. 223–235, Mar. 2022, doi: 10.1038/s44159-022-00037-z.

[37] A. Gopnik and J. W. Astington, 'Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction.', *Child Dev*, vol. 59, no. 1, pp. 26–37, 1988, doi: 10.1111/j.1467-8624.1988.tb03192.x.

[38] J. H. Flavell, E. R. Flavell, and F. L. Green, 'Development of the appearance-reality distinction', *Cogn Psychol*, vol. 15, no. 1, pp. 95–120, Jan. 1983, doi: 10.1016/0010-0285(83)90005-1.

[39] H. Rakoczy, D. Bergfeld, I. Schwarz, and E. Fizke, 'Explicit Theory of Mind Is Even More Unified Than Previously Assumed: Belief Ascription and Understanding Aspectuality Emerge Together in Development', *Child Dev*, vol. 86, no. 2, pp. 486–502, Mar. 2015, doi: 10.1111/cdev.12311.

[40] M. Taumoepeau, S. Sadeghi, and A. Nobilo, 'Cross-cultural differences in children's theory of mind in Iran and New Zealand: The role of caregiver mental state talk', *Cogn Dev*, vol. 51, pp. 32–45, Jul. 2019, doi: 10.1016/j.cogdev.2019.05.004.

[41] A. Shahaeian, C. C. Peterson, V. Slaughter, and H. M. Wellman, 'Culture and the Sequence of Steps in Theory of Mind Development', *Dev Psychol*, vol. 47, no. 5, pp. 1239–1247, Sep. 2011, doi: 10.1037/a0023899.

[42] E. E. F. Bradford, I. Jentzsch, J. C. Gomez, Y. Chen, D. Zhang, and Y. Su, 'Cross-cultural differences in adult Theory of Mind abilities: A comparison of native-English speakers and native-Chinese speakers on the Self/Other Differentiation task', *Quarterly Journal of Experimental Psychology*, vol. 71, no. 12, pp. 2665–2676, Dec. 2018, doi: 10.1177/1747021818757170.

[43] T. Callaghan *et al.*, 'Synchrony in the onset of mental-state reasoning: Evidence from five cultures', *Psychol Sci*, vol. 16, no. 5, pp. 378–384, May 2005, doi: 10.1111/j.0956-7976.2005.01544.x.

[44] S. Koerber and C. Osterhaus, 'Some but not all aspects of (advanced) theory of mind predict loneliness', *British Journal of Developmental Psychology*, vol. 38, no. 1, pp. 144–148, Mar. 2020, doi: 10.1111/bjdp.12302.

[45] C. Osterhaus and S. Koerber, 'The Development of Advanced Theory of Mind in Middle Childhood: A Longitudinal Study From Age 5 to 10 Years', *Child Dev*, vol. 92, no. 5, pp. 1872–1888, Sep. 2021, doi: 10.1111/cdev.13627.

[46] C. Osterhaus, S. Koerber, and B. Sodian, 'Scaling of Advanced Theory-of-Mind Tasks', *Child Dev*, vol. 87, no. 6, pp. 1971–1991, 2016, doi: 10.1111/cdev.12566.

[47] E. O. Hayward and B. D. Homer, 'Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence', *British Journal of Developmental Psychology*, vol. 35, no. 3, pp. 454–462, 2017, doi: 10.1111/bjdp.12186.

[48] K. Sullivan, D. Zaitchik, and H. Tager-Flusberg, 'Preschoolers Can Attribute Second-Order Beliefs', *Dev Psychol*, vol. 30, no. 3, pp. 395–402, 1994, doi: 10.1037/0012-1649.30.3.395.

[49] S. Baron-Cohen, M. O'Riordan, V. Stone, R. Jones, and K. Plaisted, 'Recognition of faux pas by normally developing children and children with asperger syndrome or high-functioning autism', *J Autism Dev Disord*, vol. 29, no. 5, pp. 407–418, 1999, doi: 10.1023/A:1023035012436.

[50] J. I. Carpendale and M. J. Chandler, 'On the Distinction between False Belief Understanding and Subscribing to an Interpretive Theory of Mind', *Child Dev*, vol. 67, no. 4, p. 1686, Aug. 1996, doi: 10.2307/1131725.

[51] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb, 'The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism', *Journal of Child Psychology and Psychiatry*, vol. 42, no. 2, pp. 241–251, Feb. 2001, doi: 10.1111/1469-7610.00715.

[52] C. Osterhaus and S. Koerber, 'Social cognition during and after kindergarten: The relations between first-order and advanced theories of mind', *European Journal of Developmental Psychology*, vol. 18, no. 4, pp. 573–592, 2021, doi: 10.1080/17405629.2020.1820861.

[53] S. A. Miller, 'Children's Understanding of Second-Order Mental States', *Psychol Bull*, vol. 135, no. 5, pp. 749–773, 2009, doi: 10.1037/a0016854.

[54] B. Sodian, S. Kristen-Antonow, and D. Kloo, 'How Does Children's Theory of Mind Become Explicit? A Review of Longitudinal Findings', *Child Dev Perspect*, vol. 14, no. 3, pp. 171–177, Sep. 2020, doi: 10.1111/cdep.12381.

[55] H. M. Wellman, 'Developing a Theory of Mind', in *The Wiley-Blackwell Handbook of Childhood Cognitive Development, Second edition*, John Wiley & Sons, Ltd, 2010, pp. 258–284. doi: 10.1002/9781444325485.ch10.

[56] R. Brooks and A. N. Meltzoff, 'The importance of eyes: how infants interpret adult looking behavior.', *Dev Psychol*, vol. 38, no. 6, pp. 958–966, 2002, doi: 10.1037/0012-1649.38.6.958.

[57] A. L. Woodward, 'Infants selectively encode the goal object of an actor's reach', *Cognition*, vol. 69, no. 1, pp. 1–34, 1998, doi: 10.1016/S0010-0277(98)00058-4.

[58] A. N. Meltzoff, 'Understanding the Intentions of Others: Re-Enactment of Intended Acts by 18-Month-Old Children', *Dev Psychol*, vol. 31, no. 5, pp. 838–850, 1995, doi: 10.1037/0012-1649.31.5.838.

[59] G. Gergely and G. Csibra, 'Teleological reasoning in infancy: The infant's naive theory of rational action a reply to Premack and Premack', *Cognition*, vol. 63, no. 2, pp. 227–233, May 1997, doi: 10.1016/S0010-0277(97)00004-8.

[60] B. M. Repacholi and A. Gopnik, 'Early reasoning about desires: evidence from 14- and 18-month-olds.', *Dev Psychol*, vol. 33, no. 1, pp. 12–21, 1997, doi: 10.1037/0012-1649.33.1.12.

[61] J. H. Flavell, S. G. Shipstead, and K. Croft, 'Young children's knowledge about visual perception: hiding objects from others.', *Child Dev*, vol. 49, no. 4, pp. 1208–1211, 1978, doi: 10.1111/j.1467-8624.1978.tb04090.x.

[62] R. Baillargeon, R. M. Scott, and Z. He, 'False-belief understanding in infants', *Trends in Cognitive Sciences*, vol. 14, no. 3. Elsevier, pp. 110–118, Mar. 01, 2010. doi: 10.1016/j.tics.2009.12.006.

[63] H. joo Song and R. Baillargeon, 'Infants' Reasoning About Others' False Perceptions', *Dev Psychol*, vol. 44, no. 6, pp. 1789–1795, Nov. 2008, doi: 10.1037/a0013774.

[64] L. Surian, S. Caldi, and D. Sperber, 'Attribution of beliefs by 13-month-old infants', *Psychol Sci*, vol. 18, no. 7, pp. 580–586, Jul. 2007, doi: 10.1111/j.1467-9280.2007.01943.x.

[65] V. Southgate, A. Senju, and G. Csibra, 'Action anticipation through attribution of false belief by 2-year-olds', *Psychol Sci*, vol. 18, no. 7, pp. 587–592, Jul. 2007, doi: 10.1111/j.1467-9280.2007.01944.x.

[66] H. joo Song, K. H. Onishi, R. Baillargeon, and C. Fisher, 'Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants', *Cognition*, vol. 109, no. 3, pp. 295–315, Dec. 2008, doi: 10.1016/j.cognition.2008.08.008.

[67] K. K. Oniski and R. Baillargeon, 'Do 15-month-old infants understand false beliefs?', *Science (1979)*, vol. 308, no. 5719, pp. 255–258, Apr. 2005, doi: 10.1126/science.1107621.

[68] Á. M. Kovács, E. Téglás, and G. Csibra, 'Can infants adopt underspecified contents into attributed beliefs? Representational prerequisites of theory of mind', *Cognition*, vol. 213, Aug. 2021, doi: 10.1016/j.cognition.2021.104640.

[69] W. A. Garnham and T. Ruffman, 'Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief?', *Dev Sci*, vol. 4, no. 1, pp. 94–100, Mar. 2001, doi: 10.1111/1467-7687.00153.

[70] W. A. Clements and J. Perner, 'Implicit understanding of belief', *Cogn Dev*, vol. 9, no. 4, pp. 377–395, Oct. 1994, doi: 10.1016/0885-2014(94)90012-4.

[71] D. Schneider, A. P. Bayliss, S. I. Becker, and P. E. Dux, 'Eye movements reveal sustained implicit processing of others' mental states', *J Exp Psychol Gen*, vol. 141, no. 3, pp. 433–438, Aug. 2012, doi: 10.1037/a0025458.

[72] L. Surian and A. Geraci, 'Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months', *British Journal of Developmental Psychology*, vol. 30, no. 1, pp. 30–44, Mar. 2012, doi: 10.1111/j.2044-835X.2011.02046.x.

[73] D. Buttelmann, M. Carpenter, and M. Tomasello, 'Eighteen-month-old infants show false belief understanding in an active helping paradigm', *Cognition*, vol. 112, no. 2, pp. 337–342, Aug. 2009, doi: 10.1016/j.cognition.2009.05.006.

[74] V. Southgate, C. Chevallier, and G. Csibra, 'Seventeen-month-olds appeal to false beliefs to interpret others' referential communication', *Dev Sci*, vol. 13, no. 6, pp. 907–912, Nov. 2010, doi: 10.1111/j.1467-7687.2009.00946.x.

[75] D. Samson, I. A. Apperly, J. J. Braithwaite, B. J. Andrews, and S. E. Bodley Scott, 'Seeing it their Way: Evidence for rapid and involuntary computation of what other people see', *J Exp Psychol Hum Percept Perform*, vol. 36, no. 5, pp. 1255–1266, Oct. 2010, doi: 10.1037/a0018729.

[76] J. Call and M. Tomasello, 'Does the chimpanzee have a theory of mind? 30 years later', *Trends Cogn Sci*, vol. 12, no. 5, pp. 187–192, 2008, doi: 10.1016/j.tics.2008.02.010.

[77] H. Meunier, 'Do monkeys have a theory of mind? How to answer the question?', *Neurosci Biobehav Rev*, vol. 82, pp. 110–123, 2017, doi: 10.1016/j.neubiorev.2016.11.007.

[78] L. A. Drayton and L. R. Santos, 'A decade of theory of mind research on cayo santiago: Insights into rhesus macaque social cognition', *Am J Primatol*, vol. 78, no. 1, pp. 106–116, Jan. 2016, doi: 10.1002/AJP.22362.

[79] D. C. W. Marticorena, A. M. Ruiz, C. Mukerji, A. Goddu, and L. R. Santos, 'Monkeys represent others' knowledge but not their beliefs', *Dev Sci*, vol. 14, no. 6, pp. 1406–1416, Nov. 2011, doi: 10.1111/j.1467-7687.2011.01085.x.

[80] C. Krachun, M. Carpenter, J. Call, and M. Tomasello, 'A competitive nonverbal false belief task for children and apes', *Dev Sci*, vol. 12, no. 4, pp. 521–535, 2009, doi: 10.1111/j.1467-7687.2008.00793.x.

[81] J. Kaminski, J. Call, and M. Tomasello, 'Chimpanzees know what others know, but not what they believe', *Cognition*, vol. 109, no. 2, pp. 224–234, 2008, doi: 10.1016/j.cognition.2008.08.010.

[82] A. Martin and L. R. Santos, 'What Cognitive Representations Support Primate Theory of Mind?', *Trends Cogn Sci*, vol. 20, no. 5, pp. 375–382, 2016, doi: 10.1016/j.tics.2016.03.005.

[83] C. Krupenye and J. Call, 'Theory of mind in animals : Current and future directions', *WIREs Cognitive Science*, vol. 10, no. 6, 2019.

[84] R. M. Seyfarth and D. L. Cheney, 'Affiliation, empathy, and the origins of Theory of Mind', *Proc Natl Acad Sci U S A*, vol. 110, no. SUPPL2, pp. 10349–10356, 2013, doi: 10.1073/pnas.1301223110.

[85] M. Devaine *et al.*, 'Reading wild minds: A computational assay of Theory of Mind sophistication across seven primate species', *PLoS Comput Biol*, vol. 13, no. 11, pp. 1–25, 2017, doi: 10.1371/journal.pcbi.1005833.

[86] J. D. Johnson, J. Li, and Z. Chen, 'Reinforcement Learning: An Introduction', *Neurocomputing*, vol. 35, no. 1–4, pp. 205–206, 2000, doi: 10.1016/s0925-2312(00)00324-6.

[87] P. Dayan and B. W. Balleine, 'Reward, motivation, and reinforcement learning', *Neuron*, vol. 36, no. 2. Cell Press, pp. 285–298, Oct. 10, 2002. doi: 10.1016/S0896-6273(02)00963-7.

[88] C. Krupenye, F. Kano, S. Hirata, J. Call, and M. Tomasello, 'Great apes anticipate that other individuals will act according to false beliefs', *Science (1979)*, vol. 354, no. 6308, pp. 110–114, Oct. 2016, doi: 10.1126/science.aaf8110.

[89] T. Hayashi *et al.*, 'Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex', *Cell Rep*, vol. 30, no. 13, pp. 4433-4444.e5, Mar. 2020, doi: 10.1016/j.celrep.2020.03.013.

[90] B. L. Roth, 'DREADDs for Neuroscientists', *Neuron*, vol. 89, no. 4. Cell Press, pp. 683–694, Feb. 17, 2016. doi: 10.1016/j.neuron.2016.01.040.

[91] C. M. Heyes, 'Theory of mind in nonhuman primates', *Behavioral and Brain Sciences*, vol. 21, no. 1, pp. 101–148, 1998, doi: 10.1017/S0140525X98000703.

[92] E. van der Vaart and C. K. Hemelrijk, '"Theory of mind" in animals: ways to make progress', *Synthese*, vol. 191, no. 3, pp. 335–354, Nov. 2014, doi: 10.1007/s11229-012-0170-3.

[93] L. S. Lewis and C. Krupenye, 'Theory of Mind in Nonhuman Primates', in *Primate Cognitive Studies*, B. L. Schwartz and M. J. Beran, Eds., Cambridge: Cambridge University Press, 2022, pp. 439–482. doi: DOI: 10.1017/9781108955836.018.

[94] C. Heyes, 'Apes Submentalise', *Trends Cogn Sci*, vol. 21, no. 1, pp. 1–2, 2017, doi: 10.1016/j.tics.2016.11.006.

[95] D. Poulin-Dubois *et al.*, 'Do infants understand false beliefs? We don't know yet – A commentary on Baillargeon, Buttelmann and Southgate's commentary', *Cognitive Development*, vol. 48. JAI, pp. 302–315, Oct. 01, 2018. doi: 10.1016/j.cogdev.2018.09.005.

[96] S. M. Schaafsma, D. W. Pfaff, R. P. Spunt, and R. Adolphs, 'Deconstructing and reconstructing theory of mind', *Trends in Cognitive Sciences*, vol. 19, no. 2. pp. 65–72, 2015. doi: 10.1016/j.tics.2014.11.007.

[97] C. Beaudoin, É. Leblanc, C. Gagner, and M. H. Beauchamp, 'Systematic Review and Inventory of Theory of Mind Measures for Young Children', *Frontiers in Psychology*, vol. 10. Frontiers Media S.A., p. 2905, Jan. 15, 2020. doi: 10.3389/fpsyg.2019.02905.

[98] T. Rusch, S. Steixner-Kumar, P. Doshi, M. Spezio, and J. Gläscher, 'Theory of mind and decision science: Towards a typology of tasks and computational models', *Neuropsychologia*, vol. 146, pp. 1–44, 2020, doi: 10.1016/j.neuropsychologia.2020.107488.

[99] J. M. Barnby, P. Dayan, and V. Bell, 'Formalising social representation to explain psychiatric symptoms', *Trends in cognitive sciences*, vol. 27, no. 3. The Authors, pp. 317–332, 2023. doi: 10.1016/j.tics.2022.12.004.

[100] F. Kano, C. Krupenye, S. Hirata, J. Call, and M. Tomasello, 'Submentalizing Cannot Explain Belief-Based Action Anticipation in Apes', *Trends in Cognitive Sciences*, vol. 21, no. 9. Trends Cogn Sci, pp. 633–634, Sep. 01, 2017. doi: 10.1016/j.tics.2017.06.011.

[101] S. A. Butterfill and I. A. Apperly, 'How to construct a minimal theory of mind', *Mind Lang*, vol. 28, no. 5, pp. 606–637, Nov. 2013, doi: 10.1111/mila.12036.

[102] I. A. Apperly and S. A. Butterfill, 'Do Humans Have Two Systems to Track Beliefs and Belief-Like States?', *Psychol Rev*, vol. 116, no. 4, pp. 953–970, Oct. 2009, doi: 10.1037/a0016923.

[103] J. Perner and T. Ruffman, 'Infants' insight into the mind: How deep?', *Science*, vol. 308, no. 5719. American Association for the Advancement of Science, pp. 214–216, Apr. 08, 2005. doi: 10.1126/science.1111656.

[104] M. Balter, '"Killjoys" challenge claims of clever animals', *Science*, vol. 335, no. 6072. American Association for the Advancement of Science, pp. 1036–1037, Mar. 02, 2012. doi: 10.1126/science.335.6072.1036.

[105] F. Kano, C. Krupenye, S. Hirata, M. Tomonaga, and J. Call, 'Great apes use self-experience to anticipate an agent's action in a false-belief test', *Proc Natl Acad Sci U S A*, vol. 116, no. 42, pp. 20904–20909, Oct. 2019, doi: 10.1073/PNAS.1910095116/SUPPL_FILE/PNAS.1910095116.SM01.MP4.

[106] T. Hayashi *et al.*, 'Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex', *Cell Rep*, vol. 30, no. 13, pp. 4433-4444.e5, Mar. 2020, doi: 10.1016/j.celrep.2020.03.013.

[107] D. J. Horschler, E. L. MacLean, and L. R. Santos, 'Do Non-Human Primates Really Represent Others' Beliefs?', *Trends in Cognitive Sciences*, vol. 24, no. 8. Elsevier, pp. 594–605, Aug. 01, 2020. doi: 10.1016/j.tics.2020.05.009.

[108] C. Heyes, 'Animal mindreading: what's the problem?', *Psychon Bull Rev*, vol. 22, no. 2, pp. 313–327, Aug. 2015, doi: 10.3758/S13423-014-0704-4/METRICS.

[109] R. Lurz, 'If chimpanzees are mindreaders, could behavioral science tell? Toward a solution of the logical problem', *Philos Psychol*, vol. 22, no. 3, pp. 305–328, Jun. 2009, doi: 10.1080/09515080902970673.

[110] D. C. Penn and D. J. Povinelli, 'On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind"', in *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2007, pp. 731–744. doi: 10.1098/rstb.2006.2023.

[111] C. Heyes, 'Animal mindreading: what's the problem?', *Psychonomic Bulletin & Review 2014 22:2*, vol. 22, no. 2, pp. 313–327, Aug. 2014, doi: 10.3758/S13423-014-0704-4.

[112] G. G. Cole, M. A. Atkinson, A. D. C. D'souza, and D. T. Smith, 'Spontaneous perspective taking in humans?', *Vision (Switzerland)*, vol. 1, no. 2, p. 17, Jun. 2017, doi: 10.3390/vision1020017.

[113] L. Kulke, B. von Duhn, D. Schneider, and H. Rakoczy, 'Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study', *Psychol Sci*, vol. 29, no. 6, pp. 888–900, Jun. 2018, doi: 10.1177/0956797617747090.

[114] J. Perner and J. Roessler, 'From infants' to children's appreciation of belief', *Trends in Cognitive Sciences*, vol. 16, no. 10. Trends Cogn Sci, pp. 519–525, Oct. 2012. doi: 10.1016/j.tics.2012.08.004.

[115] H. M. Wellman, D. Cross, and J. Watson, 'Meta-analysis of theory-of-mind development: The truth about false belief', *Child Dev*, vol. 72, no. 3, pp. 655–684, 2001, doi: 10.1111/1467-8624.00304.

[116] J. Yott and D. Poulin-Dubois, 'Are Infants' Theory-of-Mind Abilities Well Integrated? Implicit Understanding of Intentions, Desires, and Beliefs', *Journal of Cognition and Development*, vol. 17, no. 5, pp. 683–698, 2016, doi: 10.1080/15248372.2015.1086771.

[117] S. Dörrenberg, H. Rakoczy, and U. Liszkowski, 'How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures', *Cogn Dev*, vol. 46, no. January, pp. 12–30, 2018, doi: 10.1016/j.cogdev.2018.01.001.

[118] L. J. Powell, K. Hobbs, A. Bardis, S. Carey, and R. Saxe, 'Replications of implicit theory of mind tasks with varying representational demands', *Cogn Dev*, vol. 46, no. September, pp. 40–50, 2018, doi: 10.1016/j.cogdev.2017.10.004.

[119] L. Kulke, M. Reiß, H. Krist, and H. Rakoczy, 'How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span', *Cogn Dev*, vol. 46, no. September, pp. 97–111, 2018, doi: 10.1016/j.cogdev.2017.09.001.

[120] D. Poulin-Dubois and J. Yott, 'Probing the depth of infants' theory of mind: disunity in performance across paradigms', *Dev Sci*, vol. 21, no. 4, Jul. 2018, doi: 10.1111/desc.12600.

[121] R. M. Scott and R. Baillargeon, 'Early False-Belief Understanding', *Trends Cogn Sci*, vol. 21, no. 4, pp. 237–249, Apr. 2017, doi: 10.1016/J.TICS.2017.01.012.

[122] R. Baillargeon, D. Buttelmann, and V. Southgate, 'Invited Commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations', *Cognitive Development*, vol. 46. JAI, pp. 112–124, Apr. 01, 2018. doi: 10.1016/j.cogdev.2018.06.001.

[123] J. R. Stevens, 'Replicability and reproducibility in comparative psychology', *Front Psychol*, vol. 8, no. MAY, p. 253375, May 2017, doi: 10.3389/fpsyg.2017.00862.

[124] H. Kummer, V. Dasser, and P. Hoyningen-Huene, 'Exploring Primate Social Cognition: Some Critical Remarks', *Behaviour*, vol. 112, no. 1–2, pp. 84–98, 1990, doi: 10.1163/156853990X00699.

[125] J. J. Bolhuis and C. D. L. Wynne, 'Can evolution explain how minds work?', *Nature*, vol. 458, no. 7240. Nature Publishing Group, pp. 832–833, Apr. 15, 2009. doi: 10.1038/458832a.

[126] S. J. Shettleworth, 'Clever animals and killjoy explanations in comparative psychology', *Trends Cogn Sci*, vol. 14, no. 11, pp. 477–481, Nov. 2010, doi: 10.1016/j.tics.2010.07.002.

[127] D. J. Povinelli and J. Vonk, 'Chimpanzee minds: Suspiciously human?', *Trends in Cognitive Sciences*, vol. 7, no. 4. Elsevier Ltd, pp. 157–160, Apr. 01, 2003. doi: 10.1016/S1364-6613(03)00053-6.

[128] L. Barrett, P. Henzi, and D. Kendall, 'Social brains, simple minds: Does social complexity really require cognitive complexity?', in *Philosophical Transactions of the Royal Society B: Biological Sciences*, Philos Trans R Soc Lond B Biol Sci, Apr. 2007, pp. 561–575. doi: 10.1098/rstb.2006.1995.

[129] C. Heyes, 'Animal mindreading: what's the problem?', *Psychon Bull Rev*, vol. 22, no. 2, pp. 313–327, Aug. 2015, doi: 10.3758/s13423-014-0704-4.

[130] M. Tomasello, J. Call, and B. Hare, 'Chimpanzees versus humans: It's not that simple', *Trends in Cognitive Sciences*, vol. 7, no. 6. Elsevier Ltd, pp. 239–240, Jun. 01, 2003. doi: 10.1016/S1364-6613(03)00107-4.

[131] D. E. Lyons and L. R. Santos, 'Ecology, Domain Specificity, and the Origins of Theory of Mind: Is Competition the Catalyst?', *Philos Compass*, vol. 1, no. 5, pp. 481–492, Sep. 2006, doi: 10.1111/j.1747-9991.2006.00032.x.

[132] N. J. Emery and N. S. Clayton, 'How to build a scrub-jay that reads minds', in *Origins of the Social Mind: Evolutionary and Developmental Views*, Springer Japan, 2008, pp. 65–97. doi: 10.1007/978-4-431-75179-3_4.

[133] J. Vonk and T. K. Shackelford, 'Toward Bridging Gaps: Finding Commonality between Evolutionary and Comparative Psychology', in *The Oxford Handbook of Comparative Evolutionary Psychology*, Oxford University Press, 2012. doi: 10.1093/oxfordhb/9780199738182.013.0001.

[134] D. J. Povinelli and J. Vonk, 'We Don't Need a Microscope to Explore the Chimpanzee's Mind', *Mind Lang*, vol. 19, no. 1, pp. 1–28, Feb. 2004, doi: 10.1111/j.1468-0017.2004.00244.x.

[135] V. Wobber, B. Hare, J. Maboto, S. Lipson, R. Wrangham, and P. T. Ellison, 'Differential changes in steroid hormones before competition in bonobos and chimpanzees', *Proc Natl Acad Sci U S A*, vol. 107, no. 28, pp. 12457–12462, Jul. 2010, doi: 10.1073/pnas.1007411107.

[136] R. L. C. Mitchell and L. H. Phillips, 'The overlapping relationship between emotion perception and theory of mind', *Neuropsychologia*, vol. 70. Neuropsychologia, pp. 1–10, Apr. 01, 2015. doi: 10.1016/j.neuropsychologia.2015.02.018.

[137] E. Kalbe *et al.*, 'Dissociating cognitive from affective theory of mind: A TMS study', *Cortex*, vol. 46, no. 6, pp. 769–780, Jun. 2010, doi: 10.1016/j.cortex.2009.07.010.

[138] S. Baron-Cohen, 'Social and pragmatic deficits in autism: Cognitive or affective?', *J Autism Dev Disord*, vol. 18, no. 3, pp. 379–402, Sep. 1988, doi: 10.1007/BF02212194.

[139] M. L. Healey and M. Grossman, 'Cognitive and affective perspective-taking: Evidence for shared and dissociable anatomical substrates', *Frontiers in Neurology*, vol. 9, no. JUN. Frontiers Media S.A., Jun. 25, 2018. doi: 10.3389/fneur.2018.00491.

[140] A. Knafo, C. Zahn-Waxler, M. Davidov, C. Van Hulle, J. A. L. Robinson, and S. H. Rhee, 'Empathy in early childhood: Genetic, environmental, and affective contributions', in *Annals of the New York Academy of Sciences*, John Wiley & Sons, Ltd, Jun. 2009, pp. 103–114. doi: 10.1111/j.1749-6632.2009.04540.x.

[141] H. Borke, 'Interpersonal perception of young children: Egocentrism or empathy?', *Dev Psychol*, vol. 5, no. 2, pp. 263–269, Sep. 1971, doi: 10.1037/h0031267.

[142] H. M. Wellman and J. D. Woolley, 'From simple desires to ordinary beliefs: The early development of everyday psychology', *Cognition*, vol. 35, no. 3, pp. 245–275, Jun. 1990, doi: 10.1016/0010-0277(90)90024-E.

[143] H. M. Wellman and K. Bartsch, 'Young children's reasoning about beliefs', *Cognition*, vol. 30, no. 3, pp. 239–277, Dec. 1988, doi: 10.1016/0010-0277(88)90021-2.

[144] S. M. O'Connell, 'Empathy in chimpanzees: Evidence for theory of mind?', *Primates*, vol. 36, no. 3, pp. 397–410, 1995, doi: 10.1007/BF02382862.

[145] B. H. Pillow, 'Early understanding of perception as a source of knowledge', *J Exp Child Psychol*, vol. 47, no. 1, pp. 116–129, Feb. 1989, doi: 10.1016/0022-0965(89)90066-0.

[146] T. K. Ruffman and D. R. Olson, 'Children's Ascriptions of Knowledge to Others', *Dev Psychol*, vol. 25, no. 4, pp. 601–606, 1989, doi: 10.1037/0012-1649.25.4.601.

[147] A. T. Phillips, H. M. Wellman, and E. S. Spelke, 'Infants' ability to connect gaze and emotional expression to intentional action', *Cognition*, vol. 85, no. 1, pp. 53–78, 2002, doi: 10.1016/S0010-0277(02)00073-2.

[148] D. Klindt, M. Devaine, and J. Daunizeau, 'Does the way we read others' mind change over the lifespan? Insights from a massive web poll of cognitive skills from childhood to late adulthood', *Cortex*, vol. 86, pp. 205–215, Jan. 2017, doi: 10.1016/j.cortex.2016.09.009.

[149] S. Zhen and R. Yu, 'Neural correlates of recursive thinking during interpersonal strategic interactions', *Hum Brain Mapp*, vol. 42, no. 7, pp. 2128–2146, May 2021, doi: 10.1002/hbm.25355.

[150] K. H. Lagattuta, H. J. Kramer, K. Kennedy, K. Hjortsvang, D. Goldfarb, and S. Tashjian, 'Beyond Sally's Missing Marble: Further Development in Children's Understanding of Mind and Emotion in Middle Childhood', in *Advances in Child Development and Behavior*, JAI, 2015, pp. 185–217. doi: 10.1016/bs.acdb.2014.11.005.

[151] C. C. Peterson and H. M. Wellman, 'Longitudinal Theory of Mind (ToM) Development From Preschool to Adolescence With and Without ToM Delay', *Child Dev*, vol. 90, no. 6, pp. 1917–1934, Nov. 2019, doi: 10.1111/cdev.13064.

[152] C. C. Peterson, H. M. Wellman, and V. Slaughter, 'The Mind Behind the Message: Advancing Theory-of-Mind Scales for Typically Developing Children, and Those With Deafness, Autism, or Asperger Syndrome', *Child Dev*, vol. 83, no. 2, pp. 469–485, Mar. 2012, doi: 10.1111/j.1467-8624.2011.01728.x.

[153] F. G. E. Happé, 'An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults', *J Autism Dev Disord*, vol. 24, no. 2, pp. 129–154, Apr. 1994, doi: 10.1007/BF02172093.

[154] C. D. Frith and U. Frith, 'The Neural Basis of Mentalizing', *Neuron*, vol. 50, no. 4. Neuron, pp. 531–534, May 18, 2006. doi: 10.1016/j.neuron.2006.05.001.

[155] R. Cabeza, E. Ciaramelli, and M. Moscovitch, 'Cognitive contributions of the ventral parietal cortex: An integrative theoretical account', *Trends in Cognitive Sciences*, vol. 16, no. 6. Elsevier, pp. 338–352, Jun. 01, 2012. doi: 10.1016/j.tics.2012.04.008.

[156] F. Kano, J. Call, and C. Krupenye, 'Primates Pass Dynamically Social Anticipatory-Looking False-Belief Tests', *Trends Cogn Sci*, vol. 24, no. 10, pp. 777–778, Oct. 2020, doi: 10.1016/j.tics.2020.07.003.

[157] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*. 2007. doi: 10.1086/286866.

[158] R. Myerson, 'Game theory: analysis of conflict', *Choice Reviews Online*, vol. 29, no. 05, pp. 29-2753-29–2753, Jan. 1992, doi: 10.5860/choice.29-2753.
[159] W. Edwards, 'The theory of decision making', *Psychol Bull*, vol. 51, no. 4, pp. 380–417, 1954, doi: 10.1037/h0053870.
[160] R. Bogacz, 'Optimal decision-making theories: linking neurobiology with behaviour', *Trends in Cognitive Sciences*, vol. 11, no. 3. Elsevier Current Trends, pp. 118–125, Mar. 01, 2007. doi: 10.1016/j.tics.2006.12.006.
[161] C. C. Cowden, 'Game theory, evolutionary stable strategies and the evolution of biological interactions', *Nature Education*, vol. 3, no. 10, p. 6, Jan. 2012.
[162] R. Axelrod and D. Dion, 'The further evolution of cooperation', *Science (1979)*, vol. 242, no. 4884, pp. 1385–1390, Dec. 1988, doi: 10.1126/science.242.4884.1385.
[163] R. Axelrod and W. D. Hamilton, 'The evolution of cooperation', *Science*, vol. 212, no. 4489. American Association for the Advancement of Science, pp. 1390–1396, 1981. doi: 10.1126/science.7466396.
[164] J. K. Rilling, A. G. Sanfey, J. A. Aronson, L. E. Nystrom, and J. D. Cohen, 'The neural correlates of theory of mind within interpersonal interactions', *Neuroimage*, vol. 22, no. 4, pp. 1694–1703, Aug. 2004, doi: 10.1016/j.neuroimage.2004.04.015.
[165] D. Fudenberg, 'The theory of learning in games', *Computers & Mathematics with Applications*, vol. 36, no. 6, p. 144, 1998, doi: 10.1016/s0898-1221(98)90111-0.
[166] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998. doi: 10.1017/CBO9781139173179.
[167] N. Bault, M. Joffily, A. Rustichini, and G. Coricelli, 'Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process', *Proc Natl Acad Sci U S A*, vol. 108, no. 38, pp. 16044–16049, Oct. 2011, doi: 10.1073/pnas.1100892108.
[168] R. C. Wilson and A. G. E. Collins, 'Ten simple rules for the computational modeling of behavioral data', *Elife*, vol. 8, pp. 1–33, 2019, doi: 10.7554/eLife.49547.
[169] J. P. O'Doherty, J. Cockburn, and W. M. Pauli, 'Learning, Reward, and Decision Making', *Annual Review of Psychology*, vol. 68. Annual Reviews, pp. 73–100, Jan. 04, 2017. doi: 10.1146/annurev-psych-010416-044216.
[170] J. Daunizeau and L. Rigoux, 'Bayesian sequence learning', *https://mbb-team.github.io/VBA-toolbox/wiki/BSL/*, 2014. https://mbb-team.github.io/VBA-toolbox/wiki/BSL/ (accessed Jan. 12, 2022).
[171] L. Breiman and L. Breiman, 'Statistical Modeling : The Two Cultures Statistical Modeling : The Two Cultures', Institute of Mathematical Statistics, Aug. 2019. [Online]. Available: http://www.jstor.org/stable/2676681
[172] K. P. Kording, G. Blohm, P. Schrater, and K. Kay, 'Appreciating the variety of goals in computational neuroscience', pp. 1–8, 2020, [Online]. Available: http://arxiv.org/abs/2002.03211
[173] M. R. Nassar and M. J. Frank, 'Taming the beast: Extracting generalizable knowledge from computational models of cognition', *Current Opinion in Behavioral Sciences*, vol. 11. Elsevier, pp. 49–54, Oct. 01, 2016. doi: 10.1016/j.cobeha.2016.04.003.
[174] S. Palminteri, V. Wyart, and E. Koechlin, 'The Importance of Falsification in Computational Cognitive Modeling', *Trends Cogn Sci*, vol. 21, no. 6, pp. 425–433, Jun. 2017, doi: 10.1016/j.tics.2017.03.011.
[175] S. Dunne and J. P. O'Doherty, 'Insights from the application of computational neuroimaging to social neuroscience', *Current Opinion in Neurobiology*, vol. 23, no. 3. Curr Opin Neurobiol, pp. 387–392, Jun. 2013. doi: 10.1016/j.conb.2013.02.007.
[176] M. R. Nassar, K. M. Rumsey, R. C. Wilson, K. Parikh, B. Heasly, and J. I. Gold, 'Rational regulation of learning dynamics by pupil-linked arousal systems', *Nat Neurosci*, vol. 15, no. 7, pp. 1040–1046, Jun. 2012, doi: 10.1038/nn.3130.
[177] K. Samejima, Y. Ueda, K. Doya, and M. Kimura, 'Neuroscience: Representation of action-specific reward values in the striatum', *Science (1979)*, vol. 310, no. 5752, pp. 1337–1340, Nov. 2005, doi: 10.1126/science.1115270.
[178] A. G. E. Collins and M. J. Frank, 'Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory', *Proc Natl Acad Sci U S A*, vol. 115, no. 10, pp. 2502–2507, Mar. 2018, doi: 10.1073/pnas.1720963115.
[179] J. X. O'Reilly, U. Schüffelgen, S. F. Cuell, T. E. J. Behrens, R. B. Mars, and M. F. S. Rushworth, 'Dissociable effects of surprise and model update in parietal and anterior cingulate cortex', *Proc Natl Acad Sci U S A*, vol. 110, no. 38, pp. E3660–E3669, Oct. 2013, doi: 10.1073/pnas.1305373110.
[180] J. F. Cavanagh, T. V. Wiecki, A. Kochar, and M. J. Frank, 'Eye tracking and pupillometry are indicators of dissociable latent decision processes', *J Exp Psychol Gen*, vol. 143, no. 4, pp. 1476–1488, 2014, doi: 10.1037/a0035813.
[181] M. Stuart, D. C. Hoaglin, F. Mosteller, and J. W. Tukey, 'Understanding Robust and Exploratory Data Analysis.', *The Statistician*, vol. 33, no. 3, p. 320, 1984, doi: 10.2307/2988240.
[182] O. Abdoun, 'Linear models Stats Course GitLab'. https://gitlab.com/ousabd/statscourse/-/blob/56b4cb56e6c645a9804368f65c1907f85250c760/1-05_linear-models/1-05.pdf (accessed Aug. 17, 2023).
[183] M. Nowak and K. Sigmund, 'A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game', *Nature*, vol. 364, no. 6432, pp. 56–58, 1993, doi: 10.1038/364056a0.
[184] R. A. Rescorla and A. Wagner, *A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement*, no. March. 1972. [Online]. Available: https://www.researchgate.net/publication/239030972_A_theory_of_Pavlovian_conditioning_The_effectiveness_of_reinforcement_and_non-reinforcement

[185] D. Lee, M. L. Conroy, B. P. McGreevy, and D. J. Barraclough, 'Reinforcement learning and decision making in monkeys during a competitive game', *Cognitive Brain Research*, vol. 22, no. 1, pp. 45–58, Dec. 2004, doi: 10.1016/j.cogbrainres.2004.07.007.

[186] L. Zhu, K. E. Mathewson, and M. Hsu, 'Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning', *Proc Natl Acad Sci U S A*, vol. 109, no. 5, pp. 1419–1424, 2012, doi: 10.1073/pnas.1116783109.

[187] I. Erev and A. E. Roth, 'Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria', *American Economic Review*, vol. 88, no. 4, pp. 848–881, 1998.

[188] G. Viejo, B. Girard, E. Procyk, and M. Khamassi, 'Adaptive coordination of working-memory and reinforcement learning in non-human primates performing a trial-and-error problem solving task', *Behavioural Brain Research*, vol. 355, pp. 76–89, Dec. 2018, doi: 10.1016/J.BBR.2017.09.030.

[189] C. Mathys, J. Daunizeau, K. J. Friston, and K. E. Stephan, 'A Bayesian foundation for individual learning under uncertainty', *Front Hum Neurosci*, vol. 5, no. MAY, p. 9, 2011, doi: 10.3389/fnhum.2011.00039.

[190] A. O. Diaconescu *et al.*, 'Inferring on the Intentions of Others by Hierarchical Bayesian Learning', *PLoS Comput Biol*, vol. 10, no. 9, p. 1003810, 2014, doi: 10.1371/journal.pcbi.1003810.

[191] M. R. Nassar, R. C. Wilson, B. Heasly, and J. I. Gold, 'An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment', *Journal of Neuroscience*, vol. 30, no. 37, pp. 12366–12378, Sep. 2010, doi: 10.1523/JNEUROSCI.0822-10.2010.

[192] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, J. O'Doherty, and G. Pezzulo, 'Active inference and learning', *Neuroscience and Biobehavioral Reviews*, vol. 68. Pergamon, pp. 862–879, Sep. 01, 2016. doi: 10.1016/j.neubiorev.2016.06.022.

[193] K. J. Friston, R. Rosch, T. Parr, C. Price, and H. Bowman, 'Deep temporal models and active inference', *Neuroscience and Biobehavioral Reviews*, vol. 77. Pergamon, pp. 388–402, Jun. 01, 2017. doi: 10.1016/j.neubiorev.2017.04.009.

[194] J. Jara-Ettinger, 'Theory of mind as inverse reinforcement learning', *Current Opinion in Behavioral Sciences*, vol. 29. Elsevier, pp. 105–110, Oct. 01, 2019. doi: 10.1016/j.cobeha.2019.04.010.

[195] S. Collette, W. M. Pauli, P. Bossaerts, and J. O'Doherty, 'Neural computations underlying inverse reinforcement learning in the human brain', *Elife*, vol. 6, Oct. 2017, doi: 10.7554/eLife.29718.

[196] K. Khalvati *et al.*, 'Modeling other minds: Bayesian inference explains human choices in group decision-making', *Sci Adv*, vol. 5, no. 11, Nov. 2019, doi: 10.1126/sciadv.aax8783.

[197] M. Devaine, G. Hollard, and J. Daunizeau, 'The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn?', *PLoS Comput Biol*, vol. 10, no. 12, 2014, doi: 10.1371/journal.pcbi.1003992.

[198] J. Daunizeau, H. E. M. den Ouden, M. Pessiglione, S. J. Kiebel, K. E. Stephan, and K. J. Friston, 'Observing the observer (I): Meta-bayesian models of learning and decision-making', *PLoS One*, vol. 5, no. 12, p. e15554, 2010, doi: 10.1371/journal.pone.0015554.

[199] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, 'Rational quantitative attribution of beliefs, desires and percepts in human mentalizing', *Nat Hum Behav*, vol. 1, no. 4, p. 64, 2017, doi: 10.1038/s41562-017-0064.

[200] S. N. Wood, 'Statistical inference for noisy nonlinear ecological dynamic systems', *Nature*, vol. 466, no. 7310, pp. 1102–1104, Aug. 2010, doi: 10.1038/nature09319.

[201] B. van Opheusden, L. Acerbi, and W. J. Ma, 'Unbiased and efficient log-likelihood estimation with inverse binomial sampling', *PLoS Comput Biol*, vol. 16, no. 12 December, p. e1008483, Dec. 2020, doi: 10.1371/journal.pcbi.1008483.

[202] M. Järvenpää, M. U. Gutmann, A. Pleska, A. Vehtari, and P. Marttinen, 'Efficient acquisition rules for model-based approximate Bayesian computation', *Bayesian Anal*, vol. 14, no. 2, pp. 595–622, Jun. 2019, doi: 10.1214/18-BA1121.

[203] A. Kangasrääsiö, J. P. P. Jokinen, A. Oulasvirta, A. Howes, and S. Kaski, 'Parameter Inference for Computational Cognitive Models with Approximate Bayesian Computation', *Cogn Sci*, vol. 43, no. 6, p. e12738, Jun. 2019, doi: 10.1111/cogs.12738.

[204] V. M. H. Ong, D. J. Nott, M. N. Tran, S. A. Sisson, and C. C. Drovandi, 'Variational Bayes with synthetic likelihood', *Stat Comput*, vol. 28, no. 4, pp. 971–988, Jul. 2018, doi: 10.1007/s11222-017-9773-3.

[205] M. Shinn, N. H. Lam, and J. D. Murray, 'A flexible framework for simulating and fitting generalized drift-diffusion models', *Elife*, vol. 9, pp. 1–27, Aug. 2020, doi: 10.7554/ELIFE.56938.

[206] A. Fengler, L. N. Govindarajan, T. Chen, and M. J. Frank, 'Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience', *Elife*, vol. 10, Apr. 2021, doi: 10.7554/ELIFE.65074.

[207] Q. Feltgen and J. Daunizeau, 'An Overcomplete Approach to Fitting Drift-Diffusion Decision Models to Trial-By-Trial Data', *Front Artif Intell*, vol. 4, p. 531316, Apr. 2021, doi: 10.3389/frai.2021.531316.

[208] J. Boelts, J. M. Lueckmann, R. Gao, and J. H. Macke, 'Flexible and efficient simulation-based inference for models of decision-making', *Elife*, vol. 11, Jul. 2022, doi: 10.7554/ELIFE.77220.

[209] B. M. Turner and T. Van Zandt, 'Approximating Bayesian Inference through Model Simulation', *Trends in Cognitive Sciences*, vol. 22, no. 9. Elsevier, pp. 826–840, Sep. 01, 2018. doi: 10.1016/j.tics.2018.06.003.

[210] K. Cranmer, J. Brehmer, and G. Louppe, 'The frontier of simulation-based inference', *Proc Natl Acad Sci U S A*, vol. 117, no. 48, pp. 30055–30062, Dec. 2020, doi: 10.1073/pnas.1912789117.

[211] G. E. P. Box, 'Science and statistics', *J Am Stat Assoc*, vol. 71, no. 356, pp. 791–799, 1976, doi: 10.1080/01621459.1976.10480949.

[212] H. Akaike, 'A New Look at the Statistical Model Identification', *IEEE Trans Automat Contr*, vol. 19, no. 6, pp. 716–723, 1974, doi: 10.1109/TAC.1974.1100705.

[213] P. Stoica and Y. Selén, 'A review of information criterion rules', *IEEE Signal Process Mag*, vol. 21, no. 4, pp. 36–47, 2004, doi: 10.1109/MSP.2004.1311138.

[214] G. Schwarz, 'Estimating the Dimension of a Model', *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, Mar. 2007, doi: 10.1214/aos/1176344136.

[215] J. Bruineberg, J. Kiverstein, and E. Rietveld, 'The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective', *Synthese*, vol. 195, no. 6, pp. 2417–2444, Jun. 2018, doi: 10.1007/s11229-016-1239-1.

[216] K. J. Friston and K. E. Stephan, 'Free-energy and the brain', *Synthese*, vol. 159, no. 3, pp. 417–458, Dec. 2007, doi: 10.1007/s11229-007-9237-y.

[217] M. D. Lee and E. J. Wagenmakers, *Bayesian cognitive modeling: A practical course*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139087759.

[218] P. Piray, A. Dezfouli, T. Heskes, M. J. Frank, and N. D. Daw, 'Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies', *PLoS Comput Biol*, vol. 15, no. 6, p. e1007043, Jun. 2019, doi: 10.1371/journal.pcbi.1007043.

[219] J. Daunizeau and L. Rigoux, 'The variational Bayes approach', 2014. https://mbb-team.github.io/VBA-toolbox/wiki/The-variational-Bayesian-approach/ (accessed Jan. 12, 2022).

[220] K. H. Brodersen, J. Daunizeau, C. Mathys, J. R. Chumbley, J. M. Buhmann, and K. E. Stephan, 'Variational Bayesian mixed-effects inference for classification studies', *Neuroimage*, vol. 76, pp. 345–361, 2013, doi: 10.1016/j.neuroimage.2013.03.008.

[221] J. Daunizeau, K. J. Friston, and S. J. Kiebel, 'Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models', *Physica D*, vol. 238, no. 21, pp. 2089–2118, Nov. 2009, doi: 10.1016/j.physd.2009.08.002.

[222] N. D. Daw, 'Trial-by-trial data analysis using computational models', in *Decision Making, Affect, and Learning: Attention and Performance XXIII*, Oxford University Press, 2011. doi: 10.1093/acprof:oso/9780199600434.003.0001.

[223] S. J. Gershman, 'Empirical priors for reinforcement learning models', *J Math Psychol*, vol. 71, pp. 1–6, Apr. 2016, doi: 10.1016/j.jmp.2016.01.006.

[224] K. Katahira, 'How hierarchical models improve point estimates of model parameters at the individual level', *J Math Psychol*, vol. 73, pp. 37–58, Aug. 2016, doi: 10.1016/j.jmp.2016.03.007.

[225] J. Daunizeau, V. Adam, and L. Rigoux, 'VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data', *PLoS Comput Biol*, vol. 10, no. 1, p. 1003441, 2014, doi: 10.1371/journal.pcbi.1003441.

[226] K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, and K. J. Friston, 'Bayesian model selection for group studies', *Neuroimage*, vol. 46, no. 4, pp. 1004–1017, Jul. 2009, doi: 10.1016/j.neuroimage.2009.03.025.

[227] L. Rigoux, K. E. Stephan, K. J. Friston, and J. Daunizeau, 'Bayesian model selection for group studies - revisited', *Neuroimage*, vol. 84, pp. 971–985, Jan. 2014, doi: 10.1016/J.NEUROIMAGE.2013.08.065.

[228] D. J. Navarro, 'Between the Devil and the Deep Blue Sea: Tensions Between Scientific Judgement and Statistical Model Selection', *Comput Brain Behav*, vol. 2, no. 1, pp. 28–34, Mar. 2019, doi: 10.1007/S42113-018-0019-Z/FIGURES/2.

[229] H. Nilsson, J. Rieskamp, and E. J. Wagenmakers, 'Hierarchical Bayesian parameter estimation for cumulative prospect theory', *J Math Psychol*, vol. 55, no. 1, pp. 84–93, Feb. 2011, doi: 10.1016/J.JMP.2010.08.006.

[230] F. Castelli, F. Happé, U. Frith, and C. Frith, 'Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns', *Neuroimage*, vol. 12, no. 3, pp. 314–325, 2000, doi: 10.1006/nimg.2000.0612.

[231] M. K. Ho, R. Saxe, and F. Cushman, 'Planning with Theory of Mind', *Trends Cogn Sci*, vol. 26, no. 11, pp. 959–971, 2022, doi: 10.1016/J.TICS.2022.08.003.

[232] M. Devaine, G. Hollard, and J. Daunizeau, 'Theory of mind: Did evolution fool us?', *PLoS One*, vol. 9, no. 2, 2014, doi: 10.1371/journal.pone.0087619.

CHAPTER 2: Main results

This chapter comprises the 3 studies that constitute the backbone of my PhD thesis. These 3 studies are :

A. Differences in computational decision processes between human and macaque in a transparent Bach-or-Stravinsky Game.

B. Mentalising underlies strategic coordination in Guinea baboons (Papio papio)

C. Computational mechanisms underlying the emergence of theory of mind in children

A. Study 1

Differences in computational decision processes between human and macaque in a transparent Bach-or-Stravinsky Game.

Toan Nong[1,2,*], Sebastian Möller[4,5,*], Rémi Philippe[1,2], Edmund Derrington[1,2], Igor Kagan[4,5,*] & Jean-Claude Dreher[1,2,*]

1- *ISC-MJ, CNRS UMR5229, Neuroeconomics Lab, Bron, France*
2- *Lyon Univ UCBL1, Lyon, France*
3- *Leibniz Science Campus Primate Cognition, Goettingen, Germany*
4- *German Primate Center—Leibniz Institute for Primate Research, Goettingen, Germany*

Abstract

The emergence of joint coordinated action in nonhuman primates constitutes a primordial and distinctive aspect of social cognition. Learning and maintaining coordination for mutual benefit provided an evolutionary advantage in the wild to social species of primates such as great apes (e.g. chimpanzees, bonobos…), or old-world monkeys (e.g. baboons, macaques…). However, little is known about the decision process and learning mechanism that lead to simultaneous coordination, and how they differ between humans and monkeys. In particular, coordinated actions for achieving a joint goal can become more difficult when conflicting personal interests are involved. Here, we investigated whether and how rhesus macaques and adult humans solved a simultaneous transparent coordination game with inherent conflicting rules for resource repartition. We presented 9 pairs of rhesus macaques and 29 pairs of adult humans with a transparent iterated Bach-or-Stravinsky (BoS) game on a computerized shared transparent touchscreen. During the joint practice of the game, some participants exhibited spontaneous turn-taking coordination to achieve equal repartition of resources, while others displayed a despotic coordination where one player benefitted from more resources than the other. Interestingly, some pairs also achieved equal repartition of resources and optimal gains by statically coordinating on one side of the touchscreen. The

computational mechanism that best approximated humans' behavior exploited the visual cues of the nature of the partner's current choice through the transparent setup. In contrast, macaques applied a strategy that also exploited the screen transparency, but based on cues of the location of the partner's current choice. (247/250 words)

Introduction (1216 / ~1200 words)

Coordination emerged through evolution in many social species. In social environments, learning to coordinate with your own kind is essential to survive. Coordination in some situations might be described as an innate and very low-level embedded behavior, e.g. in swarming bees in search of a new home must coordinate their movements to decide where to go and where to settle [1]. This ability can also be the result of higher-level decision processes that require both social learning and interaction. For example, primates, such as chimpanzees, can successfully coordinate by communicating in problem-solving tasks [2], and, like humans, also do so in conflict situations [3].

The concept of transparent 2x2 (two players, two options) games where players can observe each other's actions was introduced to study the influence of the visibility of the other agent's actions on the player's decision process [4], [5]. In contrast to classic simultaneous or sequential games, in such game-theoretic settings, the access to the information about the current choice of the other agent is probabilistic. Indeed, in each round, there are three possible cases:

1. Player 1 knows the choice of Player 2 before making its own choice.
2. Player 2 knows the choice of Player 1 before making its own choice.
3. Neither player knows the choice of the partner.

This novel framework of game theory is perfect to study the effects of transparency on the emergence of cooperation in iterated decision games. In this particular context, purely cooperative games like the Coordination Game make less sense, as the transparency of the setup makes the evolution of such iterated games trivial because of the visual access to information about the partner's choice in the current round. Therefore, an interesting game to study with this transparent iterative setup is the Battle of the Sexes, or Bach-or-Stravinsky game (BoS) [6]. In this game each agent has a different "preferred" option, in the sense that they are more rewarded if they choose that option. However, coordination on the same option by both agents is rewarded by a bonus reward. The bonus is distributed equally to both players so that any coordinated choice is rationally better than no coordination for both agents. Nevertheless, one coordinated choice is better for the first agent and the other is better for the second. This creates an inherent conflict about who profits the most. Games that combine cooperation and conflict offer interesting opportunities to study social interactions in primates. For example, games similar to the BoS game reveal that humans try to ensure fairness by switching between the two coordinated options [7], [8]. This trial-by-trial turn-taking was observed in 5 years old, but neither in 3-year-old children, nor in chimpanzees, in a rope-pulling task that required cooperation but rewards only one agent at a time, [9]. This raises the question whether humans use unique decision processes to achieve coordination.

In the context of strategic dyadic decision games, different computational mechanisms may account for the emergence of coordination in primates. A first possibility is that their learning and decision processes are explained by reinforcement learning (RL) strategies [10]. These strategies compute and regularly update a decision value that is derived from a prediction error. The prediction error corresponds to the difference between the predicted value of the model and the actual feedback [10]. A second, more sophisticated, candidate model is the influence model [11], [12] that not only employs RL mechanisms, but also accounts for a mentalizing process that computes the influence of one's choice on another's decision. A third popular theory is that the brain performs Bayesian inference based on observations using probabilistic representations of the world, and uses the results of Bayesian inference to choose optimal actions [13], [14]. Thus, primates might learn and decide based on purely Bayesian

learning strategies. Alternatively, an even more simple and straightforward strategy would be to use simple heuristics to make decisions, such as the Win-Stay/Lose-Switch strategy, that consists of persisting with one's choices when winning, and switching choices after a loss [15].

Finally, a promising computational model was proposed to provide a formal probabilistic framework for solving tasks involving action selection and decision making under uncertainty: Partially Observable Markovian Decision Processes (POMDPs). A rigorous theoretical framework previously proposed by Rao[16] assumes that the brain performs Bayesian inference based on observations using probabilistic representations of the world and utilizes the results of Bayesian inference to choose optimal actions. A main hypothesis of this framework is that the brain performs Bayesian inferences using probabilistic representations of not only environmental states but also other individuals' intentions and behavior. The probabilistic framework of POMDPs can be used to model ways in which humans might predict hypothetical action outcomes, the intentions of others, and whether the other is cooperative or competitive. In POMDP models, the agent maintains probabilistic beliefs about hidden states of the world using Bayesian inference based on the sequence of his observations and actions. The state dynamics are supposed to be Markovian (i.e. the next state only depends on the current one), but the agent cannot directly observe the state and must instead use its sensors to make observations and maintain a probability distribution over the set of possible states, based on the observations, observation probabilities, and the underlying dynamics. It is however still unknown whether humans are the only primates that are able to perform such inferences and computations.

We therefore compared the behavior of adult humans and rhesus macaques in a transparent, iterative BoS game. To facilitate inter-species comparability, human or macaque agents were paired without explicit task instructions. We expected that the transparency and instantaneous coaction [17] would facilitate efficient coordination in both species. This should translate, in terms of computational models, into a better fit of computational learning models that actively use the visual cues accessible to the agent in the current trial, namely POMDP or 2QL. Agents that use these decision processes should also be more efficiently coordinating and therefore overall more rewarded. However, if the players (almost) did not use any visual information accessible to them, any other computational model that does not use this information would fit better. Overall, agents using these decision processes should most likely be less rewarded than the agents that used decision processes that exploited the visual information about the other agent's current action.

Based on their cognitive abilities [18], propensity for fairness [19] and perspective taking [20], humans were expected to engage in pro-social turn-taking, where the reward maximization alternated between partners. We expected that this turn-taking would be achieved by actively observing the visual access to their partner's actions with a POMDP process. As for macaques, we predicted that they would also take advantage of the transparency to coordinate more efficiently, but that they would be less likely to grant an advantage to the other player. They would instead coordinate to increase individual (and joint) rewards, but without turn-taking to balance rewards between players. We expected that this would translate into a different computational learning process that also exploits visual cues, such as the 2QL model (see models' descriptions in Methods). Alternatively, we could also find a POMDP model to fit their decision process, but with a significant difference in their learning parameters. A less likely option would be that they don't use the transparent setup in their decision process.

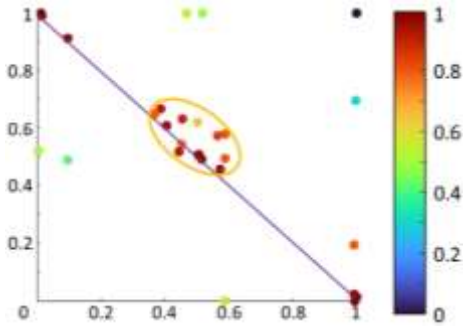
Results

Pairs of players converge towards higher rates of coordination. Theoretically, in the BoS game, the reward of the pair is maximized when both players choose the same target. In our implementation of the BoS game, players can pay attention to the other player who is facing

them to coordinate on the same target and thus profit from the coordination bonus of 2 points. In an iterative version of the game, a pair of players can be described over one session by the ratio of each player's choice preference ratio, i.e., the number of times they chose their preferred option over the total number of trials, and by the players' coordination ratio, i.e. the proportion of trials in which both players coordinated on the same option. Under this representation, the optimal states in terms of pairwise reward are those that maximize coordination between players: from choosing player A's preferred choice 100% of the time to choosing player B's preferred choice 100% of the time, including all possible intermediate states, which constitutes a diagonal where 100% coordination is achievable. We expected most participants to reach states close to the optimal, with coordination ratios close to 100%, at the end of the experiment. We represented the diagrams of the sessions of humans, and of some of the macaques in different conditions on Figure 18.

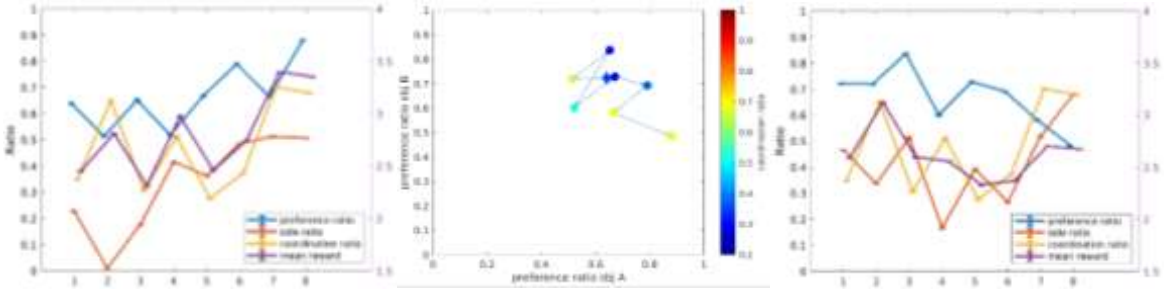
Given that none of the participants were explicitly told the rules of the game, they had to explore their options in the game across trials to discover the best course of actions. In most humans, this was achieved within one session of a few hundred trials. In macaques, it was more uncertain whether they actually understood the rules of the BoS. They took thousands of trials to reach higher coordination ratios, which was achieved after a dozen or more game sessions (naïve condition). None of the naïve macaque pairs converged to dynamic turn-taking typical of many of the human pairs. This was why some macaques were subjected to a training phase with a human confederate (confederate-training condition), and then, after having been trained, played once again with another conspecific that also underwent confederate-training (Trained condition).

(a) Naïve humans:

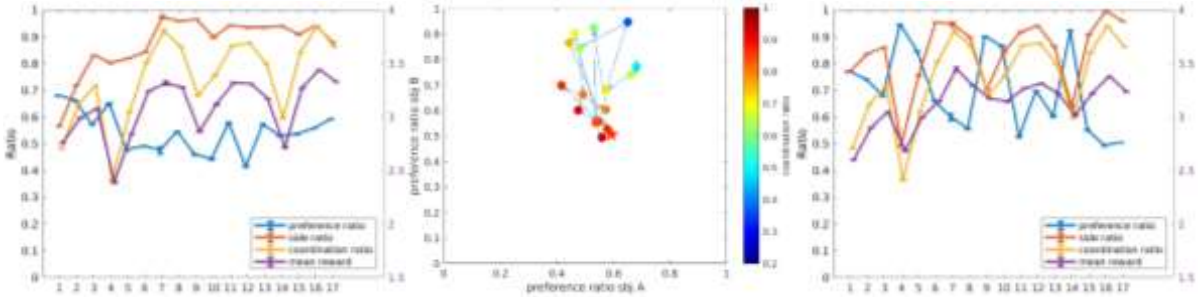


Naïve Macaques:

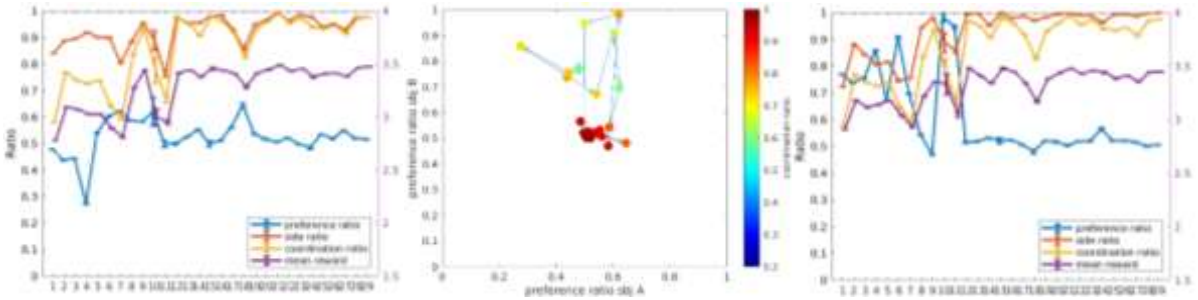
(b) (19/10/2017 - 07/11/2017) 3A: Flaffus (1st) ; 3B: Curius (1st bis):



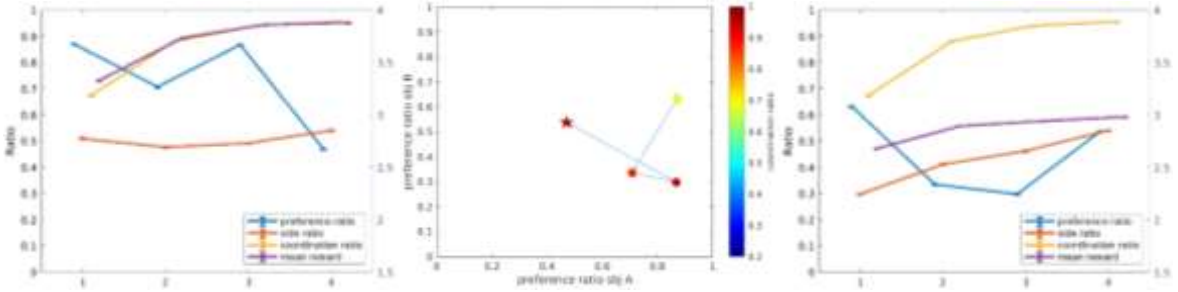
(c) (17/08/2017 - 14/02/2018) 5A: Magnus (1st bis) ; 5B: Curius (1st bis):



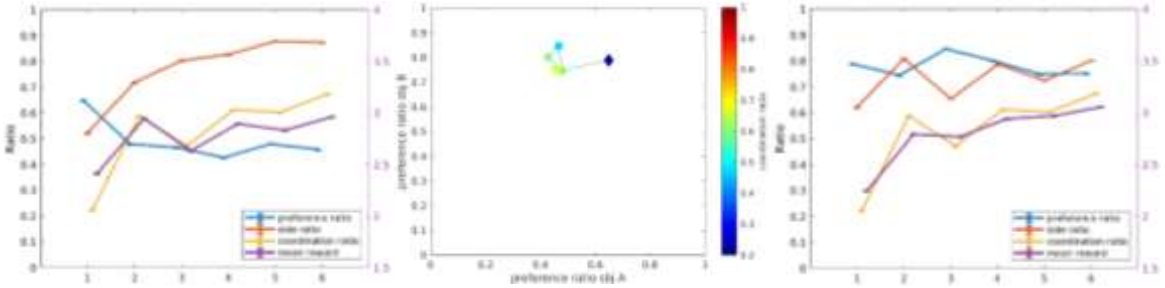
(d) (29/11/2017 - 25/01/2018) 6A: Magnus (1st bis) ; 6B: Flaffus (2nd):



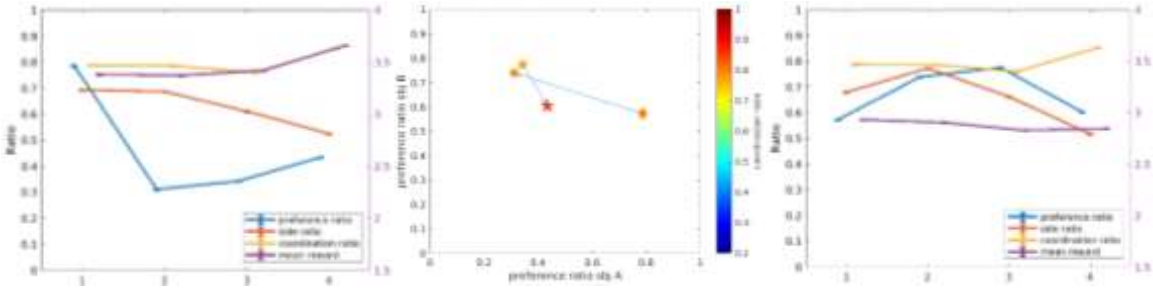
(e) (04/05/2018 - 09/05/2018) 9A: Tesla (1st) ; 9B: Flaffus (3rd):



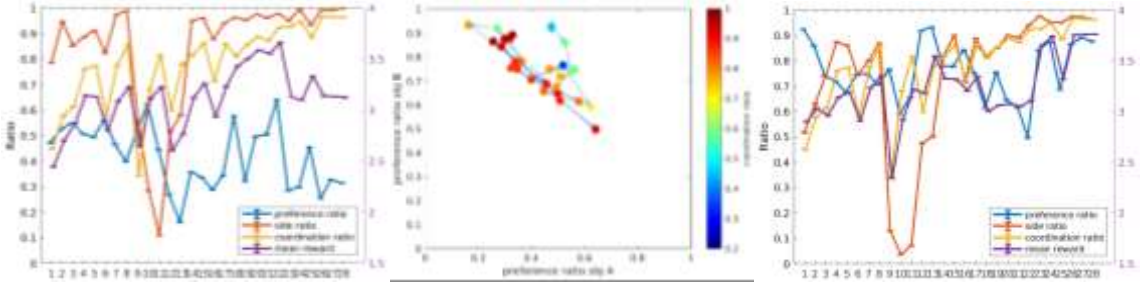
(f) (16/05/2018 - 24/05/2018) 7A: Tesla (2nd) ; 7B: Elmo (1st):



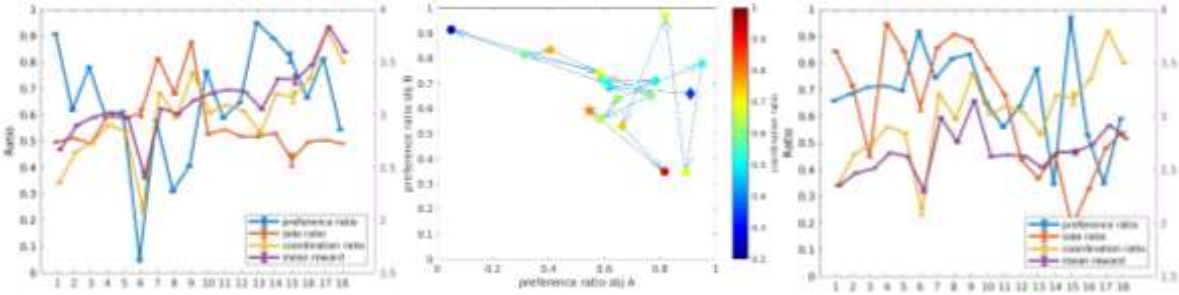
(g) (25/05/2018 - 31/05/2018) 8A: Tesla (3rd) ; 8B: Curius (3rd):



(h) (05/06/2018 - 12/10/2018) 1A: Curius (4th) ; 1B: Elmo (2nd):



(i) (23/10/2018 - 20/11/2018) 4A: Linus (1st) ; 4B: Elmo (3rd):



(j) (27/11/2018 - 11/12/2018) 2A: Curius (5th) ; 2B: Linus (2nd):

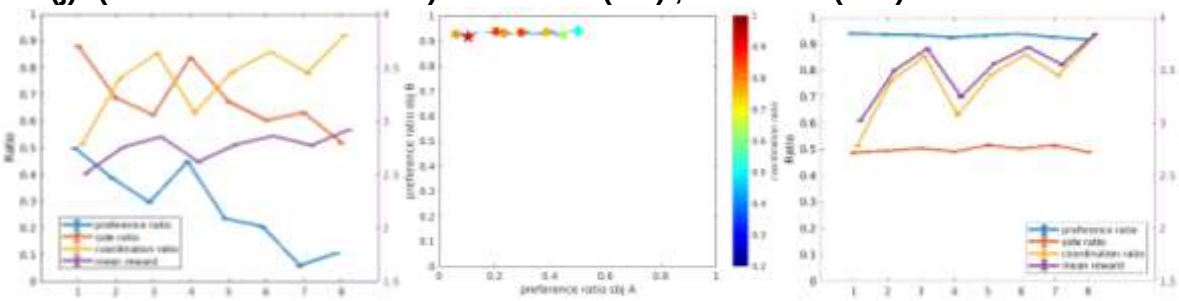


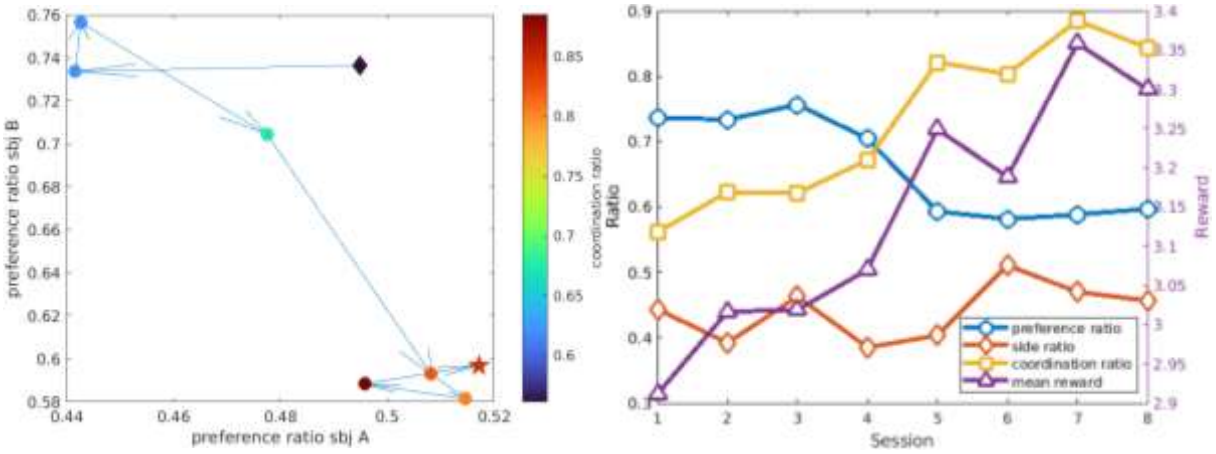
Figure 18: ***Sum-up metrics over sessions of naïve humans and macaques. (a) Naïve Human sessions.*** *Each dot represents a session of two players. The purple line represents where the theoretically optimal sessions for the whole pair are, i.e. sessions in which reward is maximized for the pair of players. In such sessions, players always coordinate on a given option at different possible ratios. The yellow ellipsoid circles sessions where the players chose relatively evenly between their preferred option and the other's preferred option. X-axis: Player A's tendency to choose their preferred choice. Y-axis: Player B's tendency to choose their preferred choice. Colored-axis: mean coordination ratio between the players within the session represented by the dot.* ***(b)-(j) Sessions for pairs 1 to 9 of naïve macaques.*** Left*: mean metrics (preferred choice, side & coordination ratios, and reward) for Player A. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.* Middle*: Evolution of mean preference ratios over sessions. Each dot represents a session of the same pair of subjects. The arrows between dots represent the order of succession of the sessions with time. The diamond represents the first session. The star represents the last session.* Right*: mean metrics (preferred choice, side & coordination ratios, and reward) for Player B. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.*

We observed in humans that 21/29 pairs achieved 60% or more coordination over the whole session. These are indicated by the dots that were the closest to the “optimal diagonal”. Among them, 14 pairs were situated around the centre of the diagonal, which means that both players chose almost as frequently their preferred choice and the preferred choice of the other player. This means that they achieved a relatively fair consensus, i.e. long-term coordination agreement, that minimizes the difference in mean reward between the two players over the session. As for macaques, depending on the individual and with whom they were paired, they could sometimes achieve high coordination ratios in the late sessions. They also went through multiple areas on the preference/coordination ratios diagram, sometimes going through unfair consensus on one side or the other of the spectrum. Some of them also achieved fair consensus at the centre of the diagram with a high coordination ratio (pairs 5, 6, 9). Others achieved high consensus in an unfair consensus (pair 2). Some achieved high coordination, but went through different phases (pair 1). And some did not achieve very high consensus consistently even after multiple sessions (pairs 3, 7, 8).

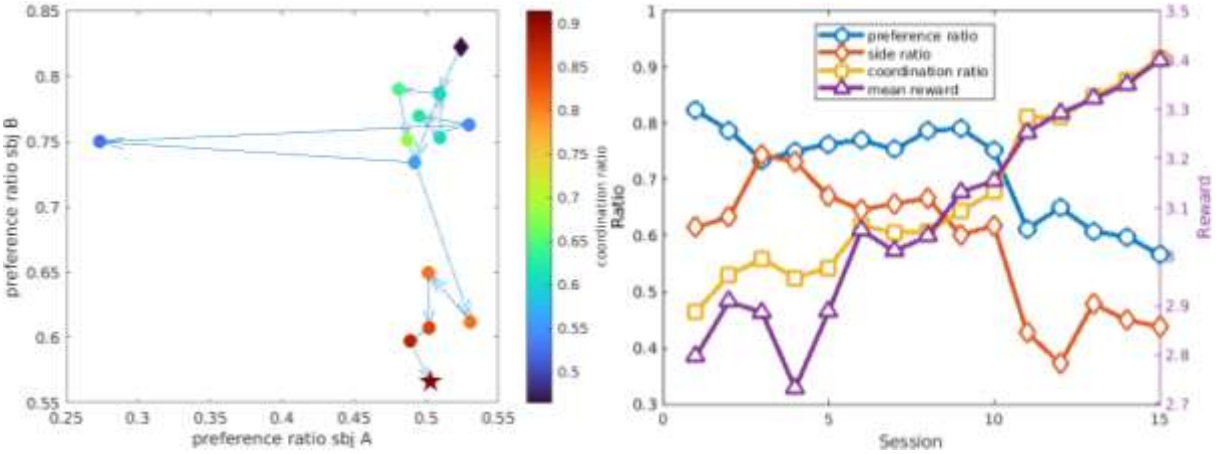
Even after multiple game sessions, the majority of macaques’ decision strategies remained very distinct from humans’. For example, some pairs achieved high coordination rates, rewards, and fairness of reward distribution by focusing statically on the same side of the display (pairs 1, 5, 6, 7). Consequently, they would always benefit from the coordination bonus (since individually-preferred target position was randomized, focusing on only one side would on average result in coordinating on the two targets with almost equal frequency, leading to a fair repartition of rewards). This somewhat elegant strategy (at least from an external standpoint) maximized the reward for the pair and the fairness of the repartition of the reward, as well as minimizing the effort since there was no need to track the target identity. Other macaque pairs managed to reach high coordination ratios when one of the players chose their preferred target almost all the time, while the other complied and followed to that target (pairs 1, 2, 9). If both behaviours are valid strategies, a question remains: are macaques able to actively alternate between one’s preferred choice, then the other’s? Therefore, a Training phase was performed to check whether macaques could actually actively coordinate with the other player, a human confederate. This was then followed by a Trained phase, where the trained macaques performed the task with each other.

In confederate-training:

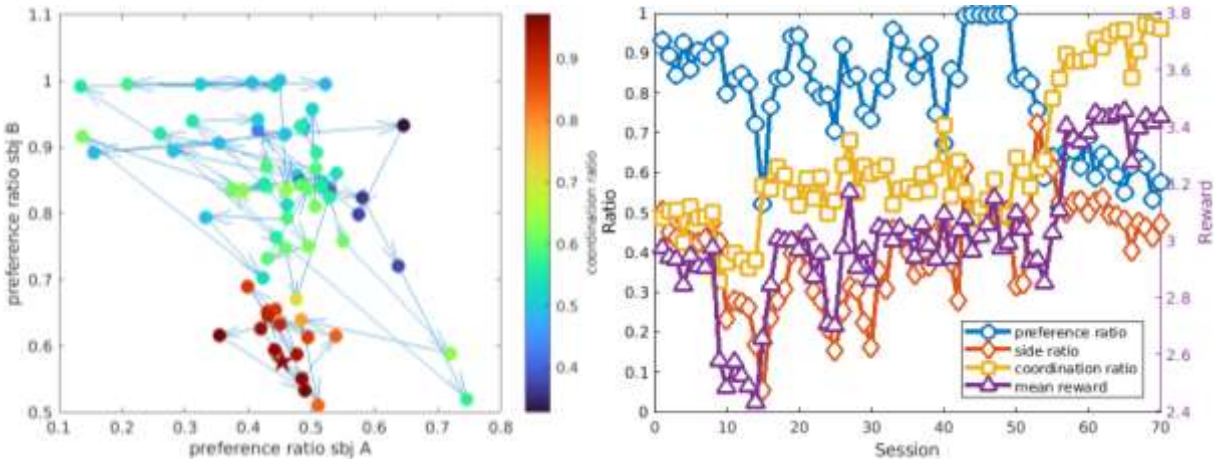
(a) 1B: Curius



(b) 2B: Flaffus



(c) 3B: Linus



(d) 4A: Elmo

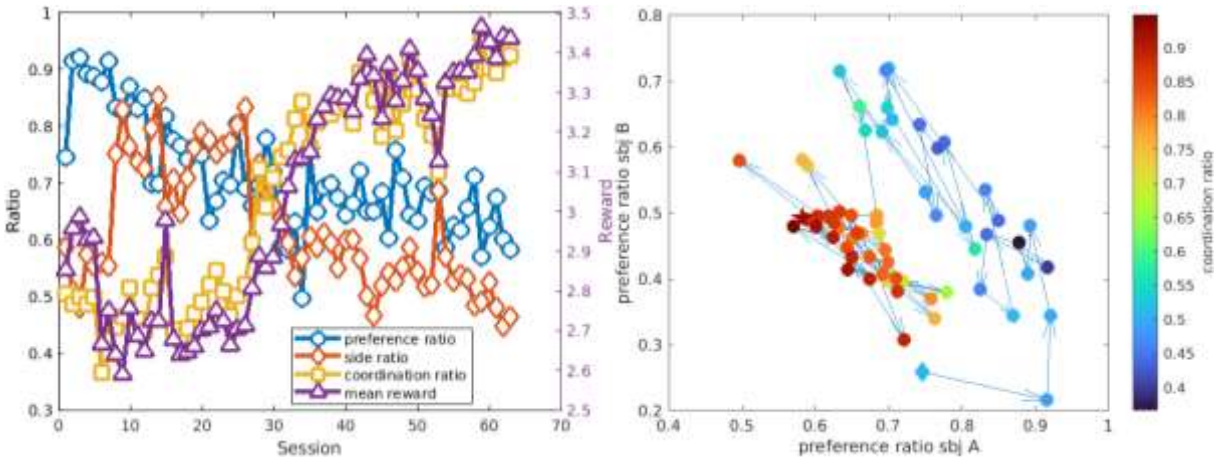


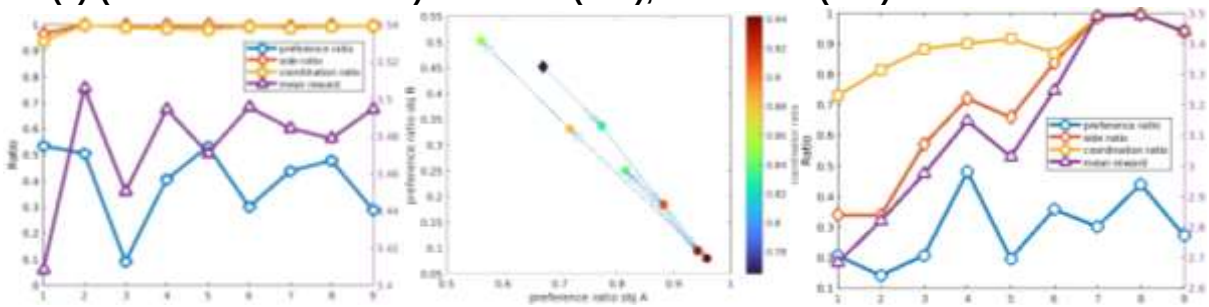
Figure 19: ***Sum-up metrics of Training sessions of macaques. (a)-(c) Left****: Evolution of mean preference ratios over sessions. Each dot represents a session of the same pair of subjects. The arrows between dots represent the order of succession of the sessions with time. The diamond represents the first session. The star represents the last session. Color-scale: The color of the dots represents the mean coordination ratio of the macaques within the session represented by the dot.* ***Right****: mean metrics (preferred choice, side & coordination ratios, and reward) for Player A. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.* ***Right****: Evolution of mean preference ratios over sessions. Each dot represents a session of the same pair of subjects. The arrows between dots represent the order of succession of the sessions with time. The diamond represents the first session. The star represents the last session. Color-scale: The color of the dots represents the mean coordination ratio of the macaques within the session represented by*

the dot. Right*: mean metrics (preferred choice, side & coordination ratios, and reward) for Player B. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.* ***(d) Left****: mean metrics (preferred choice, side & coordination ratios, and reward) for Player A. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.* ***Right****: Evolution of mean preference ratios over sessions. Each dot represents a session of the same pair of subjects. The arrows between dots represent the order of succession of the sessions with time. The diamond represents the first session. The star represents the last session. Color-scale: The color of the dots represents the mean coordination ratio of the macaques within the session represented by the dot.* Right*: mean metrics (preferred choice, side & coordination ratios, and reward) for Player B. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.*

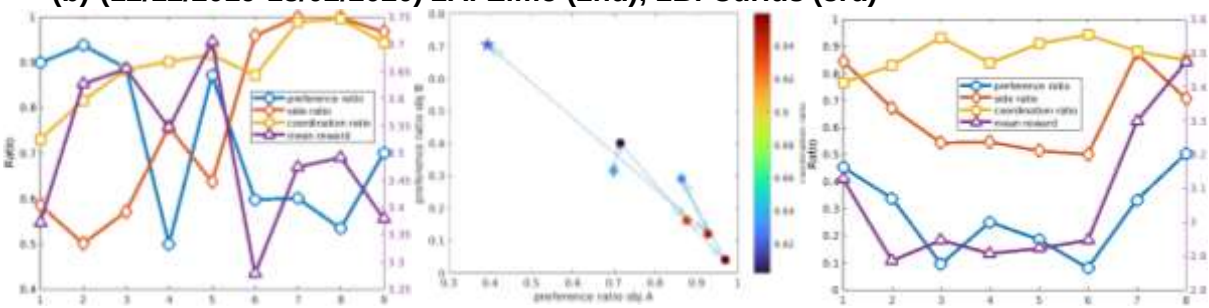
Four macaques underwent a training phase with a human confederate to make sure that they were actually able to observe the choices of the other player and actively coordinate with them. The human confederate performed stereotypical choices which required that the macaque observed or at least took into account the choices of the confederate to maximize their reward. The training lasted until the macaque could consistently display active coordination with the human confederate by switching from preferred to non-preferred in conjunction with the human confederate when needed, and vice-versa. At the beginning, for all macaques, the coordination ratio over their sessions was quite low: close to 50%, usually because the macaques actively followed or happened to choose the choice of the human confederate when it was their preferred choice, which was the case around 50% of the time (Figure 19). This was because the human confederate stereotypically chose in blocks either the preferred choice of the monkey or the non-preferred choice of the monkey, almost equally in proportion (at the beginning). After varying their alternation pattern of the confederate's choices depending on the coordination ability of the macaque, the macaques progressively, in more or less numerous sessions, improved their coordination ratio until reaching almost perfect coordination.

Confederate-trained Macaques:

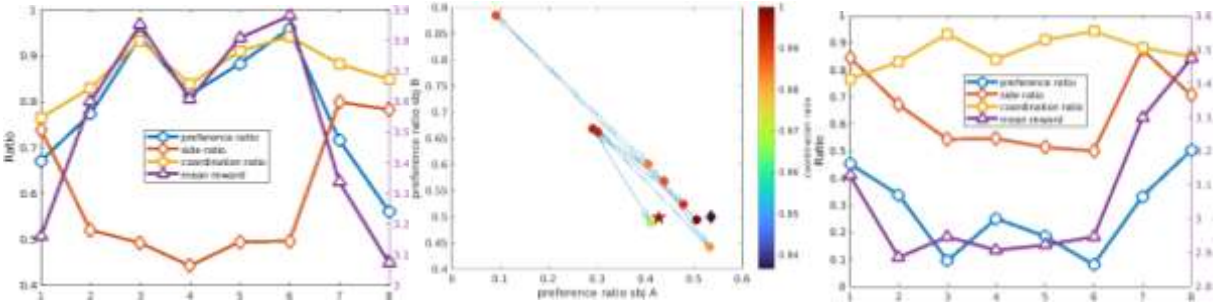
(a) (13/11/2019-27/11/2019) 1A: Linus (2nd); 1B: Curius (2nd)



(b) (12/12/2019-18/02/2020) 2A: Elmo (2nd); 2B: Curius (3rd)



(c) (28/08/2019-27/09/2019) 3A: Elmo (1st); 3B: Linus (1st)



(d) (18/04/2018-01/05/2018) 4A: Flaffus (1st); 4B: Curius (1st)

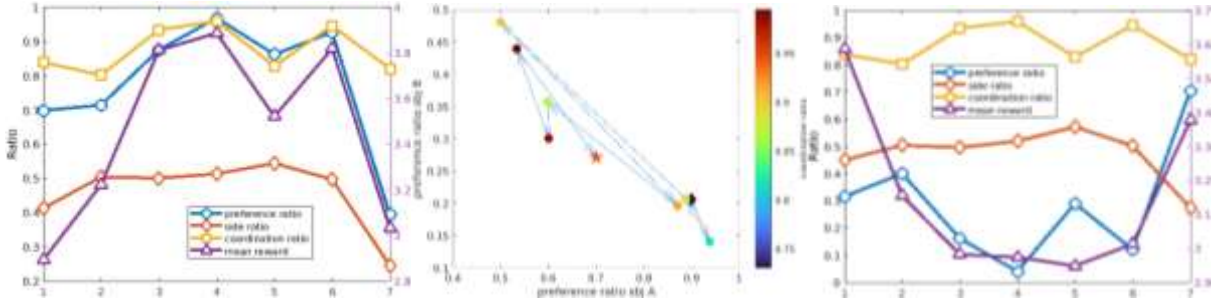


Figure 20: ***Sum-up metrics of sessions of confederate-trained macaques. (a)-(d) Left****: mean metrics (preferred choice, side & coordination ratios, and reward) for Player A. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.* ***Middle****: Evolution of mean preference ratios over sessions. Each dot represents a session of the same pair of subjects. The arrows between dots represent the order of succession of the sessions with time. The diamond represents the first session. The star represents the last session. Color-scale: The color of the dots represents the mean coordination ratio of the macaques within the session represented by the dot.* Right*: mean metrics (preferred choice, side & coordination ratios, and reward) for Player B. X-axis: session number. Y-axis: mean preferred choice ratio, mean side ratio, coordination ratio, and mean reward in each session.*

After the Training phase, the four macaques were paired with a conspecific and played the game again (Figure 20). Four pairs were constructed in this way. The macaques that went through training all played at least once with another trained monkey in this Trained condition. Their coordination rate was high from the first session. Some pairs displayed alternation of target choices with high coordination ratios (pairs 2,3,4), while some pairs applied a simple strategy of coordinating by selecting the same side as their partner, always the same side across the session (pairs 1, 2).

Difference in decision times suggests active tracking of intentions in both humans and rhesus macaques.

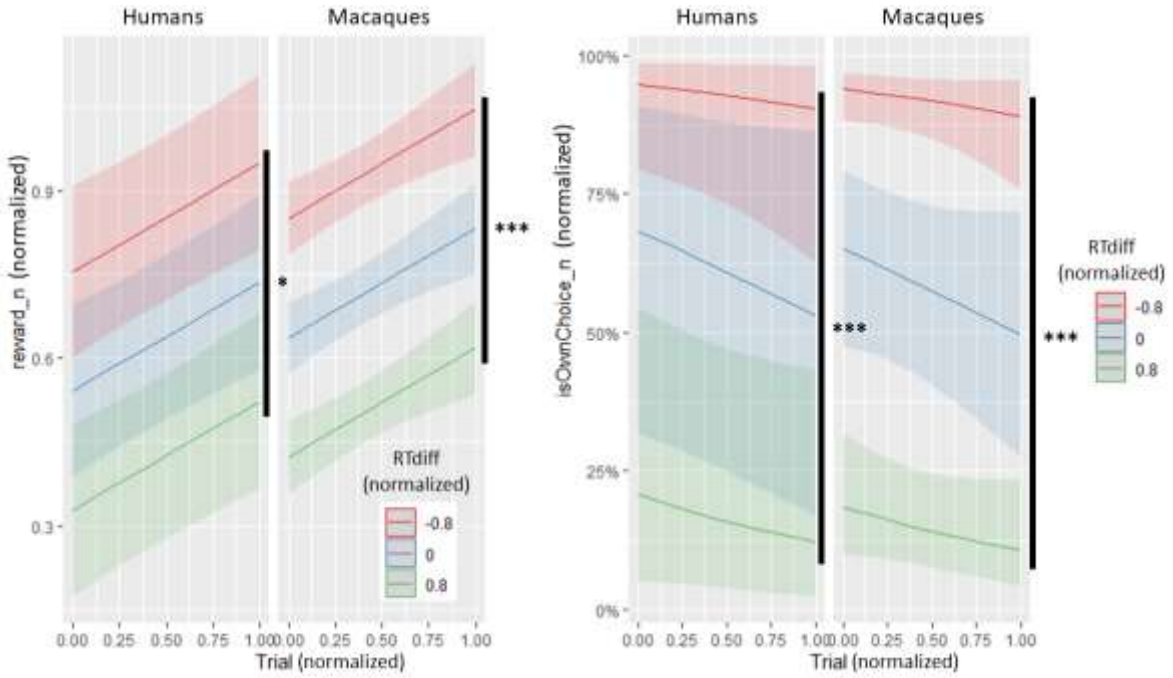


Figure 21: ***Behavioral results and correlations for naïve humans and macaques. Generalized Mixed Effects (GLME) models, marginal effects.*** *Left: predicted reward (normalized) of the subjects at trial n, reward$_n$, plotted against the normalized trial number (x-axis), for different Decision Time differences. Right: predicted probability for the subjects to choose their own target at trial n, isOwnChoice$_n$ (right), plotted against the normalized trial number (x-axis), for different Decision Times differences.*

To get a better idea of the factors that drove the decision process of both humans and macaques within a session, we conducted a correlational analysis between different regressors that could potentially predict the behavior of the participants. To do so, we used Generalized Linear Mixed-Effects (GLME) models with the lme4 package in R. We controlled for the statistical effects of each variable, ensured that we had good estimates of random effects, and accounted for variability in behavior using Nelder-Mead Optimization (see Methods). GLME1 assessed the participant's tendency to select their preferred target ("isOwnChoice_n") and GLME2 assessed their performance, normalized between 0 and 1 instead of 1 and 4 ("reward_n"). The regressors that were tested are: the reward in the previous trials (k-1, k-2, k-3), tendency to switch between targets chosen in the previous trials (k-1, k-2, k-3), tendency to choose one's preferred target in the previous trials (k-1, k-2, k-3), the trial number, the decision time difference between the players, the session number, the condition (Naïve, In Training or Trained) and the interactions between session, condition and trial.

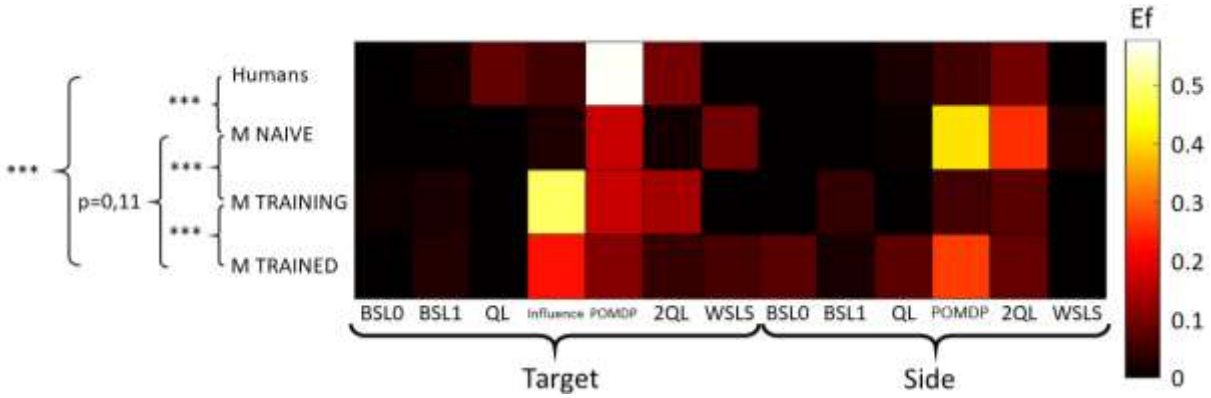
We found that the results of the previous three trials, i.e. the previous choices, outcomes and tendency to switch in the previous trials significantly predicted the decision to choose one's preferred target rather than the other's (see Supp.Tab.1). They also significantly predicted the amount of reward that the participant would get in the current trial (see Supp.Tab.1). Participants' reward improved with time, and their tendency to choose their preferred choice decreased with time within a same session on average (see Figure 21). This suggests that participants, on average, tended to learn and improve their reward with practice by accepting to choose their preferred target less frequently, but rather their partner's, to favor coordination.

Interestingly, the decision time difference (DTdiff) between the 2 players also significantly predicted participants' reward at trial n, and their tendency to choose their preferred target rather than the other's (Figure 21). A negative decision time difference meant that the participant played faster than their partner. A positive decision time indicates that they were

slower to decide. Decision time differences close to 0 meant that they chose their target almost at the same time. We found that for a negative DTdiff, the higher the absolute value, the more likely the participant was to choose their own preferred target and the higher their reward on average. On the contrary, for a positive DTdiff, the higher the absolute value, the less likely they were to choose their preferred target, and the lower was their reward on average. This suggests that participants tended to compete in speed to impose their preferred choice on the other player so that coordination would result in a higher reward for themself. A slower decision time would often result in choosing the other's preferred target to coordinate and get the bonus reward, which implies that participants indeed observed the other player's current decision. This suggests that the learning and decision process of the participants also incorporates observation of the other player's behavior. All of these results were common to both species, macaques and humans.

Computational models of the decision strategy of humans vs macaques: different learning processes between species and conditions.

(a) All naïve sessions:



(b) 4 last naïve sessions:

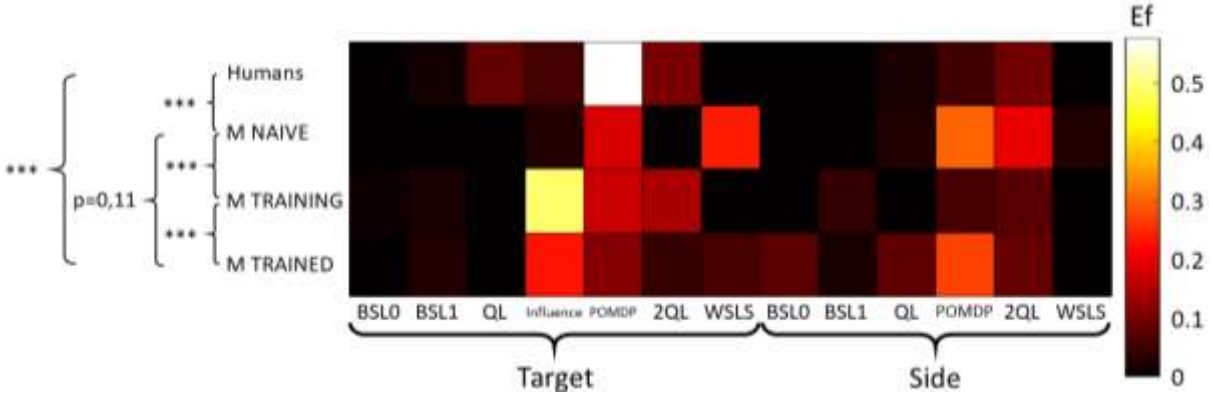


Figure 22: ***Computational learning models. Bayesian model comparison.*** *X-axis: model names; Y-axis: subjects' group (top to bottom: naive humans, naive macaques, macaques in training with a confederate, trained macaques); Color-scale: Estimated frequencies (Ef), i.e., the frequency at which one model fitted one subject best in the group population compared to the other tested models.*

In the previous correlational analysis, no learning mechanisms were identified. To elucidate the computations underlying the participants' strategic decision-making, we compared a variety of different computational models that we fitted trial-by-trial to the data of each participant. We fitted several computational models to our data to determine which model best captured the participants' behavior during their decision process. We started by testing whether the participants' decisions relied on the targets, or on the side of the cues. To do so, we fed the models with either the side of the cue as inputs, or the nature of the target (preferred or not-preferred target). We found that most humans based their decisions on the target-based POMDP model, an observational learning process based on reinforcement learning that observed the nature of the cue chosen by the other player, when accessible, to infer their best course of action (Figure 22). As for macaques, in the Naïve condition, the decision process that guided their choices varied more, depending on the pair and on their experience with the game. Some based their choices on the side of the cue (left vs right) rather than the nature of the cue itself (preferred vs non-preferred). Around 55% of the naïve sessions were fitted by models using the side of the cue and 45% the nature of the cue. We found a relatively high diversity of model representations and a broad spectrum of different types of models in naïve monkeys. Thus, although on occasions some macaques were best fitted by a target-based POMDP, most were better fitted by different models, namely target-based WSLS, side-based

POMDP and side-based 2QL. This computational modelling analysis agrees with the hypothesis that monkeys will not spontaneously employ the most common strategy among humans.

In the Training condition, macaques played against a Human confederate and had to pay attention to the target that was chosen by the human confederate to maximize their reward because the human experimenter regularly imposed their own choice in blocks of trials to train the macaques. The computational analysis found, in accordance to this expectation, that the macaques in training with a human confederate relied mostly on decision processes that were based on the nature (color) of the cue. In the majority of cases, their decision process was best fitted by an influence learning model that was based on a classical reinforcement learning model, with the integration of the influence of one's choice on the reward of the other player. This surprising result suggests that, in the majority of the training sessions, although the macaques were forced to pay attention to the nature (color) of the target, they did not pay attention, or at least did not exploit, the current choice of the human confederate, even when it was visually accessible, to modulate their choice. This could be explained by the fact that the macaques might have considered that their choice was unlikely to be followed by the confederate, that followed a strict pattern that was independent of the macaque's choices and speed to decide. Therefore, the macaques did not use the current trial, even though they had potential visual access to the trainers' choices, to make their own decisions. This may suggest that the choices of the trainer and themselves were somehow preordained and followed a predetermined pattern. Thus, their strategy to achieve coordination and optimize their rewards was to decipher the pattern by taking the players' previous options and outcomes into account, as well as the effect of their own choices on the reward of the other player.

In the Trained condition, most pairs (2, 3 and 4) continued to use the influence learning decision process, especially during the initial trial blocks. Pair 1 (Linus-Curius), and pair 2 (Elmo-Curius) in their later sessions, employed side-based decision strategies that focused almost exclusively on choosing the same side to achieve very high coordination rates. The best model of their behavior was an RL-model that also exploited the observation of the current visual cue, when accessible, to adapt and coordinate. The same decision strategy that was observed in naïve macaques for pairs 1 (Curius-Elmo) and 6 (Magnus-Flaffus).

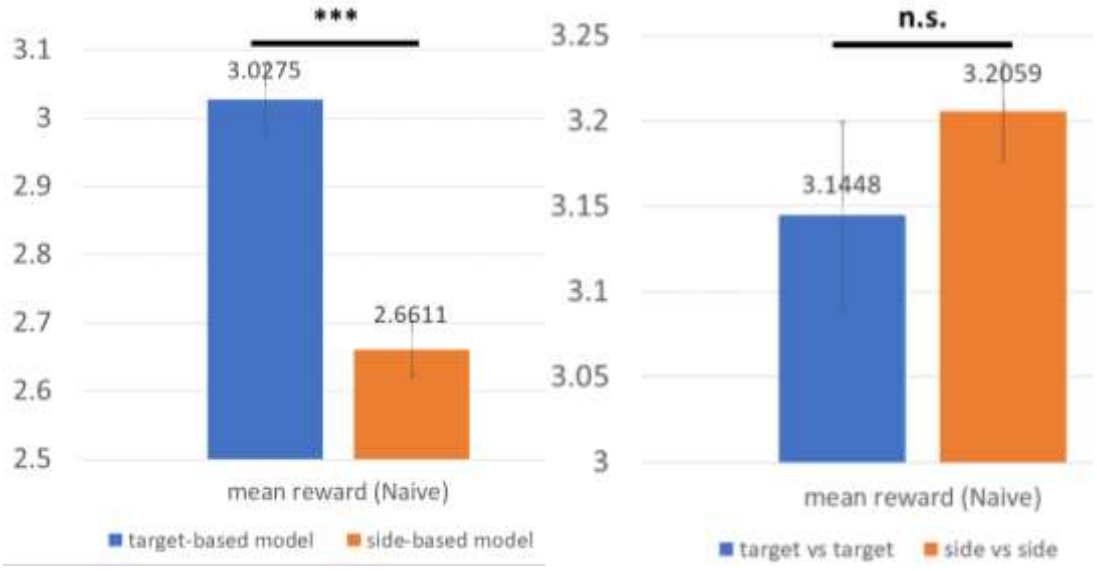


Figure 23: ***Computational learning models. Rewards of naïve macaques separated in target-based vs side-based categories.*** *Left: mean rewards over 1 session of naïve macaques that were classified as using different types (target-based vs side-based or side-based vs target-based) of decision*

processes. Right: mean rewards over 1 session of naïve macaques that were classified as using same types (target-based vs target-based or side-based vs side-based) of decision processes.

Macaques using target-based strategies that faced macaques using side-based strategies received higher rewards. The macaques that used target-based decision learning processes against side-based strategies were, on average, more rewarded than the macaques that used side-based strategies against target-based strategies ($p < 0.001$, Figure 23, Left). This might be because macaques that were target-based paid attention to selecting their preferred target and therefore likely influenced their partner into selecting the same target. The other macaque, that did not pay attention to the nature of the cue and might therefore have focused only on side-based coordination. Some might have not have understood the complete rules of the BoS game, especially in the naïve condition. Macaques that used target-based strategies against target-based strategies did not significantly differ in their reward from macaques that used side-based strategies against side-based strategies ($p = 0.279$, Figure 23, Right) overall.

Computational models that incorporate observation in their learning mechanism become more efficient during and after training to achieve higher coordination rates and rewards. We next tested whether macaques that were best fitted by a model that exploited the observation of their partner‘s current choice performed significantly better than macaques that were not exploiting the current choice observation. We found that in the Training and Trained conditions, macaques coordinated significantly more and therefore received significantly more reward when they used decision processes that exploited observational cues from their partner on current trial than macaques that did not (see Figure 24, Left). This suggests that decision processes that use observational cues correctly can foster coordination and increase rewards for the player that uses these visual cues efficiently. In the naïve condition, however, learning processes that incorporated observation of visual cues at current trial did not perform better than the processes that did not incorporate them. The monkeys seem to have failed to exploit the observation of their partner’s move correctly to maximize their reward. The coordination ratios in this naïve condition did not differ significantly.

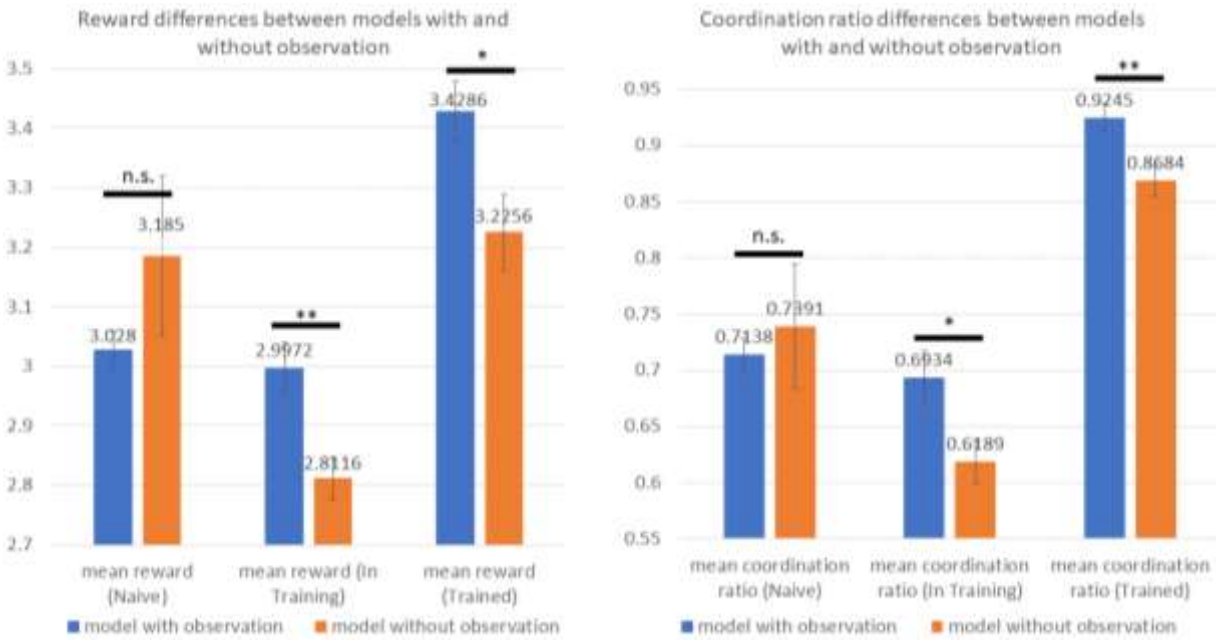


Figure 24: ***Differences in rewards and coordination ratios in different conditions for macaques.*** *Mean reward (left) and coordination ratio (right) over one session, when sessions are categorized by the nature (using observation vs not using observation) of the best model fitted to it.*

Control analyses and model checks.

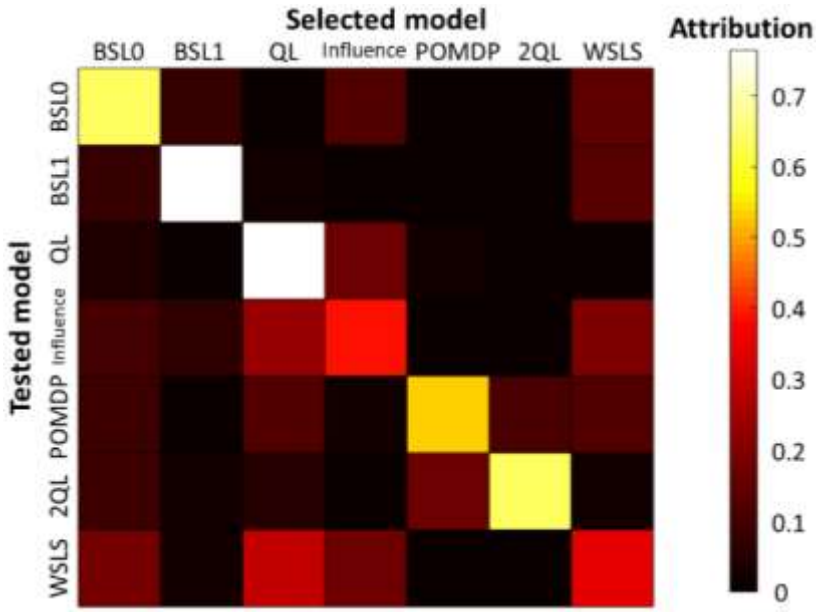


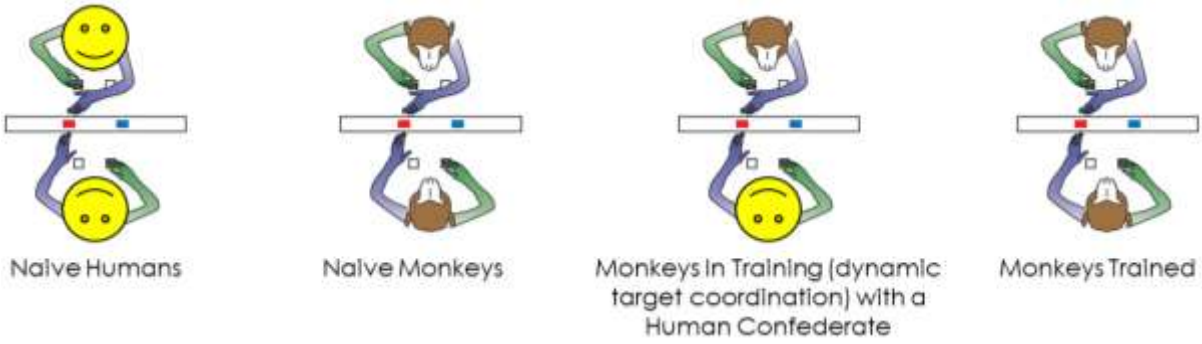
Figure 25: ***Model checks. (a) Confusion matrix.*** *X-axis: model selected as the best fitting one after model fitting and Bayesian Model Selection (BMS); Y-axis: initial simulated model that generated the data; Color-scale: model attribution ratio.*

We conducted a Confusion Analysis (CA) as a model check to ensure that the models that we compared, given our data, were likely to be confused between each other to confirm whether our model selection was reliable [21]. We therefore generated data multiple times with Monte-Carlo simulations 400 times for each of our tested models, with different sets of parameters, and conducted the same analysis as presented previously to determine whether our approach to distinguish between data generated according to each of the models was subsequently categorized appropriately by our analytical tools. This meant fitting the simulated data with all of the tested models, then performing a BMS on the models to select the one that fitted best.

If this analysis is reliable and the models are distinguishable, then the model that we simulated at the beginning should also be that selected by our analysis. The confusion matrix (Figure 25, see also the confusion table in Supp. Mat.) showed that the final selected model was indeed most frequently, that which was initially simulated by each model. This corroborates the reliability of the model attributions that we performed.

Methods

Participants



Naive Humans : 58 individuals, 29 pairs
Naive Monkeys : 6 individuals, 9 pairs
Monkeys in Training : 4 individuals
Monkeys Trained : 4 individuals, 4 pairs

BoS game

Payoff matrix		Player B (prefers blue)	
		Red	Blue
Player A (prefers red)	Red	3, 4	2, 2
	Blue	1, 1	4, 3

A's **own** choices

B's **own** choices

Table 1 : BoS payoff matrix.

In game theory, the game Bach-Or-Stravinsky (BoS), also referred as battle of the sexes, is a 2x2 (two-player, 2 options) coordination game. This game can be described as such: 2 friends agreed to meet in the evening to attend to a concert. But they forgot if they will be attending a Bach concert or a Stravinsky one. The fact that they forgot is besides common knowledge. The first one would prefer to go to the Bach concert. The other one would rather go to the Stravinsky one. Both would prefer to go to the same place rather than different ones. If they cannot communicate, where should they go ?

The payoff matrix represented in Table 1 is an example of Bach-Or-Stravinsky, where Player 1 chooses a row and Player 2 chooses a column. In each cell, the first number represents the payoff to Player 1 and the second number represents the payoff to Player 2. This representation accounts for the additional harm that might come from not only going to different locations, but going to the wrong one as well (i.e. player 1 goes to the Stravinsky concert while player 2 goes to the Bach one, satisfying neither). In this setup, they still receive a small reward as they still went out and did not stay bored at home.

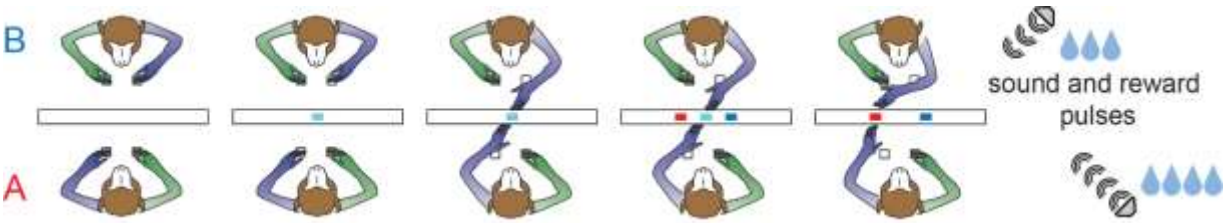
This game has two pure strategy Nash equilibria, one where both go to the Stravinsky concert and another where both go to the Bach concert. There is also a mixed strategies Nash

equilibrium in both games, where the players go to their preferred event more often than the other.

This presents an interesting case for game theory since each of the Nash equilibria is deficient in some way. The two pure strategy Nash equilibria are unfair; one player consistently does better than the other. The mixed strategy Nash equilibrium (when it exists) is inefficient: the players will miscoordinate regularly, leaving each player with an expected return smaller than the return one would receive from constantly going to one's less favored concert.

One possible resolution of the difficulty involves the use of a correlated equilibrium: for example, if the couple could flip a coin before choosing their strategies, they might agree to correlate their strategies based on the coin flip by, for example, choosing Bach for heads and Stravinsky for tails. Once the results of the coin flip are revealed, neither players have any incentives to alter their proposed actions: that would indeed result in miscoordination and a lower payoff than simply sticking to the agreed upon strategies. The result is that perfect coordination is always achieved and, prior to the coin flip, the expected payoffs for the players are exactly equal.

Experimental setup



Behavioral analysis

For the logistic regressions, we reported significant marginal effect of a given variable under the name “estimate” (for example: Cooperativity signature$_{t-1}$: estimate).

$$\text{Logistic regression: } \ln\left(\frac{p}{1-p}\right) = x_0 + x_1X_1 + x_2X_2 + \cdots$$

X_i represents independent variables and x_i the associated weights in the logistic regression. P represents the probability of a given event. The marginal effect of the variable X_1 is defined as:

$$\widehat{y_1} = mean(logit^{-1}(x_1))$$

The mean is computed across all observed data. Thus, the marginal effect called “estimate” can easily be interpreted as the discreet change of the dependent variable given a unitary change of an independent variable. For the linear regressions, reported “estimate” represents x_i i.e. the regression coefficient. Indeed, in a linear regression, marginal effect of a variable is equal to the estimated coefficient.

Learning Models

Bayesian Sequence Learning (BSL) model

The BSL model[22] tracks the (log) odds of $P(o_t=1|o_{t-K})$, where o is the partner's choice (binary outcome). This variable is updated according to a Laplace-Kalman filter, yielding 2 sufficient statistics (mean and variance) per combination of past outcomes. BSL can learn sequences of arbitrary depth K. For example:

- if K=1, then BSL tracks 2 probabilities, namely: $P(u_t=1|u_{t-1}=1)$ and $P(u_t=1|u_{t-1}=0)$. In this case, BSL needs to know about the partner's previous move u_{t-1}.
- if K=2, then BSL tracks 4 probabilities, namely: $P(u_t=1|u_{t-1}=1, u_{t-2}=1)$, $P(u_t=1|u_{t-1}=0,u_{t-2}=1)$, $P(u_t=1|u_{t-1}=1,u_{t-2}=0)$ and $P(u_t=1|u_{t-1}=0,u_{t-2}=0)$. In this case, BSL needs to know about the partner's two previous moves u_{t-1} and u_{t-2}. etc.

More generally, the BSL model tracks 2^K probabilities.

Win-Stay Lose-Switch (WSLS) model

This heuristic[15] keeps the same target as the previous one if previous reward was positive, and changes targets otherwise.

Q-Learning (QL) model

The Reinforcement Learning models[23]–[26] attempt to interpret changes in behavioral responses that arise from the subject’s exposure to environmental feedback, i.e., how people adapt their behavioral response as a result of rewards and/or punishment. The central concept in most reinforcement learning models is value, which quantifies how rewarding is a given action or item. A classic example of such a model is the so-called “Q-learning” model.

Influence Learning model

This mentalizing model was described by Hampton et al[11], who also found neural correlates with this model in humans[11]. The influence model relies on a Taylor expended reinforcement learning. This model describes computations underlying the capacity to mentalize (in the context of a strategic game). Let $p_n = P_n(o = 1)$ be the agent's prediction of the other's next move, i.e., the probability that the other will pick the first alternative option. The "influence learning" rule can be written as follows:

$$p_{n+1} = p_n + \eta * (o - p_n) - \lambda * k_1 * p_n * (1 - p_n) * (a - \frac{\beta * x_n - k_2}{k_1}) \qquad (1)$$

where o is the other's last move, a is the agent's last move, η is the weight of the agent's prediction error, λ is the weight of the other's prediction error, k_1 & k_2 are derived from the game's payoff table, x_n is the tracked log-odds of p_n, β is the other’s temperature, i.e. the tendency of the other to explore vs exploit information (the closer it is to 0, the more it explores by choosing randomly without using previous information, and the closer it is 1, the more it exploits previous information and chooses deterministically).
Note that for all subjects, the tested models fitted better than chance.

POMDP model

In POMDP models, the agent maintains probabilistic beliefs about hidden states of the world using Bayesian inference based on the sequence of his observations and actions. The state dynamics are supposed to be Markovian (i.e. the next state only depends on the current one), but the agent cannot directly observe the state and must instead use its sensors to obtain

observations and maintain a probability distribution over the set of possible states, based on the observations, observation probabilities, and the underlying dynamics.

To solve a POMDP problem, we must find the optimal action for each possible belief over the world states. The optimal action maximizes (or minimizes) the expected reward (or cost) of the agent over a possibly infinite horizon. The sequence of optimal actions is known as the optimal policy of the agent for interacting with its environment.

Anatomical, biological and functional suggestions can be drawn from the POMDP model as an extension of a reinforcement learning model. Some empirical evidence has begun to confirm these suggestions from the POMDP model[16]. The POMDP model can thus be a very interesting neuro-biologically plausible model for social decision making[27], [28].

2QL model

A model developed by Charpentier et al. [29] was used to study arbitration between imitation and emulation during human observational learning. In our experimental design, we hypothesized that the players would take a decision based on the information that they had access to. At each trial, the slowest player would thus see their partner's choice in a small timeframe, giving them a visual cue to exploit. We therefore considered 2 situations in which our players could be: they are either in the imitation mode where they have access to a visual cue of their partner's final decision and can thus choose to imitate or not, or they are in the emulation mode where they do not have any visual cue and take their decision based on the previous trials. Theoretically, the slowest player has the choice between taking into account the visual cue or not, hence the notion of arbitration between 2 modes. The 2QL model differentiates between the 2 modes: 2 QL schemes run in parallel, one for the emulation mode when the other’s choice is not definitive yet, and one for the imitation mode when the other player chose faster. In the latter case, the agent can choose to persist in their initial decision, and thus decide in the same fashion as if they were in the emulation mode. Or they can revise their revise their decision by applying a different QL scheme with different inputs now that the other’s choice is made and plainly visible.

Bayesian Model Selection (BMS)

The Bayesian Model Selection (BMS) was performed using the VBA toolbox (Variational Bayesian Analysis) in a random effect analysis relying on the free energy as the lower bound of model evidence. We use Exceedance Probability measurements (EP)[30] to select the model which is used most frequently in our population.

Discussion (821 / ~800 words)

Primates are often led to interact socially, directly face-to-face, with their conspecifics. Learning efficiently how to interact and coordinate with a conspecific in a given environment is most likely a determining factor for the survival of the fittest, and also a fundamental characteristic in which Humans might have differentiated themselves from other Primates. However, few studies have tested the emergence of coordination in transparent games. Even fewer have applied computational modelling approaches to characterize coordination in strategic decision-making games. In this paper, we applied a computational modelling approach to study the learning process of different groups of Primates under different conditions. We show that the difference in choice behaviors between adult humans and rhesus macaques comes from a difference in decision processes. We identified a target-based POMDP decision process in humans that fully exploited the information of the visual cue to

coordinate. In contrast, macaques after training were mainly better fitted by either a target-based influence model that did not exploit the current visible choice, or a side-based POMDP model that exploited the visible choice of the other agent to coordinate on the same side.

The nature of the computational models that were found to fit better macaques' behavior can be explained to a certain extent by both current literature and previous behavioral observations. The target-based influence model was indeed found to be a common model in other species of monkeys that played similar strategic decision games in a non-transparent setup [31]. The transparent nature of our settings however allowed another type of strategy that led to high rewards with low efforts: the side-based POMDP. Coordinating based on the side of the other's decision rather than on the nature of the cue requires indeed less effort and attention while also leading to high rewards. In humans, however, they were found to use in non-transparent setups more sophisticated, ToM-based (Theory-of-Mind), Bayesian, decision processes [32]. In our transparent setup, the visual access to the other's current action hijacks the usefulness to use such sophisticated computational models, though. Therefore, the use of a target-based POMDP model makes more sense as it spares the computational complexity of high-order mentalizing to focus on the exploitation of the current visual cue to coordinate.

Computational modelling is a powerful approach to understand the dynamics of individuals' behavior when faced with a decision. A strength of our computational approach was to assess and compare a large variety of competing models: Bayesian Sequence Learner, Reinforcement Learning, win-stay/lose-switch, influence model, POMDP model, arbitration-based 2QL model. Many have never previously been directly tested against each other. Here, we explicitly tested the predictions of different families of learning models against one another, investigating not only target-based vs side-based models, but also non-observing vs observing models that deployed (or not) an arbitration process between an imitation and an emulation decision process. Our results agree with studies that concluded that social learning may be driven by Bayesian algorithms to model inferences about the future actions of another [32]–[35]. Notably, this approach has recently been applied to strategic cooperation in groups[16], [36]. Together, these findings provide a mechanistic framework for learning in strategic transparent games across different species.

Learning by observation might be one of the main mechanisms that could lead to coordination in high-level choice decision processes. Such learning engages two types of prediction error: the prediction error of the observed action and the prediction error of the observed result of the action [37], [38]. This raises the question about the choice between two learning strategies by observation: action imitation or new choice emulation. Imitation is here defined as ‘‘choice’’ imitation, in line with the economics, decision neuroscience, and reinforcement learning literature [39], [40]. It requires an intrinsic reinforcement mechanism that rewards imitation itself because no reward is directly received by simply observing a conspecific [41]. As for emulation, more specifically ‘‘goal’’ emulation, it consists of a more complex and flexible Bayesian inference process [32], [42], [43]. It combines prior beliefs about the other agents with the evidence received from observation to produce posterior updated beliefs. When alternating between observation then action, humans can be modelled by a computational process that simulates both the imitation and emulation processes simultaneously, and that then arbitrates between these processes [29]. Furthermore, brain regions that correlate in activation with the imitation process were found to be the ones that have previously been implicated in tasks involving mirror neurons [44]–[48]. As for the emulation and arbitration processes, they were implemented by a network that has been linked to mentalizing tasks [11], [12], [34], [49], [50].

Observation learning was therefore possibly one of the key precursor mechanisms that led to the emergence of the ability to mentalize the mental states of others, i.e. the development

of Theory of Mind (ToM). This type of learning might be the gateway to understanding the main differences in ToM abilities between humans and other primates. The behavior of other individuals might not be viewed as simple movements anymore, but as intentional actions, planned and goal-directed, motivated by desires and guided by beliefs [51], [52].

Acknowledgements

Author contributions

Competing interests

The authors declare that they have no competing interests.

Resources availability

Lead contact

Further information and requests for resources and inquiries should be directed to and will be fulfilled by the lead contact, Jean-Claude DREHER (dreher@isc.cnrs.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Code supporting main results are available at XXXXXXXX

Supplementary Materials

Bayesian Model Selection and Data Analysis [53]

Generally, a data analysis aims to discover useful information that can support downstream conclusions and/or decisions. This is trivially true for statistical data analysis, which is typically used to summarize data, for example with descriptive statistics, or to perform inference, for example hypothesis testing. We will describe in this section a particular form of statistical data analysis: Bayesian data analysis. It essentially relies on generative models, i.e. quantitative scenarios that describe how data were generated, in order to interpret the observed data. This type of analysis has been particularly studied by Daunizeau et al. [54], leading to the implementation of the now widely used VBA toolbox on Matlab to test, optimize and compare models in Neurosciences.

One of the aim of the VBA toolbox is to derive the posterior probability density function $p(\theta|y, \sigma, m)$, which quantifies how likely is any possible value of model parameter θ, given the observed data y and the hyperparameter σ under the generative model m, by solving the "inverse problem", also called "inversing the model" :

$$p(\theta|y,\sigma,m) = \frac{p(y|\theta,\ \sigma,m) * p(\theta|\sigma,m)}{p(y|\sigma,m)}$$

where m is the model, σ is the noise' standard deviation and θ are the model parameters.

This allows us to test and optimize models with their parameters. But another aim of the VBA toolbox is to compare different models. Bayesian model comparison primarily relies on the "marginal likelihood" p(y|m) (the so-called "model evidence"), which can be derived from marginalizing all parameters and hyperparameters out of the likelihood function $p(\theta|y, \sigma, m)$ (see [54]} for more details) which quantifies how likely is the observed data y under the generative model m (we usually rather use log-likelihood, see the models comparison chapter). Interestingly, the model evidence p(y|m) is not a simple measure of model fit: there is an inherent penalization for model complexity, and p(y|m) is essentially a trade-off between goodness-of-fit and model complexity.

Learning Models' Structure [53]

We will here describe the main assumptions and notations that are used in VBA. We assume that any form of learning (including probabilistic - bayesian - belief update), can be written as an evolution equation and that any form of decision making can be understood as an action emission law, and thus written as an observation mapping (from internal states to actions). More generally, most computational models for neurobiological and behavioural data share the same structure: they are based on evolution and/or observation mappings that capture the response of relevant "states" (e.g. neural activity, beliefs and preferences, etc…) to experimentally controlled inputs.

We will first define the following notations:

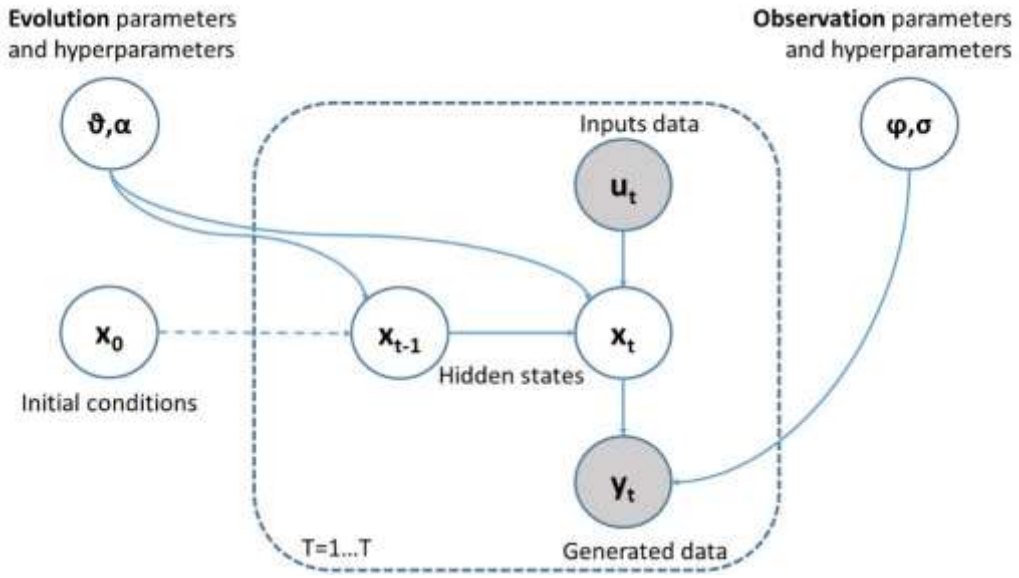
- y: experimentally measured data. These can be categorical or continuous. They are the data to compare with the output data generated by the model.
- x: hidden states. These are time-dependent model variables, and their dynamics is controlled by the evolution function (e.g. action values or Q-values).

- theta: evolution parameters. These determine the evolution function of hidden states (e.g. a learning rate).
- phi: observation parameters. These determine the observation mapping (e.g. action values or Q-values).
- u: experimentally controlled inputs. They are the data which we suppose will influence the evolution of the hidden states x of the model.
- so-called generative model. This encompasses all statistical assumptions that subtend the analysis.

VBA deals with a very general class of generative models, namely "nonlinear state-space models". These generative models take the following form, consisting of essentially two mappings:

- A quantitative assumption or model of how observations y are generated, with the unknown parameters θ, which are mapped through an observation function g such that: $y = g(x, u, \theta) + \varepsilon$,
  where ε are model residuals or measurement noise and usually follow a normal distribution with a standard deviation σ. The observation function (or mapping) g describes how observed data is generated from hidden states.
- A quantitative assumption or model of what the hidden states are and how they evolve, with the unknown set of parameters φ. The evolution function f describes how hidden states change from one time sample to the next:
  $x_{t+1} = f(x_t, u_t, \varphi) + \eta_t$
  where η_t are stochastic perturbations, often supposed to be i.i.d. Gaussian with mean zero. It is mainly the learning rule step.

The Supp.Fig. 1 summarizes the model structure. In principle, VBA only deals with Markovian systems, i.e. systems whose evolution depends solely upon their current state. But in fact, a very simple solution to this apparent issue is to augment the native state-space processes with dummy states that are copies of past instances of native states. This allows us to describe any non-Markovian system in terms of a (higher-dimensional) Markovian system.



Supp.Fig.1. Formal structure of models used in VBA. From [53].

POMDP model

In social neuroscience, there is a need to find satisfying theoretical frameworks to model feedback-dependent changes. Indeed, in a social context, the degree of uncertainty about the possible outcomes increases dramatically because the behavior of other human beings can be much more difficult to predict than the physics of the environment. Therefore, the previously mentioned theoretical framework used to model feedback-dependent changes in decision making strategies is unreliable because assumes that optimal decisions can be determined from the state of the decision makers' environment and that this state is fully observable to the decision maker. These assumptions clearly do not reflect the reality and complexity of human social decision making, as observable behaviors of other humans provide only very partial information about their likely future behaviors. Thus, a new model was proposed to provide a formal probabilistic framework for solving tasks involving action selection and decision making under uncertainty: Partially Observable Markovian Decision Processes (POMDPs).

A rigorous theoretical framework previously proposed by Rao[16] assumes that the brain performs Bayesian inference based on observations using probabilistic representations of the world and utilizes the results of Bayesian inference to choose optimal actions. A main hypothesis of this framework is that the brain performs Bayesian inferences using probabilistic representations of not only environmental states but also other individuals' intentions and behavior. The probabilistic framework of POMDPs can be used to model ways in which humans might predict hypothetical action outcomes, the intentions of others, and whether the other is cooperative or competitive.

In POMDP models, the agent maintains probabilistic beliefs about hidden states of the world using Bayesian inference based on the sequence of his observations and actions. The state dynamics are supposed to be Markovian (i.e. the next state only depends on the current one), but the agent cannot directly observe the state and must instead use its sensors to obtain observations and maintain a probability distribution over the set of possible states, based on the observations, observation probabilities, and the underlying dynamics. We assume that the brain uses Bayesian probability to perform complex inferences. The main rule of Bayesian probability is:

$$p(H|D) = \frac{p(D|H) * p(H)}{p(D)}$$

Where H is the hypothesis and D is our data. Thus, the hypothesis is not true or false like in Boolean logic, but proportional to the likelihood (p(D|H)) and multiplied by the prior probability (p(H)). The Bayesian Brain theory uses this main rule and assumes that the brain codes the probability of events occurring in its environment, collects the data and computes future inferences based on the prior probability. This can help explain many observations in sensory experiments, language learning and decision-making experiments.

To solve a POMDP problem, we must find the optimal action for each possible belief over the world states. The optimal action maximizes (or minimizes) the expected reward (or cost) of the agent over a possibly infinite horizon. The sequence of optimal actions is known as the optimal policy of the agent for interacting with its environment.

Anatomical, biological and functional suggestions can be drawn from the POMDP model as an extension of a reinforcement learning model. Some empirical evidence has begun to confirm these suggestions from the POMDP model[16]. The POMDP model can thus be a very interesting neuro-biologically plausible model for social decision making[27], [28].

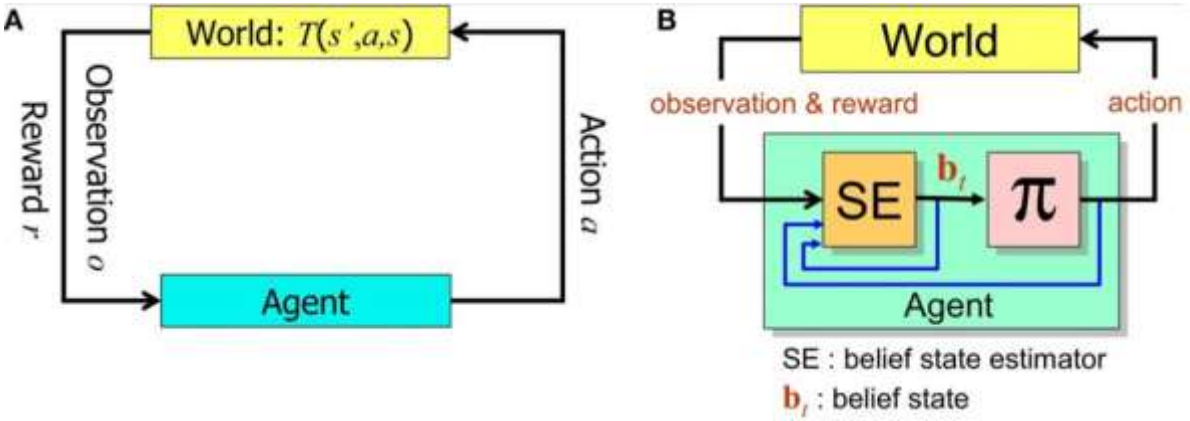


Figure. The POMDP model. ***(A)*** When the animal executes an action a in the state s', the environment ("World") generates a new state s according to the transition probability T(s',a,s). The animal receives an observation o of the new state according to P(o|s) and a reward r = R(s',a). ***(B)*** In order to solve the POMDP problem, the animal maintains a belief bt which is a probability distribution over states of the world. This belief is computed iteratively using Bayesian inference by the belief state estimator SE. An action for the current time step is provided by the learned policy π, which maps belief states to actions. From [16].

Free energy [53]

Typically, generative models have some form of nonlinearity, in the way the unknown model parameters ϑ impact the data y. In particular, the ensuing likelihood function may contain high-order interaction terms between subsets of the unknown model parameters (e.g., because of nonlinearities in the model). This implies that the high-dimensional integrals required for Bayesian parameter estimation and model comparison cannot be evaluated analytically. Also, it might be computationally very costly to evaluate them using numerical brute force or Monte-Carlo sampling schemes. This motivates the use of variational approaches to approximate Bayesian inference [55]. In brief, variational Bayes (VB) is an iterative scheme that indirectly optimizes an approximation to both the model evidence p(y|m) and the posterior density p(ϑ|y,m). The key trick is to decompose the log model evidence into:

$$log(p(y|m) = F(q) + DKL(q(\vartheta); p(\vartheta|y, m))$$

where q(ϑ) is any density over the model parameters, DKL is the Kullback-Leibler divergence and the so-called free energy F(q) is defined as:

$$F(q) = \langle ln(p(y \mid \vartheta, m))\rangle q - DKL(q(\vartheta); p(\vartheta \mid m))$$

where the expectation ⟨.⟩q is taken under q. One can see that maximizing the functional F(q) with respect to q indirectly minimizes the Kullback-Leibler divergence between q(ϑ) and the exact posterior p(ϑ|y,m). The decomposition of the log evidence is complete in the sense that if q(ϑ)=p(ϑ|y,m), then F(q)=ln(p(y|m)).

The iterative maximization of free energy is done under simplifying assumptions about the functional form of q, rendering q an approximate posterior density over model parameters and F(q) an approximate log model evidence (actually, a lower bound).

GLMs & GLMEs.

reward_n ~ 1 + isOwnChoice_n_1 + reward_n_1 + switch_n_1 + isOwnChoice_n_2 + reward_n_2 + switch_n_2 + isOwnChoicen_3 + reward_n_3 + switch_n_3 + Species + RTdiff + Session + Condition * Trial + (Session:Trial) + (1 + Trial | Subj_name)

isOwnChoice_n ~ 1 + isOwnChoice_n_1 + reward_n_1 + switch_n_1 + isOwnChoice_n_2 + reward_n_2 + switch_n_2 + isOwnChoicen_3 + reward_n_3 + switch_n_3 + Species + RTdiff + Session + Condition*Trial + (Session:Trial) + (1 + Trial | Subj_name)

Models	GLM		GLME	
Variables	$Reward_n$: Estimate (Std. Error)	$isOwnChoice_n$: Estimate (Std. Error)	$Reward_n$: Estimate (Std. Error)	$isOwnChoice_n$: Estimate (Std. Error)
Intercept			0.1057 (0.07822)	-0.729418 (0.780015)
$Reward_{n-1}$			0.2484 (2.272e-03) ***	0.117514 (0.007264) ***
$Reward_{n-2}$			1.917e-01 (2.302e-03) ***	0.098733 (0.007321) ***
$Reward_{n-3}$			1.428e-01 (2.282e-03) ***	0.012887 (0.007326) .
$isOwnChoice_{n-1}$			-8.791e-04 (3.874e-03)	0.735471 (0.011399) ***
$isOwnChoice_{n-2}$			1.184e-02 (3.889e-03) **	0.740430 (0.011405) *
$isOwnChoice_{n-3}$			-1.647e-02 (3.902e-03) ***	0.511239 (0.011521) ***
$switch_{n-1}$			-1.729e-02 (3.866e-03) ***	0.119188 (0.011527) ***
$switch_{n-2}$			-1.809e-03 (3.877e-03)	0.123541 (0.011711)

				***
$switch_{n-3}$			3.863e-03 (3.799e-03)	0.178744 (0.011576) ***
Species			9.589e-02 (7.944e-02)	-0.141797 (0.844379)
DTdiff			-2.670e-01 (9.858e-03) ***	-2.637016 (0.033562) ***
Session			2.743e-01 (1.548e-02) ***	-0.153624 (0.050344) **
Condition			6.914e-02 (4.690e-03) ***	-0.249911 (0.015013) ***
Trial			2.615e-01 (5.035e-02) ***	-0.883955 (0.503119) .
Condition:Trial			-7.655e-02 (1.279e-02) ***	0.313174 (0.041136) ***
Session:Trial			-1.122e-01 (4.681e-02) *	0.321757 (0.151981) *
			/	/

Supp. Table 1: Comparison of GLMs and GLMEs for behavioral factors. Hierar_cond, AA_mode, Coop_sig, Reward, AA_Stay and Stay are coded effects (Hierar_cond: superior = 1, equal = 0, inferior = -1; AA_mode: competitive = -1, random = 0, cooperative = 1; Coop_sig, Reward, AA_Stay & Stay: True = 1, False = 0). Trial & DT were normalized. Signif. codes: '***': $p < 0.001$; '**': $p < 0.01$; '*': $p < 0.05$.

Idea: Add model attributions for observation vs non-observation model types & target- vs side-based models

Bibliography

[1] R. B. Boone, S. J. Thirgood, and J. G. C. Hopcraft, 'Serengeti wildebeest migratory patterns modeled from rainfall and new vegetation growth.', *Ecology*, vol. 87, no. 8, pp. 1987–94, Aug. 2006, doi: 10.1890/0012-9658(2006)87[1987:swmpmf]2.0.co;2.

[2] A. P. Melis and M. Tomasello, 'Chimpanzees (Pan troglodytes) coordinate by communicating in a collaborative problem-solving task', *Proceedings of the Royal Society B*, vol. 286, no. 1901, Apr. 2019, doi: 10.1098/RSPB.2019.0408.

[3] A. Sánchez-Amaro, S. Duguid, J. Call, and M. Tomasello, 'Chimpanzees, bonobos and children successfully coordinate in conflict situations', *Proceedings of the Royal Society B: Biological Sciences*, vol. 284, no. 1856, Jun. 2017, doi: 10.1098/rspb.2017.0259.

[4] S. Möller, A. M. Unakafov, J. Fischer, A. Gail, S. Treue, and I. Kagan, 'Human and macaque pairs employ different coordination strategies in a transparent decision game', *Elife*, vol. 12, 2023, doi: 10.7554/eLife.81641.

[5] A. M. Unakafov *et al.*, 'Emergence and suppression of cooperation by action visibility in transparent games', *PLoS Comput Biol*, vol. 16, no. 1, p. e1007588, 2020, doi: 10.1371/journal.pcbi.1007588.

[6] D. M. Kilgour and N. M. Fraser, 'A taxonomy of all ordinal 2 × 2 games', *Theory Decis*, vol. 24, no. 2, pp. 99–117, Mar. 1988, doi: 10.1007/BF00132457.

[7] D. Sonsino and J. Sirota, 'Strategic pattern recognition - Experimental evidence', *Games Econ Behav*, vol. 44, no. 2, pp. 390–411, Aug. 2003, doi: 10.1016/S0899-8256(03)00040-X.

[8] C. A. Ioannou and J. Romero, 'Learning with repeated-game strategies', *Front Neurosci*, vol. 8, no. 8 JUL, p. 212, Jul. 2014, doi: 10.3389/FNINS.2014.00212/ABSTRACT.

[9] A. P. Melis, P. Grocke, J. Kalbitz, and M. Tomasello, 'One for You, One for Me: Humans' Unique Turn-Taking Skills', *Psychol Sci*, vol. 27, no. 7, pp. 987–996, Jul. 2016, doi: 10.1177/0956797616644070.

[10] J. P. O'Doherty, J. Cockburn, and W. M. Pauli, 'Learning, Reward, and Decision Making', *Annual Review of Psychology*, vol. 68. Annual Reviews, pp. 73–100, Jan. 04, 2017. doi: 10.1146/annurev-psych-010416-044216.

[11] A. N. Hampton, P. Bossaerts, and J. P. O'Doherty, 'Neural correlates of mentalizing-related computations during strategic interactions in humans', *Proc Natl Acad Sci U S A*, vol. 105, no. 18, pp. 6741–6746, 2008, doi: 10.1073/pnas.0711099105.

[12] C. A. Hill, S. Suzuki, R. Polania, M. Moisa, J. P. O'doherty, and C. C. Ruff, 'A causal account of the brain network computations underlying strategic social behavior', *Nat Neurosci*, vol. 20, no. 8, pp. 1142–1149, 2017, doi: 10.1038/nn.4602.

[13] C. Mathys, J. Daunizeau, K. J. Friston, and K. E. Stephan, 'A Bayesian foundation for individual learning under uncertainty', *Front Hum Neurosci*, vol. 5, no. MAY, p. 9, 2011, doi: 10.3389/fnhum.2011.00039.

[14] C. D. Mathys *et al.*, 'Uncertainty in perception and the Hierarchical Gaussian filter', *Front Hum Neurosci*, vol. 8, no. November, pp. 1–24, 2014, doi: 10.3389/fnhum.2014.00825.

[15] M. Nowak and K. Sigmund, ‘A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner’s Dilemma game’, *Nature*, vol. 364, no. 6432, pp. 56–58, 1993, doi: 10.1038/364056a0.

[16] R. P. N. Rao, ‘Decision making under uncertainty: A neural model based on partially observable Markov decision processes’, *Front Comput Neurosci*, vol. 4, p. 146, Nov. 2010, doi: 10.3389/fncom.2010.00146.

[17] G. S. Van Doorn, T. Riebli, and M. Taborsky, ‘Coaction versus reciprocity in continuous-time models of cooperation’, *J Theor Biol*, vol. 356, pp. 1–10, Sep. 2014, doi: 10.1016/J.JTBI.2014.03.019.

[18] E. Proto, A. Rustichini, and A. Sofianos, ‘Intelligence, Personality, and Gains from Cooperation in Repeated Interactions’, *https://doi.org/10.1086/701355*, vol. 127, no. 3, pp. 1351–1390, Jun. 2019, doi: 10.1086/701355.

[19] A. Peysakhovich, M. A. Nowak, and D. G. Rand, ‘Humans display a “cooperative phenotype” that is domain general and temporally stable’, *Nat Commun*, vol. 5, 2014, doi: 10.1038/NCOMMS5939.

[20] E. Herrmann, J. Call, M. V. Hernández-Lloreda, B. Hare, and M. Tomasello, ‘Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis’, *Science*, vol. 317, no. 5843, pp. 1360–1366, Sep. 2007, doi: 10.1126/SCIENCE.1146282.

[21] S. Palminteri, V. Wyart, and E. Koechlin, ‘The Importance of Falsification in Computational Cognitive Modeling’, *Trends in Cognitive Sciences*, vol. 21, no. 6. Trends Cogn Sci, pp. 425–433, Jun. 01, 2017. doi: 10.1016/j.tics.2017.03.011.

[22] J. Daunizeau and L. Rigoux, ‘Bayesian sequence learning’, *https://mbb-team.github.io/VBA-toolbox/wiki/BSL/*, 2014. https://mbb-team.github.io/VBA-toolbox/wiki/BSL/ (accessed Jan. 12, 2022).

[23] P. Dayan and B. W. Balleine, ‘Reward, motivation, and reinforcement learning’, *Neuron*, vol. 36, no. 2. Cell Press, pp. 285–298, Oct. 10, 2002. doi: 10.1016/S0896-6273(02)00963-7.

[24] L. Zhu, K. E. Mathewson, and M. Hsu, ‘Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning’, *Proc Natl Acad Sci U S A*, vol. 109, no. 5, pp. 1419–1424, Jan. 2012, doi: 10.1073/PNAS.1116783109/-/DCSUPPLEMENTAL.

[25] M. A. J. Apps, R. Green, and N. Ramnani, ‘Reinforcement learning signals in the anterior cingulate cortex code for others’ false beliefs’, *Neuroimage*, vol. 64, no. 1, pp. 1–9, Jan. 2013, doi: 10.1016/J.NEUROIMAGE.2012.09.010.

[26] A. Konovalov and I. Krajbich, ‘Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning’, *Nat Commun*, vol. 7, no. 1, pp. 1–11, Aug. 2016, doi: 10.1038/ncomms12438.

[27] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, J. O’Doherty, and G. Pezzulo, ‘Active inference and learning’, *Neurosci Biobehav Rev*, vol. 68, pp. 862–879, Sep. 2016, doi: 10.1016/J.NEUBIOREV.2016.06.022.

[28] K. J. Friston, R. Rosch, T. Parr, C. Price, and H. Bowman, ‘Deep temporal models and active inference’, *Neuroscience and Biobehavioral Reviews*, vol. 77. Pergamon, pp. 388–402, Jun. 01, 2017. doi: 10.1016/j.neubiorev.2017.04.009.

[29] C. J. Charpentier, K. Iigaya, and J. P. O'Doherty, 'A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning', *Neuron*, vol. 106, no. 4, pp. 687-699.e7, May 2020, doi: 10.1016/J.NEURON.2020.02.028.

[30] L. Rigoux, K. E. Stephan, K. J. Friston, and J. Daunizeau, 'Bayesian model selection for group studies - revisited', *Neuroimage*, vol. 84, pp. 971–985, Jan. 2014, doi: 10.1016/J.NEUROIMAGE.2013.08.065.

[31] M. Devaine *et al.*, 'Reading wild minds: A computational assay of Theory of Mind sophistication across seven primate species', *PLoS Comput Biol*, vol. 13, no. 11, pp. 1–25, 2017, doi: 10.1371/journal.pcbi.1005833.

[32] M. Devaine, G. Hollard, and J. Daunizeau, 'The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn?', *PLoS Comput Biol*, vol. 10, no. 12, 2014, doi: 10.1371/journal.pcbi.1003992.

[33] M. Devaine, G. Hollard, and J. Daunizeau, 'Theory of mind: Did evolution fool us?', *PLoS One*, vol. 9, no. 2, 2014, doi: 10.1371/journal.pone.0087619.

[34] W. Yoshida, B. Seymour, K. J. Friston, and R. J. Dolan, 'Neural Mechanisms of Belief Inference during Cooperative Games', *Journal of Neuroscience*, vol. 30, no. 32, pp. 10744–10751, Aug. 2010, doi: 10.1523/JNEUROSCI.5895-09.2010.

[35] J. M. Barnby, P. Dayan, and V. Bell, 'Formalising social representation to explain psychiatric symptoms', *Trends in cognitive sciences*, vol. 27, no. 3. The Authors, pp. 317–332, 2023. doi: 10.1016/j.tics.2022.12.004.

[36] K. Khalvati *et al.*, 'Modeling other minds: Bayesian inference explains human choices in group decision-making', *Sci Adv*, vol. 5, no. 11, Nov. 2019, doi: 10.1126/SCIADV.AAX8783.

[37] S. Suzuki *et al.*, 'Learning to Simulate Others' Decisions', *Neuron*, vol. 74, no. 6, pp. 1125–1137, Jun. 2012, doi: 10.1016/j.neuron.2012.04.030.

[38] C. J. Burke, P. N. Tobler, M. Baddeley, and W. Schultz, 'Neural mechanisms of observational learning', *Proc Natl Acad Sci U S A*, vol. 107, no. 32, pp. 14431–14436, 2010, doi: 10.1073/pnas.1003111107.

[39] E. Mossel, M. Mueller-Frank, A. Sly, and O. Tamuz, 'Social Learning Equilibria', *Econometrica*, vol. 88, no. 3, pp. 1235–1267, May 2020, doi: 10.3982/ECTA16465.

[40] E. Eyster and M. Rabin, 'Extensive Imitation is Irrational and Harmful', *Q J Econ*, vol. 129, no. 4, pp. 1861–1898, Nov. 2014, doi: 10.1093/QJE/QJU021.

[41] J. Joiner, M. Piva, C. Turrin, and S. W. C. Chang, 'Social learning through prediction error in the brain', *npj Science of Learning 2017 2:1*, vol. 2, no. 1, pp. 1–9, Jun. 2017, doi: 10.1038/s41539-017-0009-2.

[42] S. Collette, W. M. Pauli, P. Bossaerts, and J. O'Doherty, 'Neural computations underlying inverse reinforcement learning in the human brain', *Elife*, vol. 6, Oct. 2017, doi: 10.7554/eLife.29718.

[43] A. O. Diaconescu, C. Mathys, L. A. E. Weber, J. Daunizeau, and L. Kasper, 'Inferring on the Intentions of Others by Hierarchical Bayesian Learning', *PLoS Comput Biol*, vol. 10, no. 9, p. 1003810, 2014, doi: 10.1371/journal.pcbi.1003810.

[44] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, ‘Premotor cortex and the recognition of motor actions’, *Brain Res Cogn Brain Res*, vol. 3, no. 2, pp. 131–141, 1996, doi: 10.1016/0926-6410(95)00038-0.

[45] G. Rizzolatti and L. Craighero, ‘The mirror-neuron system’, *Annu Rev Neurosci*, vol. 27, pp. 169–192, 2004, doi: 10.1146/ANNUREV.NEURO.27.070203.144230.

[46] D. R. Lametti and K. E. Watkins, ‘Cognitive Neuroscience: The Neural Basis of Motor Learning by Observing’, *Curr Biol*, vol. 26, no. 7, pp. R288–R290, Apr. 2016, doi: 10.1016/J.CUB.2016.02.045.

[47] R. Cook, G. Bird, C. Catmur, C. Press, and C. Heyes, ‘Mirror neurons: from origin to function’, *Behav Brain Sci*, vol. 37, no. 2, pp. 177–192, 2014, doi: 10.1017/S0140525X13000903.

[48] C. Catmur, V. Walsh, and C. Heyes, ‘Associative sequence learning: the role of experience in the development of imitation and the mirror system’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1528, pp. 2369–2380, Aug. 2009, doi: 10.1098/RSTB.2009.0048.

[49] C. D. Frith and U. Frith, ‘The neural basis of mentalizing’, *Neuron*, vol. 50, no. 4, pp. 531–534, May 2006, doi: 10.1016/J.NEURON.2006.05.001.

[50] T. E. J. Behrens, L. T. Hunt, M. W. Woolrich, and M. F. S. Rushworth, ‘Associative learning of social value’, *Nature*, vol. 456, no. 7219, pp. 245–249, Nov. 2008, doi: 10.1038/NATURE07538.

[51] F. Castelli, F. Happé, U. Frith, and C. Frith, ‘Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns’, *Neuroimage*, vol. 12, no. 3, pp. 314–325, 2000, doi: 10.1006/nimg.2000.0612.

[52] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, ‘Rational quantitative attribution of beliefs, desires and percepts in human mentalizing’, *Nat Hum Behav*, vol. 1, no. 4, p. 64, 2017, doi: 10.1038/s41562-017-0064.

[53] J. Daunizeau and L. Rigoux, ‘The variational Bayes approach’, *https://mbb-team.github.io/VBA-toolbox/wiki/The-variational-Bayesian-approach/*, 2014. https://mbb-team.github.io/VBA-toolbox/wiki/The-variational-Bayesian-approach/ (accessed Jan. 12, 2022).

[54] J. Daunizeau, V. Adam, and L. Rigoux, ‘VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data’, *PLoS Comput Biol*, vol. 10, no. 1, p. 1003441, 2014, doi: 10.1371/journal.pcbi.1003441.

[55] M. J. Beal, ‘VARIATIONAL ALGORITHMS FOR APPROXIMATE BAYESIAN INFERENCE’, 1998.

B. Study 2

Mentalising underlies strategic coordination in Guinea baboons (Papio papio)

Toan Nong[1,2], Nicolas Claidière[3,4], Joël Fagot[3,4], Rémi Philippe[1,2], Edmund Derrington[1,2], Jean-Claude Dreher[1,2]

1- Institut des Sciences Cognitives Marc Jeannerod, CNRS, *Neuroeconomics Lab, Bron, France*
2- *Lyon Univ UCBL1, Lyon, France*
3- *Aix Marseille Univ, CNRS, LPC, Marseille, France*
4- *Station de Primatologie-Celphedia, CNRS UAR846, Rousset, France*

Summary

It remains controversial whether the ability to represent the world from the perspective of another (mentalising), is confined to humans[1–3]. To address this question, we combined computational models of mentalising with a new experimental setup of Guinea baboons living in a social colony, which required no direct human intervention or teaching. Baboons freely came to play a strategic coordination game via touchscreen devices with any other baboon (social condition), or alone (solo condition). In fact, in both conditions, they were interacting with an identical Artificial Agent. The ability to successfully coordinate with the Artificial Agent was better in the social than in the solo condition. A computational mentalising model that predicts the effect of one's actions on the partners' decisions[4,5] accounted for baboon's behaviour, much better than reinforcement learning[6,7], Bayesian[8–10] or heuristic models[11] that lacked mentalising components. Such computations accounted best for behaviour in the social condition only, because when they played alone, the same baboons used a win stay/lose switch strategy. Together, these findings indicate that computations required for mentalising are present in the Guinea baboon and provide an evolutionary advantageous for efficient coordination.

Many challenges arise to study whether non-human species attribute mental states to their con-specifics, i.e. have Theory of Mind (ToM) abilities[12,13]. So far, most studies that report evidence of ToM in non-human primates (NHPs) have used false-belief tests based on anticipatory looking at another agent's action[14–17]. However, these tests suffer from at least three limitations. First, it has been difficult to prove that such tests of ToM can only be explained by mentalistic accounts rather than simpler non-mentalistic accounts (construct validity)[2,3,18–21]. Second, false-belief tests can only provide a limited insight into ToM: they are sufficient to suggest ToM in the case of success, but are uninformative in case of failure[22,23], and such tests are not conceived to measure the level of sophistication of social reasoning[20,24]. Third, in this type of experiment the NHPs are often passively observing scenes of interacting individuals[14–16,25] rather than being actively engaged in the social interaction[26]. In addition, many ToM studies in NHPs resort to direct interactions with humans[1,27], thereby limiting their ecological validity and the motivation in the task[17,28]. Importantly, experiments that test ToM must result in behavioural responses that are not explainable without understanding others' knowledge, beliefs or understanding[29–31].

An approach to study ToM, which has had success in humans, has been to combine economic games with formal computational models of mentalising[4,5,8,24,32,33]. In this approach, if a model that incorporates a mentalizing process explains the data better than other candidate models that do not incorporate such a mentalizing mechanism, then the mentalizing term must have captured some unexplained variance of the behaviour. Here, we combined this computational modelling approach with a coordination task to a species of old-world monkeys, the Guinea baboon (Papio papio). We focused on coordination learning because it may constitute an essential process for non-human primates to coordinate group behavior in the wild[34–36]. We aimed to determine whether the computational mechanisms that best describe coordination learning of freely interacting baboons requires a ToM process. Such mechanistic descriptions of observed behaviour have the potential to address the first two challenges mentioned above. Indeed, a precise specification of the computational mechanisms allows us to distinguish mentalising from non-mentalising processes, and also differentiates different levels of sophistication of ToM. Moreover, our experimental set-up directly addresses motivational and ecological challenges by actively engaging baboons who freely and willingly interact with a conspecific, without any direct intervention by the experimenters.

We took advantage of a unique experimental set up in which a colony of 18 baboons, living together in a large enclosure with their offspring, have free access to Automated Social Learning Devices for Monkeys (ASLDM)[37] (**Erreur ! Source du renvoi introuvable.**). Baboons in the colony had previously participated in several other studies of social behaviour using ASLDM and were therefore familiar with this apparatus through which monkeys could see each other and interact in social games (see also **Methods, Baboons Housing**). If two baboons

entered paired adjacent workstations together (within a time delay of 5 s), this triggered the presentation of a coordination game, matching-pairs (MP), in which the same pair of icons from a large set were presented on touch screens to both baboons simultaneously, in randomized positions on the screen (Erreur ! Source du renvoi introuvable.).

In the Training condition, if the second baboon touched the icon that matched the one chosen by the first baboon, both were rewarded (Erreur ! Source du renvoi introuvable.**A**). During this initial learning phase, the identity of the partner baboon as well as their chosen icon could be seen by its partner through a transparent window (Erreur ! Source du renvoi introuvable.**C,D,E**, Erreur ! Source du renvoi introuvable.**A**). Baboons that achieved 80% successful coordination over 50 consecutive trials (n=7 out of 18) were designated to the experimental group and subsequently tested in a similar coordination game in two conditions. In a ”Social” condition, they could see their partner but an opaque screen prevented them from seeing which icon it chose (Erreur ! Source du renvoi introuvable.**B**). Thus, to coordinate, they had to infer the actions of their partner from past choices and outcomes. In the “Solo” condition, they could see that no interacting partner was present in the closed chamber next to them (Erreur ! Source du renvoi introuvable.**C**). In fact, in the Social condition, the baboons did not actually play with the partner baboon in the adjacent box. In both conditions they played with the same computer algorithm, referred to as the artificial agent (AA). The AA used their past choices and rewards history to cooperatively encourage coordination by promoting predictability, and therefore one of the best strategies to coordinate with the AA would be to maintain high choice consistency (see **Methods, Choice algorithm**). Thus, in the social condition, baboons were led to believe that they were interacting with a partner while in fact this partner was ‘fake’.

We interleaved two 200-trial blocks of the Social condition with two 200-trial blocks of the Solo condition (**Fig. 2C**). The Solo condition served as a control to allow us to compare the strategies baboons used when playing alone or with partners. The fact that in both situations they played with the same algorithm allowed us to attribute changes in choice behaviour to the presence or absence of a partner. Importantly, our experimental set-up directly addressed motivational and ecological concerns [1,19–21] by actively engaging baboons who freely and willingly came with a conspecific (in the learning and social conditions) to play. The relative dominance of each baboon in the group was measured through an ELO score (see **Methods, ELO score**), and the number of trials between each tested baboon and each other baboon in the Social task are indicated in the encounter matrix (Erreur ! Source du renvoi introuvable.**D**).

We hypothesized that if Guinea baboons are able to exhibit ToM processes, this would lead to: (a) higher coordination performance in the Social relative to the Solo condition because mentalizing processes are expected to be more engaged when baboons are led to believe they interact with another conspecific; (b) Higher choice stability in the Social (relative to the

Solo) condition to facilitate coordination by the playmate. Furthermore, we hypothesized that differences in dominance hierarchy between the two baboons in the Social condition would influence coordination performance and choice stability. Indeed, social dominance hierarchy shapes coordination behaviour in the wild to grant dominant individuals greater influence over decisions[38–40]. In addition, we tested whether a mechanistic approach can reveal the computational mechanisms underlying coordination in baboons. To do this, we tested learning algorithms known to describe non-mentalizing and mentalizing[4–11,24,32] processes in humans (see **Methods**).

First, we aggregated the raw data by social condition (Social vs Solo), as well as by dominance condition (dominant vs submissive) for each baboon and for the whole group. We plotted the coordination performance, i.e., the tendency to successfully coordinate with the AA (and thus, get rewarded) at trial n, $Reward_n$, for each baboon (**Erreur ! Source du renvoi introuvable.A, top)** and for the whole group (**Erreur ! Source du renvoi introuvable.B**). We also plotted the choice stability, i.e., the tendency to stick to one's previous choice at the current trial n, $Stay_n$, for each baboon (**Erreur ! Source du renvoi introuvable.A, bottom)** and for the whole group (**Erreur ! Source du renvoi introuvable.C**) in the Social and Solo conditions. We also plotted coordination performance for each baboon (**Erreur ! Source du renvoi introuvable.D, top)** and for the whole group (**Erreur ! Source du renvoi introuvable.E**) and choice stability for each baboon (**Erreur ! Source du renvoi introuvable.D, bottom)** and for the whole group (**Erreur ! Source du renvoi introuvable.F**) in the dominant and submissive conditions. These results point toward an effect of the Social vs Solo conditions and of the dominant vs submissive conditions on the individual as well as on the group levels.

To confirm the significance of this effect, we performed Generalized Linear Mixed-Effects (GLME) regressions including repeated measures across time and conditions. This analysis allowed us to identify the main predictors that drove baboons' behaviour. Therefore, we regressed the coordination performance (GLME model 1) and of the choice stability (GLME model 2) at the individual level with a random intercept and a random time slope for each baboon (see **Supp. Mat., GLMs et GLMEs**). These analyses take into account the non-independence of data within subjects and show that baboons achieved better coordination performance with the AA in the Social vs Solo condition (GLME model 1; coefficient: 0.363; standard error (SE): 0.130; z-value: 2.782; P<0.01**; Erreur ! Source du renvoi introuvable.**A**). They also indicate that baboons adapt their decision behaviour based on the social context by reaching higher choice stability in the Social vs Solo condition (GLME model 2; coefficient: 0.446; standard error (SE): 0.143; z-value: 3.129; P<0.01**; Erreur ! Source du renvoi introuvable.**B**). The influence of dominance hierarchy was also confirmed with these analyses. The data show that the difference in relative dominance of the baboons in the paired work stations was positively correlated with coordination performance (GLME model 1; coefficient:

0.216; standard error (SE): 0.073; z-value: 2.954; P<0.01**; Erreur ! Source du renvoi introuvable.**C**) and choice stability (GLME model 2; coefficient: 0.188; standard error (SE): 0.056; z-value: 3.356; P<0.001***; Erreur ! Source du renvoi introuvable.**D**). Thus, a given baboon achieved higher coordination performance with the AA and higher choice stability when it was dominant (vs submissive) within the dyad. This higher choice stability can be interpreted as a signal sent to the partner to communicate one's desire to stay on the same target, although this may incur a short-term cost for staying on this target[4,5,41]. Such behaviour is key to successfully coordinate[42] when agents want to trigger long-term reciprocity despite a potential short-term cost to promote cooperation from the partner[43].

The previous descriptive model analyses do not provide mechanistic understanding of the computational process that produce these differences in behaviours. Therefore, we tested a set of computational models to explain the baboons' decision behaviour mechanistically (see **Methods, Learning Models**). These models can be categorized in different classes based on the nature of their computational mechanisms: non-Bayesian vs Bayesian and non-mentalizing vs mentalizing. Bayesian (vs non-Bayesian) models use probabilistic representations of the world and infer posterior probabilities of actions from priors and observations of the environment with Bayes update rule[44,45]. Mentalising (vs non-mentalising) models incorporate a meta-representational mechanism of another individual's decision process[4,9,24,33]. Our set of candidate models comprises: (1) a Bayesian mentalising model, the 1-ToM model, which is a model of Theory-of-Mind (ToM) of depth 1[8,9] ; (2) Bayesian non-mentalising models that are called Bayesian Sequence Learning (BSL) models[10] of depth 0 and 1; (3) non-Bayesian non-mentalising models, namely a Win-Stay/Lose-switch (WSLS) strategy[11] and a reinforcement-learning based Q-Learning (QL) model[6,7] and (4) a non-Bayesian mentalising model, the Influence model[4,5], that exploits a RL-scheme that incorporates a mentalising mechanism (for more details on the models, see **Methods, Learning models**).

We anticipated that if baboons were capable of mentalising, then their behaviour in the Social, but not the Solo condition, might best be explained by a model that included a ToM mechanism. We found that, in the social condition, the Influence model[4] best explained the group's behaviour compared to all alternative models [Exceedance Probability EP=0.748] (Erreur ! Source du renvoi introuvable.**A**, social). This indicates that baboons most probably mentalised the influence of their own decision on that of the other baboon in the dyad. A previous report showed that some primate species (Orangutan, Gorilla, Chimpanzee) were better fitted by such influence models, rather than simpler RL models or heuristics[27]. In that study, however, only a few animals of each species (among which no baboon) were obliged to coordinate with a human confederate (a familiar caretaker), who applied a given strategy in a competitive game (Hide-and-Seek).

The analysis with GLMEs had revealed that the relative difference in dominance between baboons was an important factor for baboons during the Social condition. Therefore, any model explaining baboons' behaviour should account for the fact that the baboons' decision process was influenced by their relative dominance compared to the baboon in the paired workstation. We thus developed a new model derived from the influence model by allowing the relative difference in dominance between the tested baboon and its partner in each dyad, as measured by the difference between their ELO scores, to modulate any one of its learning parameters (see **Supp. Mat., Computational modelling of dominance**). This new ELO-influence model shows that when the tested baboon was more dominant in a dyad, its learning rate (η, eta parameter) was lower than when this tested baboon was the more submissive in a dyad. This result indicates that the relative difference between social dominance of the two baboons in each trial plays a key role in the decision process and specifically that more dominant individuals in the interacting dyads updated their prediction error less.

To test the social specificity of the Influence model, we also checked whether the same Influence model might fit the decision process in the Solo condition. Using Bayesian Model Selection, we found that in the Solo condition the WSLS explained behaviour better than all other models tested [Exceedance Probability EP=0.621] (Erreur ! Source du renvoi introuvable.**A**, solo). This suggests that baboons applied a more straightforward reward-based heuristic that did not rely on any ToM mechanism in the Solo condition, when there was no apparent partner to play with. A similar effect has previously been reported in humans who decipher intentional behaviour in a competitive strategic decision game (Hide-and-Seek) by attributing mental states to others *a priori* using sophisticated mentalising Bayesian learning models[8]. In contrast, in the current coordination game, baboons used a mentalizing non-Bayesian model in the social condition. Our findings also go beyond the intrinsic limits of previous studies that investigated strategic games combined with computational modelling in monkeys because these studies either used direct interactions between humans and monkeys, forced interactions between monkeys, or were very limited in the number of interacting pairs of monkeys[27,46,47]. Moreover, our computational mechanistic approach allows us to exclude alternative interpretations of our behavioural results in terms of higher attentional or arousal processes in the Social condition. Indeed, such explanations would not predict the use of distinct processes in the Social and Solo conditions.

We next tested the possibility that baboons employed the same Influence learning decision process in the Solo and Social conditions. To do so, we extracted and compared the learning parameters of the Influence model between the Social vs Solo condition. We found that the values of all the learning parameters were significantly lower in the Solo than Social condition (rmANOVA, Erreur ! Source du renvoi introuvable.**B**). This means that under the

assumption that the Influence model is used in both conditions, baboons would learn faster (higher learning rate, η (eta): Marginal Means=0.596 (Social), 0.394 (Solo); 95%CI=[0.486; 0.705] (social), [0.284; 0.504] (solo); F=15.2; P=0.024* (Holm-corrected)), attribute more influence of their choice on the outcome, (higher weight of their influence on the other, λ (lambda): Marginal Means=0.587 (Social), 0.354 (Solo); 95%CI=[0.475; 0.698] (Social), [0.243; 0.465] (Solo); F=20.0; P=0.02* (Holm-corrected)) and attribute less randomness to the other's decisions (higher inverse temperature of the other, β (beta): Marginal Means=2.67 (Social), 0.779 (Solo); 95%CI=[1.514; 3.823] (Social), [-0.376; 1.934] (Solo); F=18.4; P=0.02* (Holm-corrected)), in the social compared to the solo condition. Together, these findings reflect that the baboons have far greater uncertainty when coordinating with an "imaginary" partner in the Solo condition. This was despite the fact that the baboons were in fact playing against the same algorithm in both the Solo and Social situations. Finally, to check the robustness of the previous results, we performed a parameter simulation-recovery analysis to determine to what extent the influence model parameters are identifiable from observed data (see **Supp. Mat., Control checks**).

Although comparative research with non-human primates suggests that meta-representational ToM is a uniquely human capacity not shared by our closest living animal relatives[1,3,18,48], it has remained controversial to what extent more basic forms of ToM might be evolutionarily more ancient and shared with other species[1,17,20,21,28,47,49]. A primitive form of ToM is to mentalise one's exerted influence on the other's action[4,5,27,50,51]. Our results suggest that Guinea baboons are able to use this basic form of ToM to learn to coordinate, and characterize the computational mechanisms that give rise to this ability. Thus, ToM processes used for coordination learning may have evolved in the common ancestor of the Old-World monkey (Cercopithecidae) and apes (Hominoidea) that diverged 30 million years ago. Additionally, our findings indicate that for dyadic coordination, adaptive behaviour also depends on the relative difference between social dominance of the two baboons. Given the importance of adapting one's behaviour to others' social status for social interactions, such a learning mechanism could be selected for by evolution[38–40]. This corroborates evolutionary theories proposing that leadership can play a central role in the evolution of cooperation[52,53] and can facilitate coordination[40,47,54,55].

Author Contributions

J.C. D., N.C., E.D., T.N., J.F., R.P. designed the experiments. N.C. and J.F. implemented the experiment and collected the data. T.N. analysed the data and wrote the initial draft of the paper. T.N. , E.D., and J.C. D. wrote the final version of the paper. All authors contributed to analytical insights and commented on statistical tests, discussed the results and implications.

Author Information

The authors declare no competing interests. Correspondence and requests for materials should be addressed to T.N. (toan.nong@isc.cnrs.fr) and J.C. D. (dreher@isc.cnrs.fr).

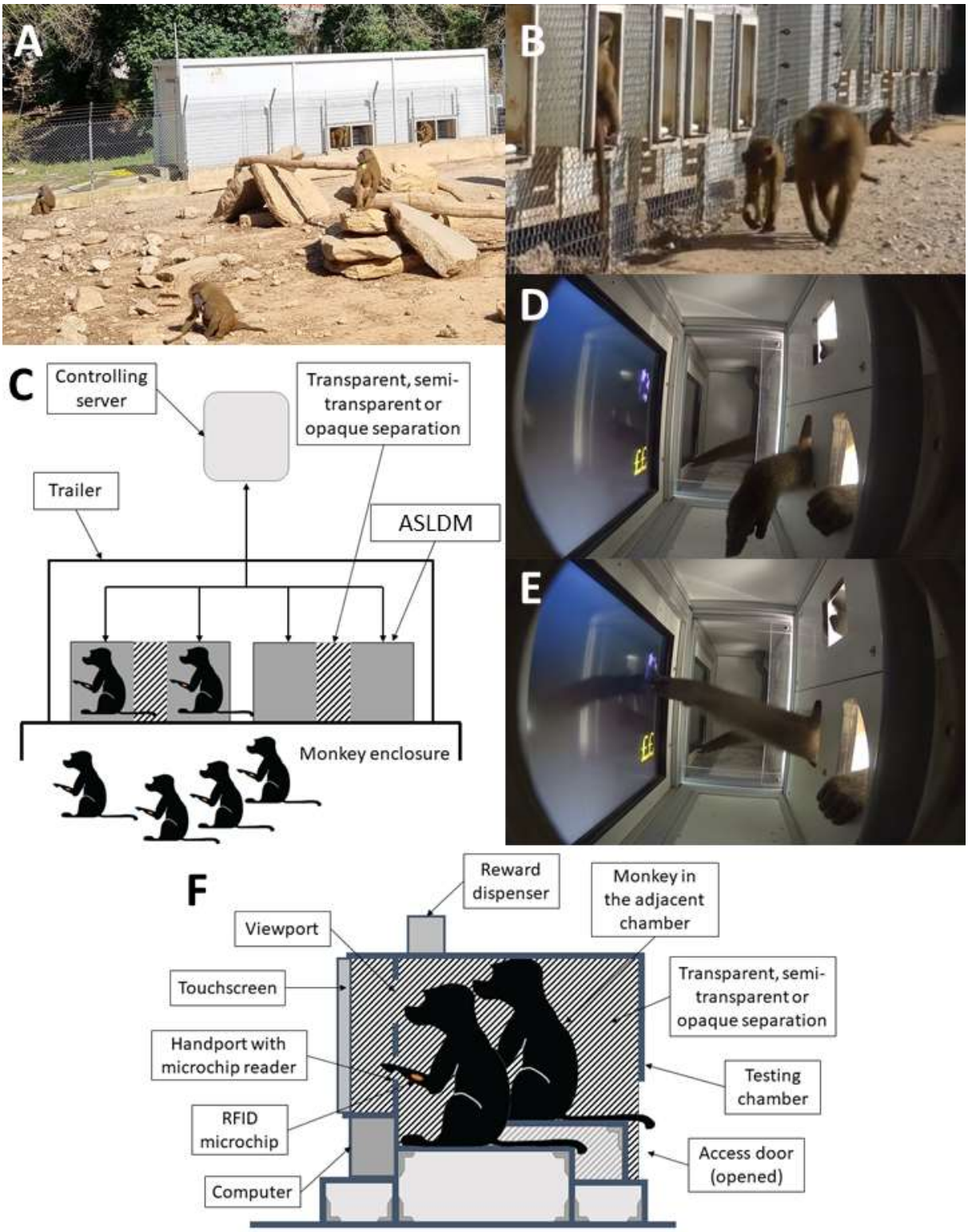


Figure 26. Experimental setup. The baboons live in a colony of 18 adults in a large enclosure with free access to Automated Social Learning Devices for Monkeys (ASLDM) (Large white trailer) which contain the workstations equipped with computer touch sensitive screens (to allow the baboons to play the economic games) and tag readers that automatically recognize which baboon is playing. **(A) Picture of the enclosure. (B) Picture of the entrance of the trailers housing the ASLDM testing units** (see **Methods, Baboons Housing**). **(C) Schematic of the S-ALDM system's organization.** Adapted from Claidière et al [37]. **(D-E) A baboon interacts with a touchscreen during the Training Phase.** The other

baboon’s arm is visible in the background through the transparent separation screen. **(F) Details of the ASLDM testing unit** (see **Methods, Baboons Housing**)**.**

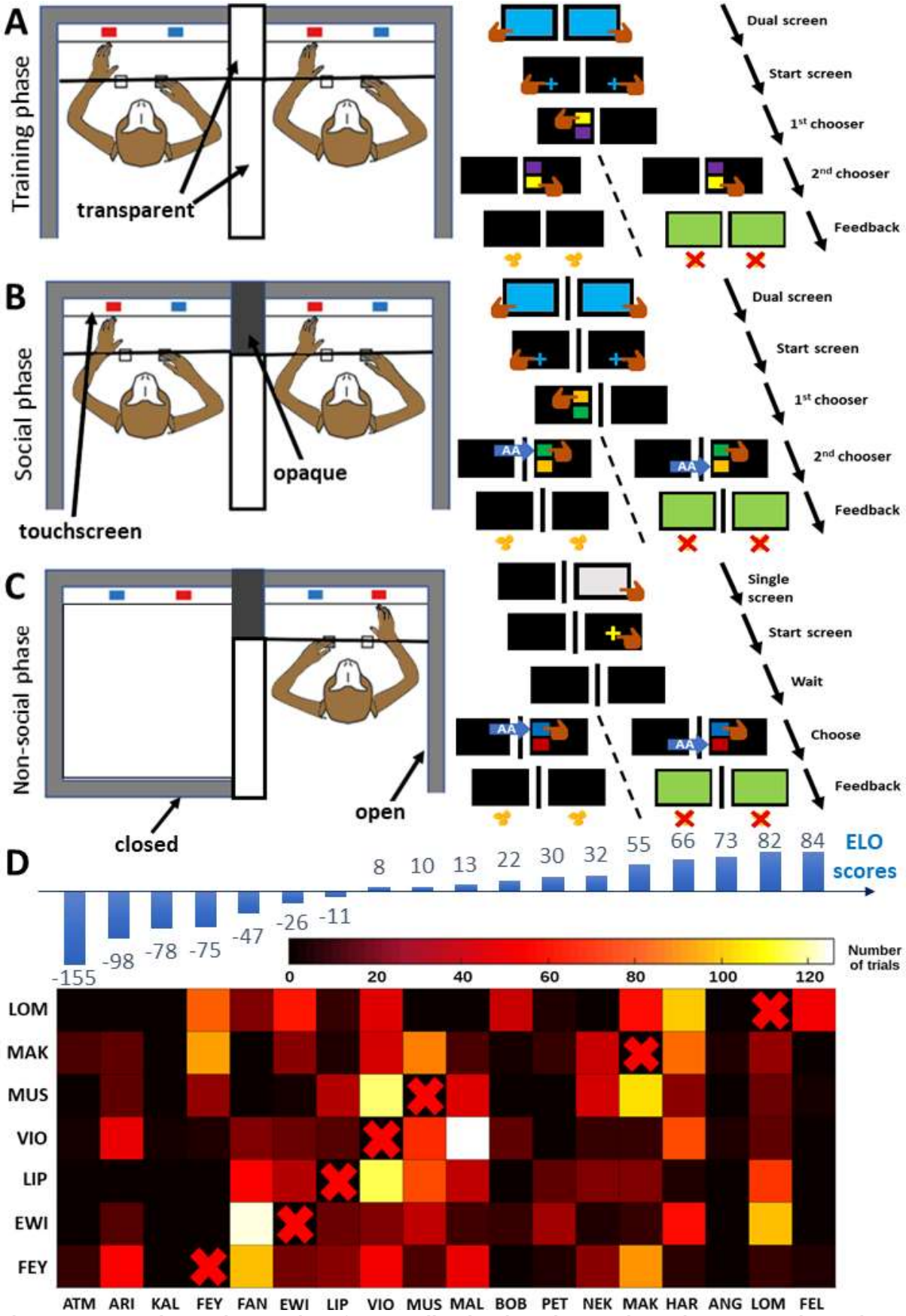


Figure 27. Experimental procedure for coordination learning, and matrix representing pairs of baboons coordinating. To be presented with the coordination learning task, two baboons were required to arrive at adjacent ASLDM together. Adapted from Möller et al[47]. **(A) Learning phase**:

Baboons learned to make coordinated responses. One baboon, as first chooser, pressed one icon from the two presented (randomized at each trial out of a possible menu of 20) on a touch screen. As second chooser, the partner baboon had to press the same icon from the two on their own screen for both baboons to receive rewards, wheat grains, if they coordinated (see **Methods, Learning phase**). **(B) Social condition**: baboons were required to arrive together and could see each other in adjacent ASLDM, but could not see each other's choices. The tested baboon was allocated the role of second chooser in any pair until it had completed 200 social block trials. The choices of the "first chooser" were irrelevant. Rewards were determined by an algorithm (AA, Artificial Agent), represented by the blue arrow, which determined which of the baboon's choices would be rewarded. Grains were distributed to both baboons if the tested baboon chose the same option as the AA. The pair of two icons was the same for the entire block of 200 trials (see **Methods, Social Condition**). **(C) Solo condition**: in Solo trial blocks the work station adjacent to the tested baboon was closed and the baboons could see there was no other baboon inside. The two icons were presented and the baboon had to choose one to coordinate with the same algorithm as was used for the Social condition. The tested baboon was rewarded only if he chose the same option as the AA. The pair of two icons was the same for the entire block of 200 trials (see **Methods, Solo Condition**). **(D) ELO dominance hierarchy and Pair meeting matrix.** Blue histograms above each matrix column show the hierarchical status of the corresponding baboon, as determined by ELO scores (see **Methods, Evaluation of Dominance Hierarchy**) from encounters between baboons. The matrix beneath shows the colour-coded frequency of partners for the social trials between each tested baboon (left-most column) and any other baboon in the colony. The partner baboon had not necessarily achieved the 80% success rate on the training trials, and therefore may not be included as 'participant' in the study. The red crosses represent pairings that are not possible, i.e., a baboon with themself.

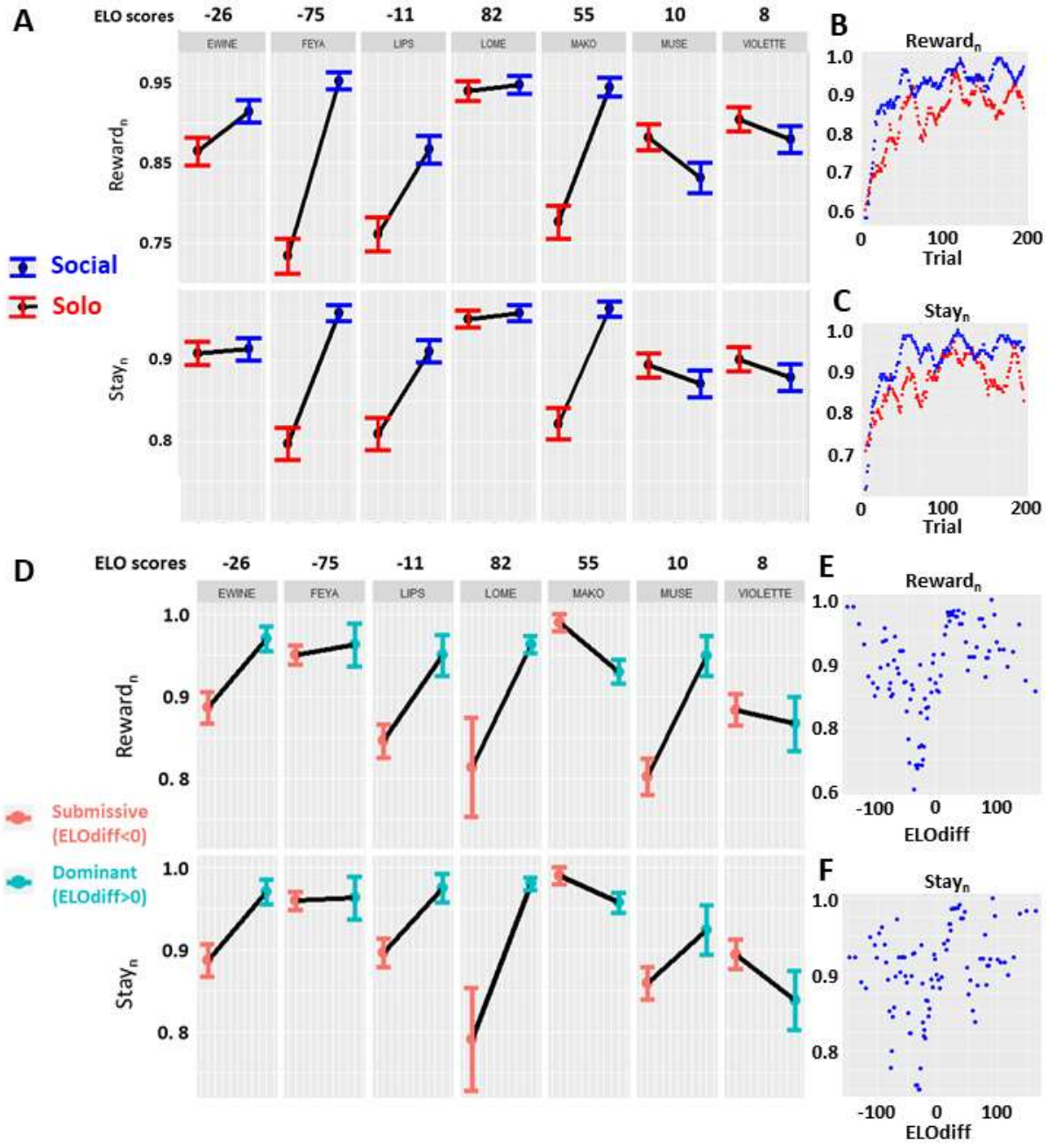


Figure 28. Influence of conditions (Social vs Solo) and dominance differences (ELO_{diff}) on baboons' behaviour. (A) Social vs Solo conditions. Coordination performance, i.e. the tendency to successfully coordinate with the AA, $Reward_n$ (top) and choice stability, i.e. the tendency to stick to the previous choice, $Stay_n$ (bottom) of each baboon at trial n, averaged over all sessions and trials. Each column represents a tested baboon that came to play with another baboon in the adjacent box. Error bars: Standard Error Means (S.E.M.) **(B) Coordination performance $Reward_n$ averaged by trial.** The data points are averaged on all baboons and sessions, and with a moving average window across 20 consecutive trials. **(C) Choice stability $Stay_n$ averaged by trial.** The data points are averaged on all baboons and with a moving average window across 20 consecutive trials. **(D) Submissive vs dominant conditions.** Coordination performance $Reward_n$ (top) and choice stability $Stay_n$ (bottom) of each baboon at trial n, averaged over all Social sessions and trials. Submissive condition: the tested baboon is more submissive than their partner in the adjacent box (ELOdiff<0). Dominant condition: the tested baboon is

more dominant than their partner in the adjacent box (ELOdiff>0). Error bars: S.E.M. **(E) Coordination performance Reward$_n$ averaged by difference in ELO scores ELO$_{diff}$.** The data points are averaged on all baboons and sessions, and with a moving average window across 5 consecutive points. **(F) Choice stability Stay$_n$ averaged by difference in ELO scores ELO$_{diff}$.** The data points are averaged on all baboons and sessions, and with a moving average window across 5 consecutive points.

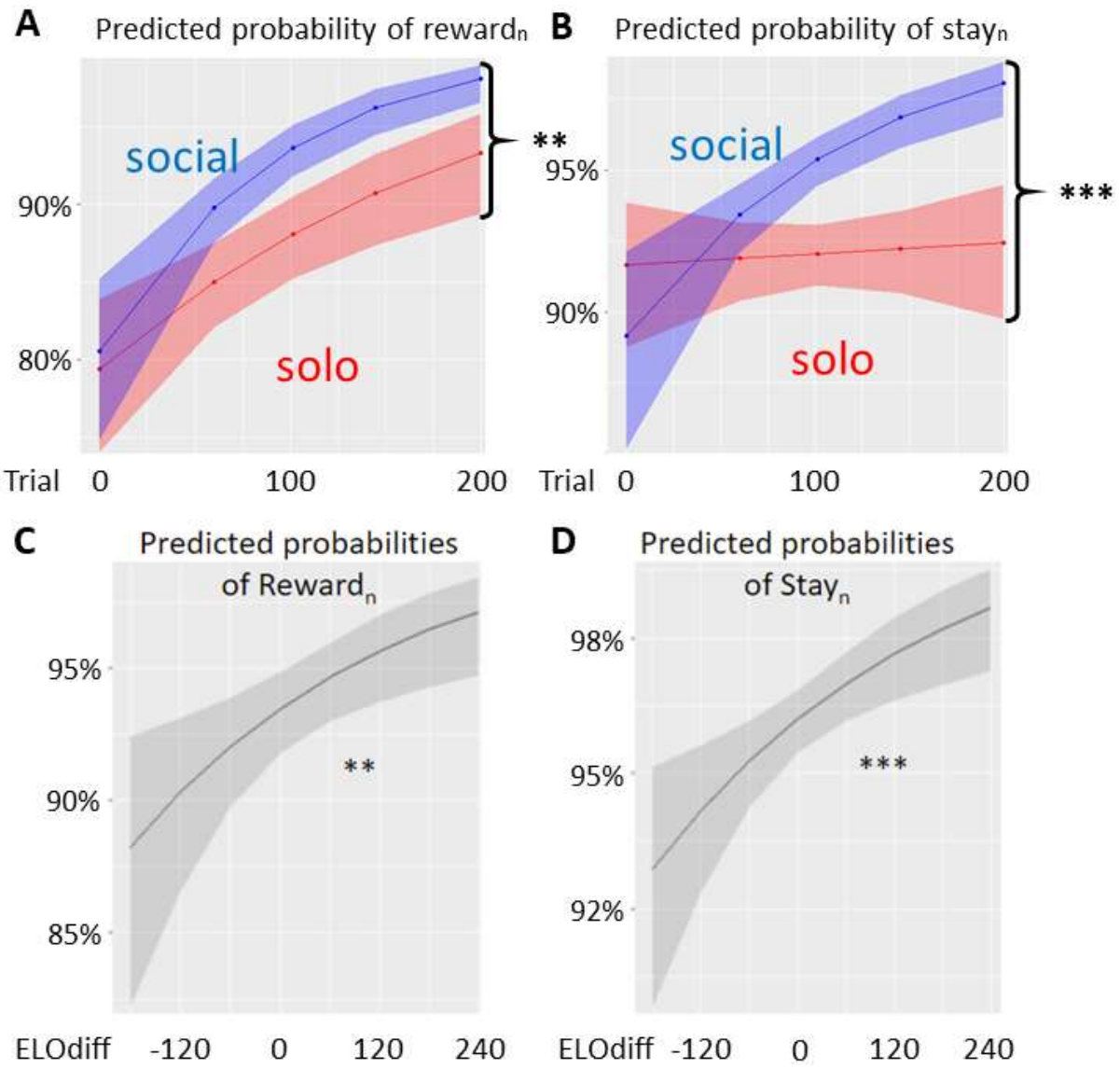


Figure 29. Correlational model analysis. Marginal Effects of Generalized Linear Mixed-Effects (GLME) models 1 and 2 depending on Trial and Condition, or relative Dominance (see **Methods, Generalized Linear Models**). **(A)** Predicted probability to coordinate at trial n, Reward$_n$, computed with GLME model 1, depending on Trial and Condition. Model prediction with standard deviations are in straight lines. **(B)** Predicted probability of staying (as opposed to switching) at trial n relative to trial n-1, stay$_n$, computed with GLME model 2, depending on Trial and Condition. Model prediction with standard deviations are in straight lines. (**C**) Predicted probability to coordinate at trial n, Reward$_n$, computed with GLME model 1, depending on relative Dominance measured with relative difference of ELO scores (see Methods). **(D)** Predicted probability of staying (as opposed to switching) at trial n relative to trial n-1, stay$_n$, computed with GLME model 2, depending on relative Dominance measured with relative difference of ELO scores (see **Methods, ELO scores**).

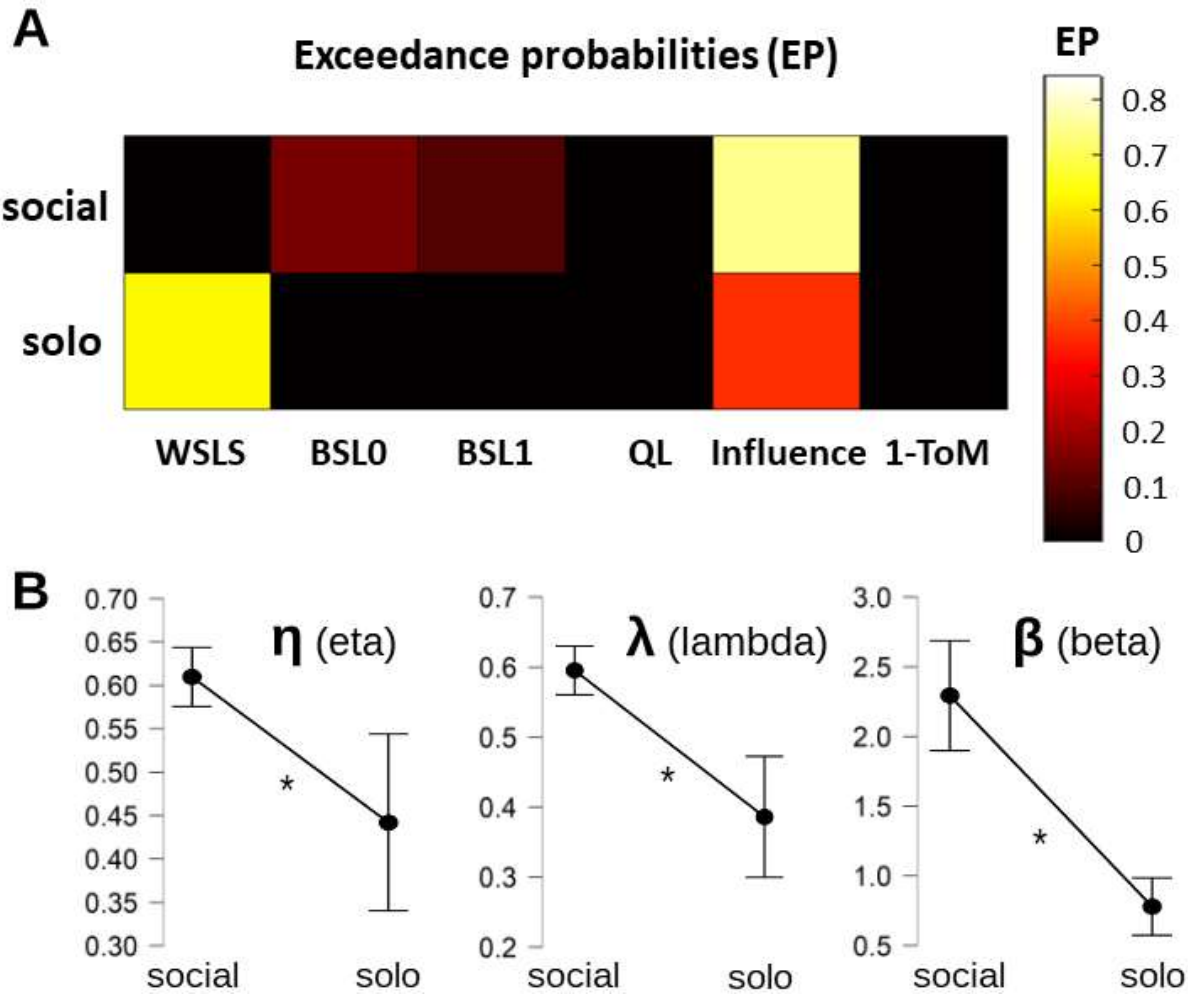


Figure 30. Results of computational modelling. (A) Bayesian Model Selections (BMS). X-axis: model names; Y-axis: social condition; Color-scale: Exceedance Probabilities (EP), i.e., the probability that one model is more frequent in the group population than any other model (see **Methods, Learning Models**). **(Social)** In the Social condition, the influence model is the best fit for our group of baboons. **(Solo)** In the Solo condition, the WSLS heuristic fits best. **(B) Fixed-Effect Model Analysis**. Baboons are all fitted by the Influence Learning model, for each session, in the Social and Solo Condition. Then, their estimated learning parameters are compared (rmANOVA). X-axis: condition; Y-axis: parameter value. Error bars: standard errors.

References

1. Call, J. & Tomasello, M. Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn Sci* **12**, 187–192 (2008).

2. Horschler, D. J., MacLean, E. L. & Santos, L. R. Do Non-Human Primates Really Represent Others' Beliefs? *Trends in Cognitive Sciences* vol. 24 594–605 Preprint at https://doi.org/10.1016/j.tics.2020.05.009 (2020).

3. Heyes, C. Animal mindreading: what's the problem? *Psychon Bull Rev* **22**, 313–327 (2015).

4. Hampton, A. N., Bossaerts, P. & O'Doherty, J. P. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci U S A* **105**, 6741–6746 (2008).

5. Hill, C. A. *et al.* A causal account of the brain network computations underlying strategic social behavior. *Nat Neurosci* **20**, 1142–1149 (2017).

6. O'Doherty, J. P., Cockburn, J. & Pauli, W. M. Learning, Reward, and Decision Making. *Annual Review of Psychology* vol. 68 73–100 Preprint at https://doi.org/10.1146/annurev-psych-010416-044216 (2017).

7. Zhu, L., Mathewson, K. E. & Hsu, M. Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proc Natl Acad Sci U S A* **109**, 1419–1424 (2012).

8. Devaine, M., Hollard, G. & Daunizeau, J. The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? *PLoS Comput Biol* **10**, (2014).

9. Daunizeau, J. *et al.* Observing the observer (I): Meta-bayesian models of learning and decision-making. *PLoS One* **5**, e15554 (2010).

10. Éltető, N., Nemeth, D., Janacsek, K. & Dayan, P. Tracking human skill learning with a hierarchical Bayesian sequence model. *PLoS Comput Biol* **18**, 1–28 (2022).

11. Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58 (1993).

12. Rakoczy, H. Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology* **1**, 223–235 (2022).

13. Taylor, D. *et al.* (submitted) Reading minds or reading scripts? De-intellectualising theory of mind. *Proceedings of the Royal Society B: Biological Sciences* **12**, (2023).

14. Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. Great apes anticipate that other individuals will act according to false beliefs. *Science (1979)* **354**, 110–114 (2016).

15. Hayashi, T. *et al.* Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex. *Cell Rep* **30**, 4433-4444.e5 (2020).

16. Kano, F., Krupenye, C., Hirata, S., Tomonaga, M. & Call, J. Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proc Natl Acad Sci U S A* **116**, 20904–20909 (2019).

17. Lewis, L. S. & Krupenye, C. Theory of Mind in Nonhuman Primates. in *Primate Cognitive Studies* (eds. Schwartz, B. L. & Beran, M. J.) 439–482 (Cambridge University Press, Cambridge, 2022). doi:10.1017/9781108955836.018.

18. Martin, A. & Santos, L. R. What Cognitive Representations Support Primate Theory of Mind? *Trends Cogn Sci* **20**, 375–382 (2016).

19. Royka, A. & Santos, L. R. Theory of Mind in the wild. *Curr Opin Behav Sci* **45**, 1–6 (2022).

20. van der Vaart, E. & Hemelrijk, C. K. 'Theory of mind' in animals: ways to make progress. *Synthese* **191**, 335–354 (2014).

21. Meunier, H. Do monkeys have a theory of mind? How to answer the question? *Neurosci Biobehav Rev* **82**, 110–123 (2017).

22. Penn, D. C. & Povinelli, D. J. On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. in *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 362 731–744 (2007).

23. Lurz, R. If chimpanzees are mindreaders, could behavioral science tell? Toward a solution of the logical problem. *Philos Psychol* **22**, 305–328 (2009).

24. Rusch, T., Steixner-Kumar, S., Doshi, P., Spezio, M. & Gläscher, J. Theory of mind and decision science: Towards a typology of tasks and computational models. *Neuropsychologia* **146**, 1–44 (2020).

25. Kano, F., Call, J. & Krupenye, C. Primates Pass Dynamically Social Anticipatory-Looking False-Belief Tests. *Trends Cogn Sci* **24**, 777–778 (2020).

26. Canteloup, C., Piraux, E., Poulin, N. & Meunier, H. Do Tonkean macaques (Macaca tonkeana) perceive what conspecifics do and do not see? *PeerJ* **2016**, e1693 (2016).

27. Devaine, M. *et al.* Reading wild minds: A computational assay of Theory of Mind sophistication across seven primate species. *PLoS Comput Biol* **13**, 1–25 (2017).

28. Krupenye, C. & Call, J. Theory of mind in animals : Current and future directions. *WIREs Cognitive Science* **10**, (2019).

29. Crockford, C., Wittig, R. M., Mundry, R. & Zuberbühler, K. Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current Biology* **22**, 142–146 (2012).

30. Hare, B., Call, J. & Tomasello, M. Do chimpanzees know what conspecifics know? *Anim Behav* **61**, 139–151 (2001).

31. Hare, B., Call, J., Agnetta, B. & Tomasello, M. Chimpanzees know what conspecifics do and do not see. *Anim Behav* **59**, 771–785 (2000).

32. Devaine, M., Hollard, G. & Daunizeau, J. Theory of mind: Did evolution fool us? *PLoS One* **9**, (2014).

33. Yoshida, W., Dolan, R. J. & Friston, K. J. Game theory of mind. *PLoS Comput Biol* **4**, e1000254 (2008).

34. Samuni, L., Preis, A., Deschner, T., Crockford, C. & Wittig, R. M. Reward of labor coordination and hunting success in wild chimpanzees. *Commun Biol* **1**, 1–9 (2018).

35. Strandburg-Peshkin, A., Papageorgiou, D., Crofoot, M. C. & Farine, D. R. Inferring influence and leadership in moving animal groups. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 373 Preprint at https://doi.org/10.1098/rstb.2017.0006 (2018).

36. Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D. & Crofoot, M. C. Shared decision-making drives collective movement in wild baboons. *Science (1979)* **348**, 1358–1361 (2015).

37. Claidière, N., Gullstrand, J., Latouche, A. & Fagot, J. Using Automated Learning Devices for Monkeys (ALDM) to study social networks. *Behav Res Methods* **49**, 24–34 (2017).

38. King, A. J. & Cowlishaw, G. Leaders, followers and group decision-making. *Communicative and Integrative Biology* vol. 2 147–150 Preprint at https://doi.org/10.4161/cib.7562 (2009).

39. King, A. J., Douglas, C. M. S., Huchard, E., Isaac, N. J. B. & Cowlishaw, G. Dominance and Affiliation Mediate Despotism in a Social Primate. *Current Biology* **18**, 1833–1838 (2008).

40. Van Vugt, M., Hogan, R. & Kaiser, R. B. Leadership, followership, and evolution: Some lessons from the past. *American Psychologist* **63**, 182–196 (2008).

41. Behrens, T. E. J., Hunt, L. T. & Rushworth, M. F. S. The computation of social behavior. *Science* vol. 324 1160–1164 Preprint at https://doi.org/10.1126/science.1169694 (2009).

42. Glover, S. & Dixon, P. The role of predictability in cooperative and competitive joint action. *J Exp Psychol Hum Percept Perform* **43**, 644–650 (2017).

43. Rilling, J. K. *et al.* A neural basis for social cooperation. *Neuron* **35**, 395–405 (2002).

44. Nassar, M. R., Wilson, R. C., Heasly, B. & Gold, J. I. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience* **30**, 12366–12378 (2010).

45. Khalvati, K. *et al.* Modeling other minds: Bayesian inference explains human choices in group decision-making. *Sci Adv* **5**, (2019).

46. Ong, W. S., Madlon-Kay, S. & Platt, M. L. Neuronal correlates of strategic cooperation in monkeys. *Nat Neurosci* **24**, 116–128 (2021).

47. Möller, S. *et al.* Human and macaque pairs employ different coordination strategies in a transparent decision game. *Elife* **12**, (2023).

48. Heyes, C. Apes Submentalise. *Trends Cogn Sci* **21**, 1–2 (2017).

49. Formaux, A., Sperber, D., Fagot, J. & Claidière, N. Guinea baboons are strategic cooperators. *Sci Adv* **9**, (2023).

50. Barnby, J. M., Dayan, P. & Bell, V. Formalising social representation to explain psychiatric symptoms. *Trends in cognitive sciences* vol. 27 317–332 Preprint at https://doi.org/10.1016/j.tics.2022.12.004 (2023).

51. Barnby, J. M. *et al.* Beyond Theory of Mind: A formal framework for social inference and representation. *PsyXiv* (2023).

52. King, A. J., Johnson, D. D. P. & Van Vugt, M. The Origins and Evolution of Leadership. *Current Biology* vol. 19 R911–R916 Preprint at https://doi.org/10.1016/j.cub.2009.07.027 (2009).

53. Van Vugt, M. & Smith, J. E. A Dual Model of Leadership and Hierarchy: Evolutionary Synthesis. *Trends in Cognitive Sciences* vol. 23 952–967 Preprint at https://doi.org/10.1016/j.tics.2019.09.004 (2019).

54. Halevy, N., Y. Chou, E. & D. Galinsky, A. A functional model of hierarchy. *Organizational Psychology Review* **1**, 32–52 (2011).

55. Ronay, R., Greenaway, K., Anicich, E. M. & Galinsky, A. D. The Path to Glory Is Paved With Hierarchy: When Hierarchical Differentiation Increases Group Effectiveness. *Psychol Sci* **23**, 669–677 (2012).

Materials and Methods

Baboons Housing

The baboons belonged to a group of eighteen Guinea baboons (Papio papio) from the CNRS primate Centre, Rousset-sur-Arc, France. They are housed in a large enclosure (Fig. 1) (25 x 30 m2 outdoor enclosure connected to an indoor area and to two experimental trailers providing access to computerized Automated Social Learning Devices for Monkeys (ASLDM) test systems [1–3]. The group had ad libitum access to water, and feeding was provided daily at 5 pm. The baboons are free to participate (or not) in different types of learning experiments or economic games to receive food rewards. The games are presented on computer touch screens to which the baboons have free access when experiments are being run. The previous versions of ASLDM test systems are described in detail in previous studies [1,4–6]. Briefly, each ASLDM workstation consisted of a 70 cm × 70 cm × 80 cm test chamber equipped with an RFID microchip scanner to identify the baboon inside thanks to RFID microchips implanted subcutaneously in their arms, a 19-inch computer touch screen displaying the task, and a food dispenser to deliver a reward inside the test chamber when the subject responded correctly to the task. Often several different experiments are in progress at any one time and the game that a particular baboon or pair of baboons must play to receive rewards is signalled to them by a screen colour code. The game they play as well as the baboons' responses, reaction times and rewards are signalled, recorded and controlled automatically. Finally, the extent to which a baboon can see or observe the decisions of the baboon in an adjacent ASLDM can be controlled by opaque or transparent screens that may allow (or not) a baboon to observe its partner/competitor and or the partners decisions in social games. In the Solo condition of the coordination game the workstations adjacent to that in which a tested baboon was playing were closed so that it was evident to the tested baboon that there was no partner baboon inside.

The coordination game (CG)

Learning phase

To play the coordination game (Learning phase and Social condition) two baboons had to arrive together at two adjacent ASDLM (within a 5 second delay). The Learning phase would then initiate. At the beginning of each trial, a coloured screen that would indicate the nature of the task, i.e. which game they were supposed to play and its social nature, first appears. The game was a simple two-by-two (two players two choices) coordination game in which the players had the choice between two different targets on their screens. If both baboons chose the same target, they were both rewarded. Otherwise, neither was. Two target images from a panel of 20 possible targets were randomly displayed one above the other to each baboon for each trial.

One of the baboons was randomly assigned the role of First Chooser while the other was Second Chooser. They maintained their roles for the following trials for as long as the dyad continued to play together. The First Chooser would see their targets appear first, right after both baboons have touched the coloured screen to initiate the trial. In the meantime, the Second Chooser would wait before a black screen before their turn. Once the First Chooser has chosen, the other baboon's targets appear on the screen of the Second Chooser. Each baboon has a 2s timeframe to make their choice, otherwise the trial is cancelled and another one starts. Baboons had to touch the same targets on their own computer screens to be rewarded, and were able to watch the choices of their partner. The condition for a baboon to pass the Training Phase and be included for the study was to coordinate successfully as the Second Chooser at least 80% of the time on a block of 50 consecutive trials. This protocol

served to identify baboons that understood they were required to touch the same target as their partner in order to be rewarded. Baboons were allowed 4 weeks to achieve this criterion. The 7 baboons that passed this phase were then further tested over two trial blocks of 200 trials each of the Social condition, interleaved with two 200 trial blocks of the Solo condition (always Social, Solo, Social, Solo for each baboon).

Social Condition

In this condition, as for the Learning phase, two baboons, one of which was the tested baboon, were required to arrive together (within a 5 s delay), at adjacent ASLDM. The same coloured screen as that used during the Learning phase appears at the beginning of each trial to signal that a social partner is also present in the adjacent ASLDM. In the Social condition, the two baboons were shown the same two icons on each trial for the 200-trial block, and were required to select one of them by touching it. The icons changed between the first and second Social blocks. These were new icons that had not been used in the panel for the Learning phase. However, in the Social condition although the baboons could see their partner in the adjacent ASLDM, an opaque screen prevented the baboons from seeing the icon actually selected by the partner. Thus, to be rewarded, the tested baboon would seek to guess the icon selected by the partner. The baboons did not know that in fact the correct icon, which would be rewarded, was in fact selected by the algorithm (see below). Since at any one time only a single baboon was tested, the rewards of its partner were yoked to the rewards won by the tested baboon. Whenever the tested baboon entered an ASLDM together with another baboon in the adjacent ASLDM, the Social Condition trial block started/continued for the tested baboon, for as long as they remained playing, until the 200-trial block was completed. Then another baboon was selected as the tested baboon. If a session between two baboons extended over the 200-trial limit of the tested baboon's trial block, the baboons continued to be rewarded as a function of their responses, however, the trials over the 200-trial limit were ignored for the purposes of analysis. In this way Social trial blocks often stretched over several sessions with different partners for each participating baboon, because baboons could either be interrupted by other baboons displacing them from the ASLDM or could decide to interrupt the session to do something else. The Solo condition was not initiated for any baboon until the 7 participating baboons had all completed their 200 trial Social condition block. After they completed the first Solo condition 200 trial block for all the baboons they each performed a second 200 trial Social condition block.

Solo Condition

Once the 7 baboons had completed the first Social condition trial block the first Solo condition trial block began for each of them. The Solo condition was initiated when one of the 7 participant baboons arrived alone at an ASLDM with the adjacent ASLDM closed. The baboons could see there was no baboon in the adjacent work station, nevertheless, the touch screen colour code indicated that they were to play the coordination game to win rewards. The baboons were presented two novel icons (the same two for the whole 200 Solo condition trial block and different from any of the icons in the previous phases) on the touch screen and were required to touch one of them. The icons changed between the first and second Solo blocks. If they chose the one that was indicated by the same algorithm as made the target choices in the Social condition, they were rewarded. When the Solo condition was completed for the 7 participating baboons the second Social trial block with a single tested baboon began. When the seven participating baboons completed the second Social trial block the second (and final) Solo trial block began.

Choice algorithm (Artificial Agent)

The AA calculated the probability p for the baboon to select a particular card based on the history of the two previous choices and their outcomes. It then chose the card that it deemed the most probable to be chosen by the participant (the agent is indeed cooperative). The artificial agent (AA) selected its card according to the probability for the player to choose a specific card after a given sequence of events. By sequence of events, we mean a specific set of two consecutive choices and outcomes. If we consider two choices, A and B, and two possible outcomes, win (W) and lose (L), an example for a sequence of events could be: player chose A and won at trial n-2, then chose A and lost at n-1. We will sum up this sequence as: AWAL.

To determine its choice at trial n, the AA will check in its memory table what choice the participant usually followed with after such sequence of events, based on the previous history of actions of the player. For example, let's assume that the sequence of events AWAL just occurred. The AA must now decide its move at trial n. Let's also assume that this sequence of events already happened four times in the past, and that the player chose to play card A 75% of the time and card B 25% of the time. Then, the algorithm in the cooperative mode will decide to play card A with a 75% probability for its next move, and card B with a 25% probability. Vice-versa in the competitive mode, i.e. card A with 25% of probability and card B with 75% of probability. If we assume that the AA as well as the player choose card A for trial n, then the updated history of choices for the specific sequence of events AWAL would become: card A chosen at 80%, and card B at 20%. Also, the current sequence of events will become ALAW.

Let's assume that the current memory table is the one shown in Supp. Tab. 1. If we take the same example as described in the previous paragraph, the current history of events is "…AWAL…". After the AA and participant choose A, the history of events becomes "…AWALAW" and the memory table becomes Supp. Tab. 2. Now let's assume that the sequence history is "…AWALAWBLBLAWBW…". The sequence of the subject's choices in this example would be "…AAABBAB…". Consequently, in this example, the AA's choices would be "…ABAAAAB…". Following the same reasoning as presented in the previous paragraph, we can update the memory table. It will become at the end of this sequence of event the memory table shown in Supp. Tab. 3.

For the AA, the sequence of choices and probabilities would be "…A(/)B(/)A(75%)A(100%)A(100%)A(100%)B(50%)…", where the percentage of probability of the choice of the algorithm is given in parentheses and where "/" are undefined probabilities in this example because they depend on previous events before this sequence occurred and that we did not define. Consequently, the best strategy to coordinate with the AA would be to become as predictable as possible. One of the possible ways to do so would be to maintain high choice stability.

A_WA_W A: 1 ; B: 0	**A_WA_L** A: 3 ; B: 1	**A_LA_W** A: 1 ; B: 0	**A_LA_L** A: 1 ; B: 2
B_WB_W A: 0 ; B: 0	**B_WB_L** A: 0 ; B: 0	**B_LB_W** A: 0 ; B: 2	**B_LB_L** A: 1 ; B: 0
B_WA_W A: 0 ; B: 0	**B_WA_L** A: 10 ; B: 2	**B_LA_W** A: 0 ; B: 0	**B_LA_L** A: 0 ; B: 0
A_WB_W A: 5 ; B: 2	**A_WB_L** A: 3 ; B: 0	**A_LB_W** A: 0 ; B: 0	**A_LB_L** A: 2 ; B: 3

Supp. Tab. 1: Example of a memory table of the history of the sequences of events considered by the algorithm. The algorithm (Artificial Agent) keeps a history of the choices of the player after a given sequence of events that occurred during a session. In each cell, the sequence of events is **in bold**. An example of a history of the player's following choice is given below the sequence of events. "A: 1 ; B: 0" means "After the sequence of events in bold, the player chose card A once and never card

B until now". The algorithm updates this table at each trial and chooses its card based on it. **W = Win ; L = Loss ; A = Card A chosen ; B = Card B chosen.**

A_WA_W A: 1 ; B: 0	**A_WA_L** A: 4 ; B: 1	**A_LA_W** A: 1 ; B: 0	**A_LA_L** A: 1 ; B: 2
B_WB_W A: 0 ; B: 0	**B_WB_L** A: 0 ; B: 0	**B_LB_W** A: 0 ; B: 2	**B_LB_L** A: 1 ; B: 0
B_WA_W A: 0 ; B: 0	**B_WA_L** A: 10 ; B: 2	**B_LA_W** A: 0 ; B: 0	**B_LA_L** A: 0 ; B: 0
A_WB_W A: 5 ; B: 2	**A_WB_L** A: 3 ; B: 0	**A_LB_W** A: 0 ; B: 0	**A_LB_L** A: 2 ; B: 3

Supp. Tab. 2: First updated memory table of the history of the sequences of events. The algorithm (Artificial Agent) keeps a history of the choices of the player after a given sequence of events that occurred during a session. In each cell, the sequence of events is **in bold**. If we consider the sequence of events "…$A_WA_LA_W$…", this is the updated table from **Supp. Tab. 7**. **W = Win ; L = Loss ; A = Card A chosen ; B = Card B chosen.**

A_WA_W A: 1 ; B: 0	**A_WA_L** A: 4 ; B: 1	**A_LA_W** A: 1 ; B: 1	**A_LA_L** A: 1 ; B: 2
B_WB_W A: 0 ; B: 0	**B_WB_L** A: 0 ; B: 0	**B_LB_W** A: 0 ; B: 2	**B_LB_L** A: 2 ; B: 0
B_WA_W A: 0 ; B: 0	**B_WA_L** A: 10 ; B: 2	**B_LA_W** A: 0 ; B: 1	**B_LA_L** A: 0 ; B: 0
A_WB_W A: 5 ; B: 2	**A_WB_L** A: 3 ; B: 1	**A_LB_W** A: 0 ; B: 0	**A_LB_L** A: 2 ; B: 3

Supp. Tab. 3: Last updated memory table of the history of the sequences of events. The algorithm (Artificial Agent) keeps a history of the choices of the player after a given sequence of events that occurred during a session. In each cell, the sequence of events is **in bold**. If we consider the sequence of events "…$A_WA_LA_WB_LB_LA_WB_W$…", this is the updated table from **Supp. Tab. 7**. **W = Win ; L = Loss ; A = Card A chosen ; B = Card B chosen.**

Evaluation of Dominance Hierarchy

ELO scores were measured by the ASLDM system as described in previous studies[1,7]. Spatial use of ten ASLDM test systems provides information on the social structure of the baboon colony [8]. The social network inferred from the co-presence of individuals within the spatially adjacent workstations correlated strongly and reliably with the social network inferred from the affiliative behaviours (e.g., grooming or greeting behaviours) observed inside the enclosure in which the group of baboons lived. More particularly, it has been demonstrated that the use of ASLDM systems also provides information on dominance hierarchy [9]. Supplanting behaviours, defined as a sequence of behaviour in which one animal A approaches a workstation occupied by another animal B, who then leaves, are interesting behavioural markers of the dominance hierarchy.

ELO scores

ELO scores [10–14] were measured by the ASLDM system according to the protocol described by Claidière et al. and Ballesta et al. [1,7]. It roughly consists of following when one baboon displaces another from the work station to determine a relative hierarchical status by ELO score. The

Elo-rating method [15] was used to calculate dominance hierarchy. Briefly, each individual started with the same score and at each new contest the score of the winner increased by a certain amount and decreased for the loser by the same amount. To calculate the dominance hierarchy, we used RandomElo function from the Elo-rating package [16] that ran on the statistical environment R (R version 4.0.2; R Core Team, 2015), and followed the procedure described by Neumann & Kulik (2014)[17], with starting value of 1000 and a constant k optimized using maximum log-likelihood.

Analysis of Behavioral Data

Analyses of behavioural data were performed in R (version 4.0.2, for linear mixed-effects modelling: lme4 version 1.1-27). For computational modelling of behavioural data, we used MATLAB 2018b, and the VBA toolbox (v1.9.2). For the rmANOVA on the learning parameters, it was performed on JASP with the default priors from the JASP software.

Generalized Linear Models

Generalized linear mixed-effects models were computed with the lme4 package in R and ensured that we had good estimates of random effects, and accounted for variability in behaviour using Bound Optimization by Quadratic Approximation. Our GLME models showed a main effect of past choices and outcomes on the decision of the baboons (see **Supp.Mat., Supp.Tab.3**). Marginal effects[18] were computed and then plotted (see **Supp.Mat., GLMs&GLMEs**). For the linear regressions, reported “estimate” represents the regression coefficient.

Learning Models

Models were individually fit using Variational Based method with the VBA toolbox. All priors were set to their default values All priors were set to default value except for the sigma priors (covariance matrix of priors) that were set to 10 times the identity matrix.. With this method, we were able to find free parameters that minimized the free energy of the model[19].

The Bayesian Model Selection (BMS) was performed using the VBA toolbox (Variational Bayesian Analysis) in a random effect analysis relying on the free energy as the lower bound of model evidence. We use Exceedance Probability measurements (EP) [20] to select the model which was used most frequently in our population.

The models that we tested can be categorized in different classes based on the nature of their computational mechanisms: non-Bayesian vs Bayesian and non-mentalizing vs mentalizing. Bayesian (vs non-Bayesian) models use probabilistic representations of the world and infer posterior probabilities of actions from priors and observations of the environment with Bayes update rule[21–23]. Mentalising (vs non-mentalising) models incorporate a meta-representational mechanism of another individual’s decision process[24–27]. Our set of candidate models comprises: (1) a Bayesian mentalising model, the 1-ToM model, which is a model of Theory-of-Mind (ToM) of depth 1 and which belongs to the class of k-ToM models[24,28,29] ; (2) Bayesian non-mentalising models, Bayesian Sequence Learning (BSL) models[30] of depth 0 and 1; (3) non-Bayesian non-mentalising models, namely a Win-Stay/Lose-switch (WSLS) strategy[31] and a reinforcement-learning based Q-Learning (QL) model[32,33] and (4) a non-Bayesian mentalising model, the Influence model[26,34], that exploits a RL-scheme in that incorporates a mentalising mechanism.

k-ToM models

We will give here only the intuition behind k-ToM models. For detailed mathematical description of the model, refer to Devaine et al [28,29]. k-ToM models use recursive Bayesian inferences of depth k to predict the future choice of the other player. In brief, the repeated observation of the other's actions, noted a_{other}, gives the agent the opportunity to learn the other's behavioral tendency p_{other}.

At the lowest level of sophistication, i.e. k=0, the agent does not attribute any mental state to the other, but rather tracks her behavioral tendency p_{other} without mentalizing. They simply assume that their opponents choose their actions with a probability that varies in time with a certain volatility. They keep track of this probability by updating their prior estimation with Bayes' rule at each trial. Note that this would equate to a BSL0 agent.

ToM comes into play for higher recursion depth k>0, i.e. higher levels of sophistication, when k-ToM agents consider that the other is also a Bayesian agent whose decision policy p_{other} is driven by her hidden beliefs and desires. Distinct recursion levels will induce differences in the way agents update their subjective prediction p_{other}. For example, a 1-ToM agent considers that she is facing a 0-ToM agent, and will therefore predict 0-ToM's next move, based on her inference on the 0-ToM agent's beliefs and the choices' payoffs of both players. Eventually, a 1-ToM agent will learn how a 0-ToM agent learns about herself, and act accordingly. Similarly, a n-ToM agent will assume that they are facing a (n-1)-ToM agent and adapt accordingly.

Bayesian Sequence Learning:

The BSL model [21,23,30] learns online the structure of sequences, i.e. patterns of outcomes that are specified in terms of transition probabilities, by relying on Bayes' rule. It tracks the (log) odds of $P(o_t=1|o_{t-K})$, where u is the partner's move (binary outcome), K the depth of the sequence reasoning. This variable is updated according to a Laplace-Kalman filter [35], yielding 2 sufficient statistics (mean and variance) per combination of past outcomes. The BSL model can learn sequences of arbitrary depth K. For example:

- if K=1, then BSL tracks 2 probabilities, namely: $P(u_t=1|u_{t-1}=1)$ and $P(u_t=1|u_{t-1}=0)$. In this case, BSL needs to know about the other's previous move u_{t-1}.
- if K=2, then BSL tracks 4 probabilities, namely: $P(u_t=1|u_{t-1}=1, u_{t-2}=1)$, $P(u_t=1|u_{t-1}=0,u_{t-2}=1)$, $P(u_t=1|u_{t-1}=1,u_{t-2}=0)$ and $P(u_t=1|u_{t-1}=0,u_{t-2}=0)$. In this case, BSL needs to know about the other's two previous moves u_{t-1} and u_{t-2}. etc.

More generally, the BSL model tracks 2^K probabilities of the other's choices.
In this scheme, the only evolution parameter is BSL's prior volatility about the (log) odds.
The choice of the BSL agent (a=1 or a=0) is then based upon its prediction of its partner's next move (o=1 or o=0), given the game payoff table. The evolution function takes 1 parameter which is the BSL's prior volatility about the environment and the observation function takes 2 parameters which are the agent's temperature, which represents the agent's tendency to explore vs exploit, and the agent's bias towards one choice.

Win-Stay Lose-Switch (WSLS) model

This heuristic [31] keeps the same target as the previous one if previous reward was positive, and changes targets otherwise.

Q-learning:

The constituents of this classical model[32,33,36–38] in its simplest form are:

- a set of (action/item) value states. In two-armed bandit problems, there are two of these (n in a n-armed). In general, there will be as many values as there are available actions. Some behavioral biases can be captured by the initial conditions on these hidden states.
- a learning rate. This parameter controls the impact of reward prediction error on the value update. Note that one may want to ask whether the learning rate depends upon experimental factors (pathological condition, gain/loss domains, etc…)
- a behavioral temperature and a bias. These parameters control the exploitation vs exploration ratio of the agent, and capture choice randomness.

Q-learning models simply assume that subjects update the value of possible actions. In its simplest form, the Q-learning algorithm expresses the change in value Q(t+1)-Q(t) from trial t to trial t+1 as being linearly proportional to the prediction error. This yields the following learning rule:

$$Q_{t+1} = Q_t + \alpha * (r_{t+1} - Q_t)$$

Where rt is the reward delivered to the subject at trial t, and α is the (unknown) "learning rate" of the subject.

The Q-learning evolution function thus takes the agent's previous action and the feedback received for the previous action as data inputs. We complement with a softmax decision rule in the Q-learning observation function. As a reminder, this is an equation that expresses the probability Pt(ai) of the subject to choose action ai at trial t:

$$P_t(a_i) = \frac{\exp(\beta * Q_t(a_i))}{\sum_j \beta * Q_t(a_j)}$$

Where β is the (unknown) temperature.

Fitting the above Q-learning model to behavioral data means finding estimates of the learning rate α, the inverse temperature β, and the initial values Q_0 that best explains the observed choices.

Influence model

This mentalizing model was described and adapted from Hampton et al, who also identified neural correlates with this model in humans [26,34]. The influence model relies on a Taylor expended reinforcement learning. This model describes computations underlying the capacity to mentalize (in the context of a strategic game). Let $p_n = P_n(o = 1)$ be the agent's prediction of the other's next move, i.e., the probability that the other will pick the first alternative option. The "influence learning" rule can be written as follows:

$$p_{n+1} = p_n + \eta * (o - p_n) - \lambda * k_1 * p_n * (1 - p_n) * (a - \frac{\beta * x_n - k_2}{k_1})$$

where o is the other's last move, a is the agent's last move, η is the weight of the agent's prediction error, λ is the weight of the other's prediction error, k_1 & k_2 are derived from the game's payoff table, x_n is the tracked log-odds of p_n, β is the other's temperature, i.e. the tendency of the other to explore vs exploit information (the closer it is to 0, the more it explores by choosing randomly without using previous information, and the closer it is to 1, the more it exploits previous information and chooses deterministically).

ELO-Influence (ELO-modulated influence model):

The ELO-influence model is inspired from the influence model[26]. Its "influence learning" rule is modified in by modulating one of the initial learning parameters (η, λ or β) with a factor ELO_{mod} that depended on ELO_{diff}, the difference in ELO-scores of the interacting baboons. The three possible variations of the learning rules are:

$$p_{n+1} = p_n + ELO_{mod} * \eta * (o - p_n) - \lambda * k_1 * p_n * (1 - p_n) * (a - \frac{\beta * x_n - k_2}{k_1})$$

$$p_{n+1} = p_n + \eta * (o - p_n) - ELO_{mod} * \lambda * k_1 * p_n * (1 - p_n) * \left(a - \frac{\beta * x_n - k_2}{k_1}\right)$$

$$p_{n+1} = p_n + \eta * (o - p_n) - \lambda * k_1 * p_n * (1 - p_n) * (a - \frac{ELO_{mod} * \beta * x_n - k_2}{k_1})$$

Where $ELO_{mod} = x * (ELO_{diff} - y)$ and x, y are learning parameters that were optimized to the baboons' behaviour.

References

1. Claidière, N., Gullstrand, J., Latouche, A. & Fagot, J. Using Automated Learning Devices for Monkeys (ALDM) to study social networks. *Behav Res Methods* **49**, 24–34 (2017).

2. Formaux, A., O'Sullivan, E., Fagot, J. & Claidière, N. Understanding Imitation in Papio papio: The Role of Experience and the Presence of a Conspecific Demonstrator. *Cogn Sci* **46**, e13117–e13117 (2022).

3. Formaux, A., Paleressompoulle, D., Fagot, J. & Claidiere, N. The experimental emergence of convention in a non-human primate. *Philosophical Transactions of the Royal Society B: Biological Sciences* this issue (2021).

4. Gullstrand, J., Claidière, N. & Fagot, J. Computerized assessment of dominance hierarchy in baboons (Papio papio). *Behav Res Methods* **53**, 1923–1934 (2021).

5. Fagot, J., Marzouki, Y., Huguet, P., Gullstrand, J. & Claidière, N. Assessment of social cognition in non-human primates using a network of computerized automated learning device (ALDM) test systems. *Journal of Visualized Experiments* **2015**, e52798 (2015).

6. Gelardi, V., Fagot, J., Barrat, A. & Claidière, N. Detecting social (in)stability in primates from their temporal co-presence network. *Anim Behav* **157**, 239–254 (2019).

7. Ballesta, S. *et al.* Assessing the reliability of an automated method for measuring dominance hierarchy in non-human primates. *Primates* **62**, 595–607 (2021).

8. Claidière, N., Trouche, E. & Mercier, H. Argumentation and the diffusion of counter-intuitive beliefs. *J Exp Psychol Gen* **146**, 1052–1066 (2017).

9. Gullstrand, J., Claidière, N. & Fagot, J. Computerized assessment of dominance hierarchy in baboons (Papio papio). *Behavior Research Methods 2021 53:5* **53**, 1923–1934 (2021).

10. Albers, P. C. H. & De Vries, H. Elo-rating as a tool in the sequential estimation of dominance strengths. *Animal Behaviour* vol. 61 489–495 Preprint at https://doi.org/10.1006/anbe.2000.1571 (2001).

11. Neumann, C. *et al.* Assessing dominance hierarchies: validation and advantages of progressive evaluation with Elo-rating. *Anim Behav* (2011).

12. Newton-Fisher, N. E. Modeling Social Dominance: Elo-Ratings, Prior History, and the Intensity of Aggression. *Int J Primatol* **38**, 427–447 (2017).

13. Fagot, J. & Bonté, E. Automated testing of cognitive performance in monkeys: Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). *Behav Res Methods* **42**, 507–516 (2010).

14. Neumann, C. & Fischer, J. Extending Bayesian Elo-rating to quantify the steepness of dominance hierarchies. *Methods Ecol Evol* **14**, 669–682 (2023).

15. A. Elo. *The Rating of Chess Players, Past and Present*. (Arco, 1987).

16. Neumann, C. *et al.* Assessing dominance hierarchies: validation and advantages of progressive evaluation with Elo-rating. *Anim Behav* **82**, 911–921 (2011).

17. Neumann, C. & Lars, K. EloRating: Animal Dominance Hierarchies by Elo Rating. *R package version 0.43.* http://CRAN.R-project.org/package=EloRating (2014).

18. Busenbark, J. R., Graffin, S. D., Campbell, R. J. & Lee, E. Y. A Marginal Effects Approach to Interpreting Main Effects and Moderation. *Organ Res Methods* **25**, 147–169 (2022).

19. Daunizeau, J., Adam, V. & Rigoux, L. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Comput Biol* **10**, 1003441 (2014).

20. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies - Revisited. *Neuroimage* **84**, 971–985 (2014).

21. Nassar, M. R., Wilson, R. C., Heasly, B. & Gold, J. I. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience* **30**, 12366–12378 (2010).

22. Khalvati, K. *et al.* Modeling other minds: Bayesian inference explains human choices in group decision-making. *Sci Adv* **5**, (2019).

23. Diaconescu, A. O. *et al.* Inferring on the Intentions of Others by Hierarchical Bayesian Learning. *PLoS Comput Biol* **10**, 1003810 (2014).

24. Daunizeau, J. *et al.* Observing the observer (I): Meta-bayesian models of learning and decision-making. *PLoS One* **5**, e15554 (2010).

25. Yoshida, W., Dolan, R. J. & Friston, K. J. Game theory of mind. *PLoS Comput Biol* **4**, e1000254 (2008).

26. Hampton, A. N., Bossaerts, P. & O'Doherty, J. P. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci U S A* **105**, 6741–6746 (2008).

27. Rusch, T., Steixner-Kumar, S., Doshi, P., Spezio, M. & Gläscher, J. Theory of mind and decision science: Towards a typology of tasks and computational models. *Neuropsychologia* **146**, 1–44 (2020).

28. Devaine, M., Hollard, G. & Daunizeau, J. The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? *PLoS Comput Biol* **10**, (2014).

29. Devaine, M. *et al.* Reading wild minds: A computational assay of Theory of Mind sophistication across seven primate species. *PLoS Comput Biol* **13**, 1–25 (2017).

30. Daunizeau, J. & Rigoux, L. Bayesian sequence learning. https://mbb-team.github.io/VBA-toolbox/wiki/BSL/ (2014).

31. Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58 (1993).

32. O'Doherty, J. P., Cockburn, J. & Pauli, W. M. Learning, Reward, and Decision Making. *Annual Review of Psychology* vol. 68 73–100 Preprint at https://doi.org/10.1146/annurev-psych-010416-044216 (2017).

33. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction, 2nd Ed. Reinforcement learning: An introduction, 2nd ed.* (The MIT Press, Cambridge, MA, US, 2018).

34. Hill, C. A. *et al.* A causal account of the brain network computations underlying strategic social behavior. *Nat Neurosci* **20**, 1142–1149 (2017).

35. Quang, P. B., Musso, C. & Gland, F. Le. The Kalman Laplace filter: A new deterministic algorithm for nonlinear Bayesian filtering. in *2015 18th International Conference on Information Fusion (Fusion)* 1566–1573 (2015).

36. Dayan, P. & Balleine, B. W. Reward, motivation, and reinforcement learning. *Neuron* vol. 36 285–298 Preprint at https://doi.org/10.1016/S0896-6273(02)00963-7 (2002).

37. Behrens, T. E. J., Hunt, L. T., Woolrich, M. W. & Rushworth, M. F. S. Associative learning of social value. *Nature* **456**, 245–249 (2008).

38. Johnson, J. D., Li, J. & Chen, Z. Reinforcement Learning: An Introduction. *Neurocomputing* **35**, 205–206 (2000).

Supplementary material

GLMs & GLMEs

Generalized linear mixed-effects models were computed with the lme4 package in R and ensured that we had good estimates of random effects, and accounted for variability in behaviour using Bound Optimization by Quadratic Approximation. Our GLME models showed a main effect of past choices and outcomes on the decision of the baboons.

We at first modeled our GLM based on the following formulas:
$Reward_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-1}$ + $Stay_{n-2}$ + ELOdiff + (session*Condition*Trial)
$Stay_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-1}$ + $Stay_{n-2}$ + ELOdiff + (session*Condition*Trial)

After applying a dropterm process based on the AICs of the models, the final models were:
$Reward_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELOdiff + (session*Condition*Trial)
$Stay_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELOdiff + (session*Trial) + (session*Condition) + (Condition*Trial)

We used the glm() function of the lme4 package on R with a binomial family.

We then checked if our results were still robust if we included random effects. We used the glmer() function of the lme4 package on R with a binomial family and the bobyqa optimizer. This led to the following formula and results:

$Reward_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELOdiff + (session*Condition*Trial) + (1 + Trial | Nom)

$Stay_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELOdiff + (session*Trial) + (session*Condition) + (Condition*Trial) + (1 + Trial | Nom)

Models	**GLM**		**GLME**	
Variables	**$Reward_n$: Estimate (Std. Error)**	**$Stay_n$: Estimate (Std. Error)**	**$Reward_n$: Estimate (Std. Error)**	**$Stay_n$: Estimate (Std. Error)**
Intercept	-0.254 (0.180)	-0.736 (0.164) ***	-0.100 (0.191)	-0.723 (0.164) ***
$Reward_{n-1}$	0.879 (0.104) ***	2.358 (0.106) ***	0.827 (0.104) ***	2.357 (0.107) ***
$Reward_{n-2}$	1.246 (0.101) ***	0.681 (0.119) ***	1.195 (0.101) ***	0.681 (0.119) ***
$Stay_{n-2}$	0.431 (0.115) ***	0.429 (0.130) ***	0.401 (0.115) ***	0.421 (0.131) **
ELOdiff	0.199 (0.048) ***	0.182 (0.055) ***	0.216 (0.073) **	0.188 (0.056) ***

session	-0.080 (0.115)	0.292 (0.131) *	-0.074 (0.116)	0.291 (0.132) *
Condition	0.352 (0.129) ***	0.450 (0.142) **	0.363 (0.130) **	0.446 (0.143) **
Trial	0.260 (0.082) **	-0.082 (0.080)	0.296 (0.101) **	-0.079 (0.089)
session:Condition	0.867 (0.224) ***	0.343 (0.223)	0.877 (0.225) ***	0.357 (0.224)
session:Trial	0.052 (0.113)	0.246 (0.106) *	0.059 (0.114)	0.241 (0.107) *
Condition:Trial	0.189 (0.124)	0.465 (0.110) ***	0.200 (0.125)	0.468 (0.110) ***
session:Condition:Trial	0.406 (0.210)	/	0.398 (0.211)	/

Supp. Tab. 4*: Comparison of GLMs and GLMEs for behavioral factors.** Condition, Session, Reward and Stay are coded effects (Condition: Social = 1, Non-social = -1; Session: First = -1, Second = 1; Reward & Stay: True = 1, False = 0). Signif. codes: '': p < 0.001; '**': p < 0.01; '*': p < 0.05.

Detailed results for GLMEs:

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial (logit)
Formula: Score ~ 1 + Score_n_1 + Score_n_2 + Stay_n_2 + ELOdiff + (session*Condition*Trial) + (1 + Trial | Nom)
   Data: dat
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))
     AIC      BIC   logLik deviance df.resid
  3522.2   3621.5  -1746.1   3492.2     5501
Scaled residuals:
    Min      1Q  Median      3Q     Max
-12.1634  0.1594  0.2698  0.3570  1.7088
Random effects:
 Groups Name        Variance Std.Dev. Corr
 Nom    (Intercept) 0.08426  0.2903
        Trial       0.02223  0.1491   0.68
Number of obs: 5516, groups:  Nom, 7

Fixed effects:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.09978    0.19065  -0.523 0.600724
Score_n_1         0.82703    0.10443   7.919 2.39e-15 ***
Score_n_2         1.19512    0.10107  11.824  < 2e-16 ***
```

```
Stay_n_2               0.40113    0.11501   3.488 0.000487 ***
ELOdiff               0.21637    0.07325   2.954 0.003139 **
session               -0.07381    0.11563  -0.638 0.523249
Condition              0.36283    0.13044   2.782 0.005409 **
Trial                 0.29621    0.10108   2.930 0.003385 **
session:Condition      0.87650    0.22504   3.895 9.82e-05 ***
session:Trial          0.05931    0.11403   0.520 0.602954
Condition:Trial        0.19961    0.12461   1.602 0.109170
session:Condition:Trial  0.39801    0.21065   1.889 0.058836 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
          (Intr) Scr__1 Scr__2 Sty__2 ELOdf sessin Condtn Trial  sssn:C sssn:T Cndt:T
Score_n_1   -0.346
Score_n_2   -0.297 -0.090
Stay_n_2    -0.406 -0.095 -0.130
ELOdiff     0.029 -0.033 -0.019 -0.049
session     -0.331  0.029  0.028 -0.001  0.007
Condition   -0.251 -0.016 -0.013 -0.028  0.163  0.455
Trial       0.401 -0.080 -0.077 -0.053 -0.006 -0.129 -0.107
sessn:Cndtn  0.225 -0.056 -0.055 -0.041  0.007 -0.516 -0.560  0.075
session:Trl -0.072  0.012  0.016 -0.012  0.004  0.238  0.103 -0.568 -0.124
Conditn:Trl -0.048  0.004  0.001 -0.033  0.020  0.106  0.318 -0.520 -0.181  0.464
sssn:Cndt:T  0.073 -0.032 -0.035 -0.020  0.032 -0.132 -0.181  0.313  0.508 -0.543 -0.591

optimizer (bobyqa) convergence code: 0 (OK)
boundary (singular) fit: see ?isSingular
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
 Family: binomial (logit)
Formula: Stay ~ 1 + Score_n_1 + Score_n_2 + Stay_n_2 + ELOdiff + (session*Trial) +
(session*Condition) + (Condition*Trial) + (1 + Trial | Nom)
   Data: dat
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))
     AIC      BIC   logLik deviance df.resid
  2775.9   2868.5  -1373.9   2747.9     5502
Scaled residuals:
    Min      1Q  Median      3Q     Max
-8.8525  0.1542  0.2137  0.2608  1.7489
Random effects:
 Groups Name        Variance Std.Dev. Corr
 Nom    (Intercept) 0.00000  0.0000
        Trial       0.01046  0.1023    NaN
Number of obs: 5516, groups:  Nom, 7

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.72323    0.16443  -4.398 1.09e-05 ***
Score_n_1          2.35710    0.10668  22.094  < 2e-16 ***
Score_n_2          0.68099    0.11946   5.700 1.20e-08 ***
Stay_n_2           0.42086    0.13050   3.225 0.001259 **
ELOdiff           0.18809    0.05605   3.356 0.000791 ***
session            0.29110    0.13150   2.214 0.026848 *
Trial             -0.07938    0.08906  -0.891 0.372789
```

```
Condition          0.44637    0.14265   3.129 0.001753 **
session:Trial      0.24119    0.10654   2.264 0.023584 *
session:Condition  0.35735    0.22377   1.597 0.110276
Trial:Condition    0.46773    0.10981   4.259 2.05e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) Scr__1 Scr__2 Sty__2 ELOdf sessin Trial  Condtn sssn:T sssn:C
Score_n_1   -0.295
Score_n_2   -0.435 -0.094
Stay_n_2    -0.540 -0.079 -0.131
ELOdiff      0.061 -0.024 -0.027 -0.060
session     -0.409  0.054  0.036  0.005  0.084
Trial        0.146 -0.092 -0.081 -0.056 -0.006 -0.021
Condition   -0.312 -0.022 -0.004 -0.020  0.127  0.419  0.017
session:Trl  0.009 -0.006 -0.012 -0.031  0.055  0.132 -0.524 -0.059
sessn:Cndtn  0.289 -0.057 -0.042 -0.051 -0.033 -0.554 -0.129 -0.573  0.208
Trial:Cndtn  0.023 -0.009  0.000 -0.057  0.039 -0.028 -0.477  0.210  0.134  0.154

optimizer (bobyqa) convergence code: 0 (OK)
boundary (singular) fit: see ?isSingular
```

Marginal effects and data:

The marginal effects of the social vs non-social condition on the tested baboons' stay vs switch decisions corroborates the influence of social context in decision learning (Fixed Effect of Condition on the stay vs switch decisions $stay_n$: Estimate=0.446; SE=0.143; z value=3.13; P=1.75E-3**). This result was also significant for the baboons' performance: we found a significant difference between conditions (Fixed Effect of Condition on the performance $reward_n$: Estimate=0.363; SE=0.130; z value=2.78; P=5.40E-3**). As for the marginal effects of difference of dominance on the tested baboons' stay vs switch decision, once again we found a significant effect (Fixed Effect of Dominance on the stay vs switch decisions $stay_n$: Estimate=0.188; SE=0.056; z value=3.36; P=7.9E-4***). We also found a significant effect of the difference of dominance on the performance, (Fixed Effect of Dominance on the performance $reward_n$: Estimate=0.216; SE=0.073; z value=2.95; P=3.14E-3**).

ELOdiff vs ELOself vs ELOother:

Our logistic regression indicated that coordination learning was influenced by the relative difference in dominance between individuals. Some prior work suggests that prosocial behavior in macaques depends on absolute rank rather than the difference in rank[1]. To test this alternative hypothesis, we included the absolute ELO score of the baboon rather than the relative difference in ELO scores in the logistic regression and found no significant results (Supp. Mat. & Meth.). We also tested the absolute ELO score of the other baboon instead of the difference and found a significant effect of this regressor, but less significant than the relative difference in ELO scores. This could be explained by the fact that in the experiment with macaques[1] the decision to be prosocial or not was made by one individual in the pair, who chose to reward either only themselves, or both themselves and the other. When the baboons learned the co-ordination task for our experiment, both their decisions were important, since they had to co-ordinate to obtain rewards. We presume that having learned this during the similar social task during training, the baboons would continue to believe this was the case during the social task.

Predicted variable	$Reward_n$			$Stay_n$		
Variables	**p-value**	**AIC**	**BIC**	**p-value**	**AIC**	**BIC**
ELOdiff	0.003139 **	3522.2	3621.5	0.000791 ***	2775.9	2868.5
ELOself	0.067841	3527.9	3627.1	0.05413	2780.2	2872.8
ELOother	0.018506 *	3525.0	3624.2	0.007483 **	2779.9	2872.5

***Supp. Tab. 5*: Comparison between GLME models using ELOdiff vs ELOself vs ELOother.** We replaced in ELOdiff in the GLME models (see "GLMs & GLMEs" section) by the personal ELO score of the tested baboon ELOself, and then by the ELO score of the other baboon present in the adjacent box ELOother. We compared the different models by AIC, BIC, and p-value for each ELO variable. The model using ELOdiff was the best explaining model.

Second analysis with WCST-score, age and association weight

This analysis included the age of the baboons, the association weight between baboons and the perseveration score of the subject at the WCST:
We at first modeled our GLM based on the following formulas:
$Reward_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-1}$ + $Stay_{n-2}$ + ELO_diff + assoc_weight + mean_persev_self + Age + (session*Condition*Trial)
$Stay_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-1}$ + $Stay_{n-2}$ + ELO_diff + assoc_weight + mean_persev_self + Age + (session*Condition*Trial)

After applying a dropterm process based on the AICs of the models, the final models were:
$Reward_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELO_diff + Age + (session*Condition*Trial)
$Stay_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELO_diff + assoc_weight + mean_persev_self + (session*Trial) + (session*Condition) + (Condition*Trial)

We then checked if our results were still robust if we included random effects. We used the glmer() function of the lme4 package on R with a binomial family and the bobyqa optimizer. This led to the following formula and results:

$Reward_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELOdiff + Age + (session*Condition*Trial) + (1 + Trial | Nom)

$Stay_n$ ~ 1 + $Reward_{n-1}$ + $Reward_{n-2}$ + $Stay_{n-2}$ + ELOdiff + assoc_weight + mean_persev_self + (session*Trial) + (session*Condition) + (Condition*Trial) + (1 + Trial | Nom)

Models	GLM		GLME	
Variables	**$Reward_n$: Estimate (Std. Error)**	**$Stay_n$: Estimate (Std. Error)**	**$Reward_n$: Estimate (Std. Error)**	**$Stay_n$: Estimate (Std. Error)**
Intercept	-0.234 (0.151)	-1.361 (0.275) ***	-0.108 (0.183)	-1.40 (0.278) ***

$Reward_{n-1}$	0.865 (0.104) ***	2.36 (0.107) ***	0.827 (0.104) ***	2.35 (0.10690) ***
$Reward_{n-2}$	1.23 (0.101) ***	0.678 (0.119) ***	1.20 (0.101) ***	0.675 (0.119) ***
$Stay_{n-2}$	0.435 (0.115) ***	0.412 (0.130) **	0.404 (0.115) ***	0.398 (0.131) **
ELOdiff	0.256 (0.0515) ***	0.286 (0.067) ***	0.241 (0.0723) ***	0.305 (0.069) ***
Age	0.153 (0.0500) **	/	0.128 (0.0881)	/
assoc_weight	/	-0.279 (0.166) .	/	-0.290 (0.168) .
mean_persev_self	/	0.0581 (0.0285) *	/	0.0665 (0.0293) *
session	-0.0805 (0.115)	0.297 (0.131) *	-0.0731 (0.116)	0.297 (0.132) *
Condition	0.363 (0.129) **	0.648 (0.172)	0.371 (0.131) **	0.657 (0.174)
Trial	0.268 (0.0817) **	-0.075 (0.080) ***	0.298 (0.101) **	-0.0676 (0.094) ***
session:Condition	0.877 (0.225) ***	0.328 (0.223) *	0.874 (0.225) ***	0.345 (0.224) *
session:Trial	0.0453 (0.114)	0.244 (0.106)	0.057 (0.114)	0.238 (0.107)
Condition:Trial	0.179 (0.124)	0.491 (0.111) ***	0.198 (0.125)	0.499 (0.111) ***
session:Condition: Trial	0.427 (0.211) *	/	0.403 (0.211) .	/

Supp. Tab. 6*: Comparison of GLMs and GLMEs for behavioral factors.** Condition, Session, Reward and Stay are coded effects (Condition: Social = 1, Non-social = -1; Session: First = -1, Second = 1; Reward & Stay: True = 1, False = 0). Signif. codes: '': p < 0.001; '**': p < 0.01; '*': p < 0.05. The association weight, assoc_weight, corresponds to the number of times where two monkeys played in interaction together, divided by their total number of trials.

The association weight, assoc_weight, corresponds to the number of times where two monkeys played in interaction together divided by their number of trials. The perseveration score mean_persev_self indicates the tendency for a baboon to persevere in a wrong choice (perseverative error) at a Wisconsin Card Sorting Test [2,3]. It measures their cognitive flexibility. These results indicate that the cognitive flexibility of an individual significantly increases their tendency to stay with their previous choice.

lm(formula = EloScore_other ~ 1 + mean_persev, data = persev_ELO_data)

Residuals:

Min	1Q	Median	3Q	Max
-143.430	-25.512	3.131	53.957	70.667

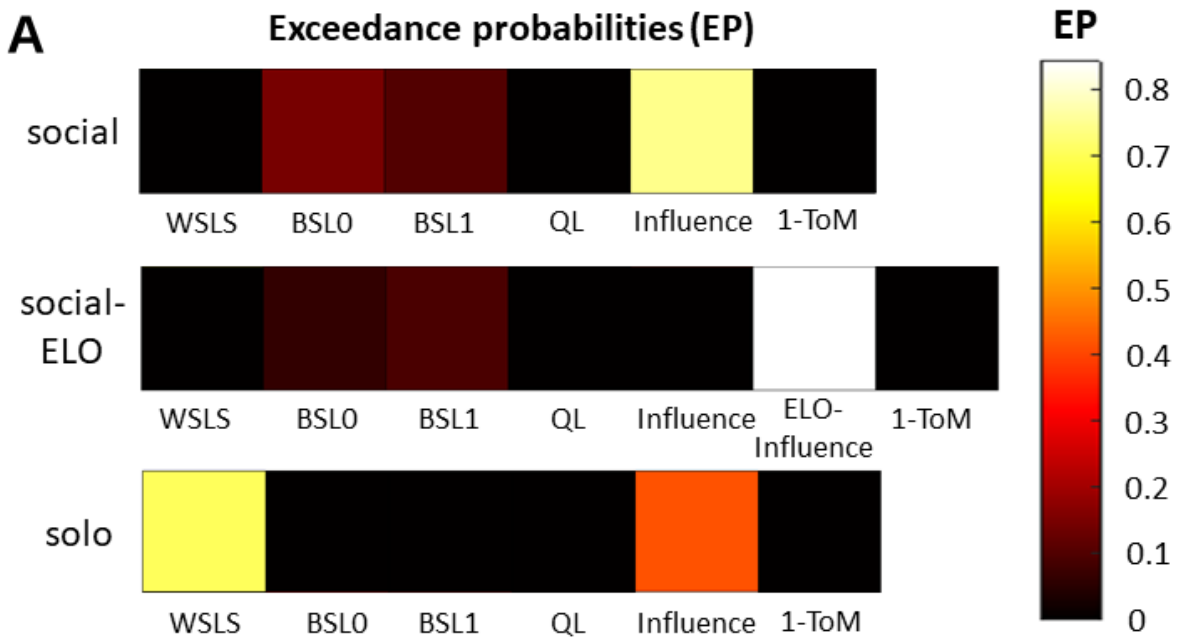
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	23.953	29.935	0.800	0.436
mean_persev	-2.460	3.048	-0.807	0.432

Residual standard error: 67.36 on 15 degrees of freedom
Multiple R-squared: 0.04161, Adjusted R-squared: -0.02228
F-statistic: 0.6513 on 1 and 15 DF, p-value: 0.4323

Computational modelling of dominance:

We first developed three new models derived from the influence model by allowing the relative difference in their ELO dominance score to modulate any one of its three learning parameters. The first parameter "eta" modulates the prediction error between the opponent's expected action and actual action (sometimes also called the learning rate). On the other hand, the two other parameters lambda and beta modulate the influence update, i.e. how much a player influences his/her opponent (see **Methods, Influence model**). We then fitted each model to the social condition data, and selected the one that fitted best. We found that modulating the parameter "eta", which is the learning rate of the tested baboon in the Influence model, best explained the data (BMS: EP_{eta} = 0.952, EP_{lambda} = 0.046, EP_{beta} = 0.003). This new "eta"-modulated Influence model, thereafter called ELO-Influence model, incorporates a prediction error term that is modulated by differences of dominance, and that runs in addition to a ToM component, the influence update, which is the same as in the classical Influence model. A Bayesian Model Selection (BMS) performed with this new "eta"-modulated Influence model, thereafter called ELO-Influence model, in addition to the previous set of candidate models in the social condition, revealed that the ELO-influence model was the best fitting model [EP = 0.84] compared to all those tested previously (**Supp. Fig. 1A**).

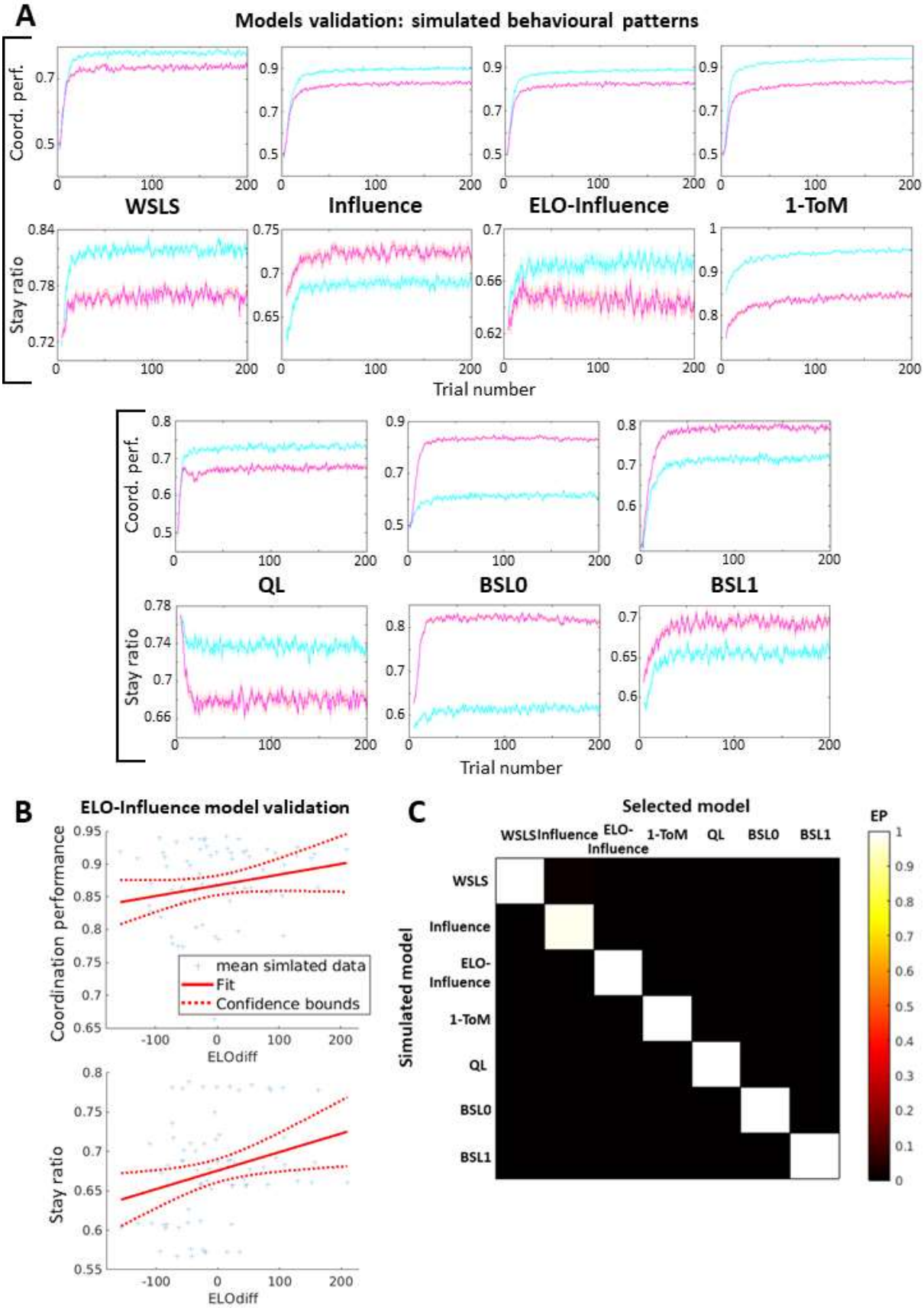


***Supp. Fig. 1*: Results of computational modelling. (A) Bayesian Model Selections (BMS).** X-axis: model names; Y-axis: social condition; Color-scale: Exceedance Probabilities (EP), i.e., the probability that one model is more frequent in the group population than any other model [4]. **(social)** In the social condition, the influence model is the best fit for our group of baboons. **(social-ELO)** In the updated social condition, the ELO-modulated Influence model was the most probable model. **(solo)** In the solo condition, the WSLS heuristic fits best.

Control analyses:

A good practice in the field of computational modelling is to check to what extent the selected set of candidate computational models qualitatively generate data close to the actual behavioural data [5,6]. This step is called model validation. To do so, we first proceeded to 10000 Monte Carlo simulations of each model that played with the AA. If the models generate plausible data, then we would expect that distinct decision patterns would be produced between distinct (simulated) conditions, i.e. data generated from a same model but with model parameters that were drawn from different statistical distributions. We also expected that the coordination performance and the tendency to stay would increase fast at the beginning, then would converge to an asymptotic value for each model. This is globally what is observed for the simulated data of each candidate model in our set of models (**Supp. Fig. 2A**). As for the ELO-Influence model specifically, a distinct decision pattern, i.e. coordination performance and tendency to stay, should appear between dominant vs submissive interactions, i.e. high vs low difference in relative ELO- scores. More precisely, a high (vs low) relative difference in ELO-scores would result in higher mean coordination performance (**Supp. Fig. 2B**, left) and tendency to stay (**Supp. Fig. 2B**, right).

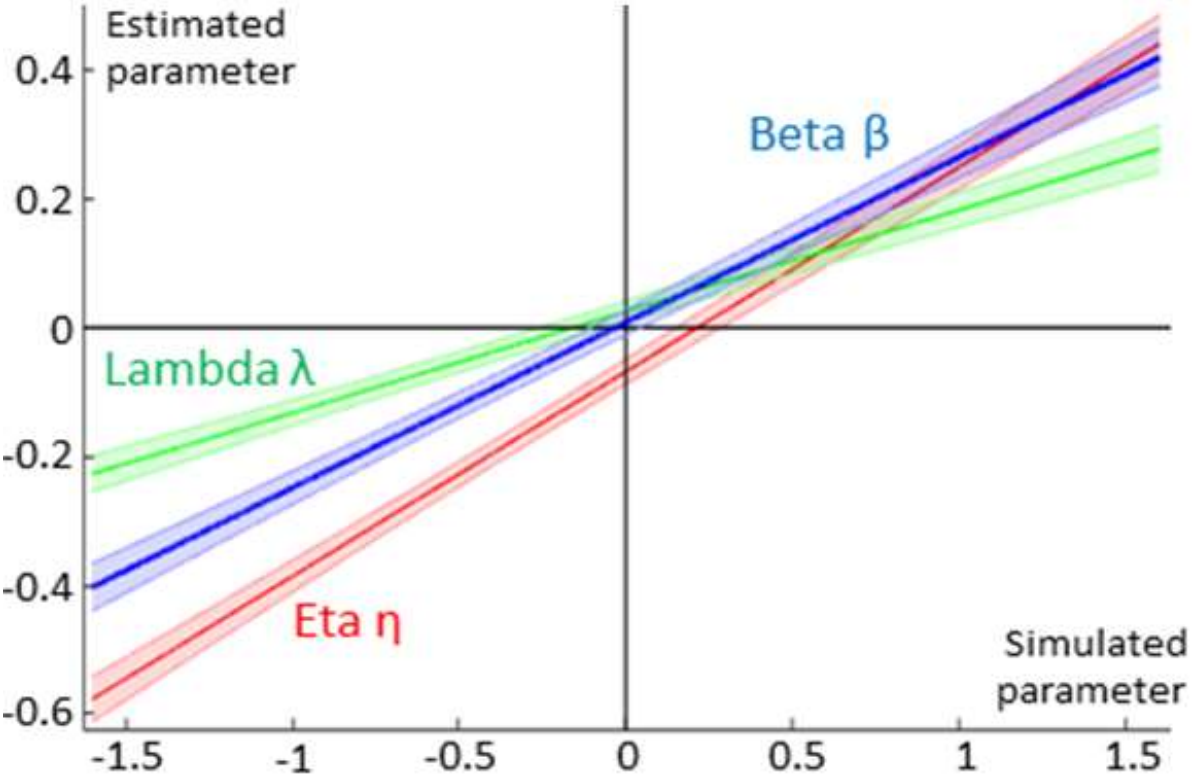
The distinguishability of the models was ensured by a model confusion analysis [5]: we generated 1400 Monte Carlo simulations of each model that played with the AA, then fitted the simulated data to our set of candidate models. If the models are indeed distinguishable with the method of model selection that we use, then each model that was initially simulated should be recovered by this method. The confusion matrix that reports the Exceedance Probability [4] for each model is diagonal (**Supp. Fig. 2C**), which indicates a recoverability of the models at the group level [5]. This allows us to compare these models through the Bayesian Model Selection (BMS) method [4,7] to determine which computational models explain the behaviour best.



Supp. Fig. 2: Sanity checks. **(A) Models validation: simulated conditions.** *x-axis: trial number. y-axis: coordination performance, perf (top); tendency to maintain same choice as previously, stay (bottom). Each model was simulated 10000 times playing with the AA in two different conditions (magenta and cyan, 5000 Monte-Carlo simulations per condition), i.e. for sets of parameters drawn from*

different gaussian distributions. Ribbons: standard errors. **(B) ELO-Influence model validation: effect of social dominance.** *x-axis: difference in ELO scores between the simulated agents, ELOdiff. y-axis: coordination performance, perf (top); tendency to maintain same choice as previously, stay (bottom). Data were simulated over 200 trials with 5000 Monte-Carlo simulations and averaged over ELOdiffs (blue crosses). A linear model was then fitted (red line) for both perf (top) and stay (bottom) with the corresponding 95% confidence interval (dashed red lines).* **(C) Confusion matrix**. *X-axis: model selected as that fitting best after model fitting and Bayesian Model Selection (BMS); Y-axis: initial simulated model that generated the data; Color-scale: Exceedance Probabilities (EP).*

Finally, to check the robustness of the rmANOVA, we performed a parameter simulation-recovery analysis to determine to what extent the influence model parameters are identifiable from observed data [5]. We therefore generated simulated data with multiple Monte-Carlo simulations (approx. 10000 in total) from the influence model using different sets of learning parameters sampled from a uniform distribution. The sampling was done (uniformly) between the most extreme empirical values found in our data analysis for each of our three learning parameters. We then fitted the generated data with the influence model with the same method as previously. We then regressed the estimated parameters on the simulated parameters. We found for each learning parameter a positive correlation between estimated and simulated parameters (**Supp. Fig. 3**). The regression indicates that the estimated parameters were actually underestimated compared to those used to generate simulated data. Therefore, we might have underestimated the values of the learning parameters of the influence model in our previous rmANOVA (Erreur ! Source du renvoi introuvable.**B**). In reality, the differences of the estimated parameters might be larger between conditions. This result therefore shows that the above-mentioned findings are robust, and, if anything, underestimate the difference between the social and the non-social conditions.



Supp. Fig. 3: **Parameters recovery for the Influence model.** X-axis: value of the simulated parameters of the influence model that generated the simulated data; Y-axis: value of the estimated parameter after model fitting and parameter estimation on the simulated data.

References

1. Massen, J. J. M., van den Berg, L. M., Spruijt, B. M. & Sterck, E. H. M. Generous leaders and selfish underdogs: Pro-sociality in despotic macaques. *PLoS One* **5**, 1–5 (2010).
2. Berg, E. A. A Simple Objective Technique for Measuring Flexibility in Thinking. *https://doi.org/10.1080/00221309.1948.9918159* **39**, 15–22 (2010).
3. Gullstrand, J., Claidière, N. & Fagot, J. Age effect in expert cognitive flexibility in Guinea baboons (Papio papio). *Behavioural Brain Research* **434**, (2022).
4. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies - Revisited. *Neuroimage* **84**, 971–985 (2014).
5. Wilson, R. C. & Collins, A. G. E. Ten simple rules for the computational modeling of behavioral data. *Elife* **8**, 1–33 (2019).
6. Palminteri, S., Wyart, V. & Koechlin, E. The Importance of Falsification in Computational Cognitive Modeling. *Trends Cogn Sci* **21**, 425–433 (2017).
7. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).

C. Study 3

Main Manuscript for

Computational mechanisms underlying the emergence of theory of mind in children

Toan Nong[1,2], Yansong Li[3,4,*], Chen Qu[6,*], Jae Hyung Woo[7,**], Yuqian Wang[3,**], Chengguo Miao[3,**], Xiaoyi Liu[3,**], Qingxia Liu[6,**], Jean-Baptiste Van der Henst[5], Rémi Philippe[1,2],Edmund Derrington[1,2], Alireza Soltani[7], Jean-Claude Dreher[1,2]

5- *ISC-MJ, CNRS UMR5229, Neuroeconomics Lab, Bron, France*
6- *University of Lyon 1, UCBL1, Lyon, France*
7- *Reward, Competition and Social Neuroscience Lab, Department of Psychology, School of Social and Behavioral Sciences, Nanjing University, Nanjing, China*
8- *Institute for Brain Sciences, Nanjing University, Nanjing, China*
9- *CRNL-CNRS-UCBL1, Bron, France*
10- *Center for Studies of Psychological Application, School of Psychology,South China Normal University, Guangzhou, China*
11- *Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA.*
12- **,**: equal contributions*

Corresponding authors: dreher@isc.cnrs.fr and yansongli@nju.edu.cn

Author Contributions: J-C.D. and T.N. developed the general concept, experiment and models. Y.W., C.M., X.L., Q.L. performed the experiments under the supervision of Y.L. and Q.C. T.N. programmed the task under the supervision of J.-C.D. T.N. developed the models and implemented the algorithms, and analyzed the data under the supervision of J.-C.D. J.H.W. helped to analyze the data under the supervision of A.S. T.N. wrote the draft. E.D., J.B., J.H.W., J.-B.V.D.H., Y.L, A.S., and J.-C. D. edited the manuscript.

Competing Interest Statement: The authors declare that they have no competing interests.

Classification: Paste the major and minor classification here. Dual classifications are permitted, but cannot be within the same major classification.

Keywords: Paste the keywords here. There should be at least three and no more than five.

This PDF file includes:

Main Text
Figures 1 to 4

Abstract (250 words max)

Dyadic social interactions often require one to adapt to the fluctuating cooperative or competitive intentions of others, which can change over time without being explicitly signaled. This ability is a critical component of theory of mind (ToM), which develops gradually during childhood. However, the computations underlying this ability remain to be described at a mechanistic level. Here, we used a combination of experimental and modeling approaches in children (3 to 9 years old) to elucidate the computational mechanisms underlying this ability and to identify how these computations develop during childhood. We implemented a card-matching task in which 192 children (100 females) thought they were playing with another player. In fact, they played with an algorithm that alternated between cooperative and competitive strategies across blocks of trials without warning. Behavior of younger children (kindergarten, age 3-5) was more compatible with a simple decision process that tracked the probability that the other agent would choose a particular card, based on their previous actions. By 8 years of age, most children mainly followed a mixed-intentions influence learning (MIIL) model, that arbitrates between cooperative and competitive intentions to adapt their decisions. These results show that the combination of the simulation of one's own influence on others and of an arbitration process between cooperative and competitive strategies is an important developmental milestone occurring at about 8 years of age. Our findings characterize the development of the computations required to adapt to the fluctuating cooperative or competitive intentions of others during childhood.

Significance statement (120 words max)

This study identifies the computations required for strategic interactions in children using a formal computational approach. We characterize the nature and milestones of the development of computational mechanisms involved in adapting to the intentions of others in children which is part of Theory-of-Mind. Most children become able to mentalize their influence on another's decision and adapt to another's intentions at about 8 years old. The model accounting best for behavior around this age was a model that arbitrated between a mixture of a competitive and a cooperative version of an influence learning model. This influence learning model is a reinforcement learning model that also integrates the influence of its decision on the other's outcome.

Introduction

Often considered as an essential requirement for sophisticated social interactions, Theory-of-Mind (ToM) is the ability to attribute mental states, such as emotions, beliefs, or desires, to other individuals(1). Different milestones have been proposed in the development of ToM from birth to early adulthood (2,3). Its most basic form, also known as perception–goal psychology (4–6) emerges from around 9 months of age and allows individuals to represent that others may have different perceptual perspectives of the world and act accordingly. Later, meta-representations emerge from around 4 years of age, in the form of belief–desire psychology, and involve an appreciation that others subjectively represent the world in fine-grained aspectual ways that may be false or incompatible with one's own view (7–10). At around 7 years of age, children's meta-representational understanding can be further refined to the so-called Advanced Theory-of-Mind (AToM) (8,10–14). AToM describes multiple aspects of an advanced understanding of the minds of others, including social reasoning that involves understanding of others' intention (8,11). The ability to track how others' behavior might change, and to adapt to others' cooperative/competitive intentions that fluctuate over time is essential to keep an advantage in social interactions. In such unsignaled changing contexts, understanding other's intentions requires not only to represent the beliefs about other's intentions, as in cooperative or competitive contexts alone, but also to adapt to whether the other is competing or cooperating, in order to deploy an appropriate behavioral response (15). The computational mechanisms underlying this ability to adapt to the fluctuating cooperative or competitive intentions of others remains unknown during development in children. Characterizing these computations is important because it should help build a mechanistic understanding of how this component of AToM develops in childhood (8), and may also help build Artificial Intelligence (AI) systems able to seamlessly interact with humans (16–18).

To examine the computations underlying adaptation to fluctuating cooperative/competitive intentions of others in children, we built on a recent literature that uses economic games to test children's decision abilities (19–22). We specifically focus on the ability to adapt to fluctuating cooperative versus competitive intentions of others, which may be an important cognitive component required for the social reasoning component of AToM. Formal computational models of ToM functions have been proposed in adults (15,23–30). More specifically, a model, known as the Mixed-Intention Influence Learning (MIIL) model (15) accounts for the influence of one's choice on another's decisions and also arbitrates between competitive and cooperative opponent's decision strategies using two experts and a controller. One expert assesses the competitive intentions and the other assesses cooperative intentions whereas the controller determines the relative weights of these experts on choice based on their relative reliabilities. This model outperforms other learning models in predicting social choice behavior in adults when they have to alternate between unsignaled cooperative and competitive games. Yet, it is unclear whether this or simpler models of ToM apply to children and how they develop during childhood.

Here, we sought to identify whether, and when, these specific computational mechanisms appear during children's development, before they develop an AToM. To that end, we used computational modeling and a strategic decision task, similar to the one performed in adults (15), that requires participants to detect fluctuating cooperative/competitive intentions of others. Pairs of children completed a task in which they played a card-matching game on a tablet (**Fig.1**). Each child from these pairs played facing a same-gender child (**Fig.1a&b**). Both were told that they were playing with each other, but were not allowed to communicate. Unbeknownst to them, they were in fact playing with an Artificial Agent (AA) that switched between cooperative and competitive modes of interaction (see **Fig.1d**). The goal of the game was to score as many points as possible by finding the same card as the other player (i.e. AA) (see **Fig 1c**). The rules to earn points were not explicitly given to the children, rather they had to infer them by trial and error. Therefore, this task allowed us to investigate the computations used by children to adapt to fluctuating cooperative and competitive intentions of the other and to identify at what developmental age they are able to do so. Importantly, the specific mode of interaction (i.e., cooperation vs competition) was not indicated to children.

To identify the computational mechanisms used to adapt to others' intentions, we fit individual's choice data using each of the candidate models to determine the best/winning model for each age groups. We also analyzed the learning parameters of the winning models to characterize the evolution of their hidden states in different age groups. This allowed us to describe, at a mechanistic level, how adaptation to changing intentions of others develops and evolves between different age groups. We expected the ability to differentiate between the competitive vs cooperative modes of interaction to increase with age. Based on the literature about strategic games (19) and AToM (8,11,31), we expected children around the age of 7-8 to adapt to the hidden cooperative or competitive intentions of their

partners. At the computational level, we hypothesized that younger age groups of children should be better fitted by simpler computational models, whereas the most sophisticated model (the MIIL model) could be the winning model only in the oldest age group. Alternatively, if the same model fits all age groups best, the evolution of the learning parameters of that model with age should explain better performance corresponding to better adaptation to competitive vs cooperative intentions in older children.

Results

Behavioral signature of tracking intentions. To study possible signature of tracking other's intentions, we considered multiple variables that could affect choice behavior. These included the previous stay vs switch choices of the child player and the artificial agent (AA) over the previous three trials (at *t-1*, *t-2*, *t-3*), the three previous outcomes (i.e. whether the previous three trials were won or lost), the age and sex of the child, decision time, the hierarchical condition (i.e., if they were paired with a child in the same school grade, an older child, or a younger child), the number of trials within a session of the experiment, the strategy of the AA at that time (competitive or cooperative), the previous cooperativity signatures at *t-1* and *t-2* (see below), and finally the interactions between age and cooperativity signatures as regressors. We first investigated the probability to "stay", i.e., to select the same target as that from the previous trial, for the three different age groups (**Fig. 2a**). We examined how children used the history of previous interactions to make their choices.

To measure cooperation, we defined the "cooperativity signature" as a binary variable that is equal to 1 if: (a) the participant won on the previous trial and the artificial agent (AA) stayed on the same target for the next trial, or (b) the participant lost on the previous trial and the AA switched to the other target for the next trial. Otherwise, the cooperativity signature was set to 0. Indeed, from the perspective of the participant, if the AA is a cooperative partner, then when both players win by choosing the same target, they should choose to keep the same target. Instead, if they lose due to choosing different targets, the AA's switch can be seen as its willingness to conform to the participant's choice. Cooperativity signature equal to 1 corresponds to the AA following win-stay/lose-switch strategy.

Overall, participants' performance improved with time and within a session of the experiment. Children tended to stay more frequently over time in the cooperative blocks. Moreover, the average probability of "stay" (i.e., to select the same target as that from the previous trial) over time revealed that older children tended to stay more during the cooperative blocks (**Fig. 2a**). Examining how children used the history of previous interactions to make their choices, we found that the outcomes of the previous three trials, i.e., the decisions of both the participant and the AA to stay and whether the participant was successful, significantly predicted the probability that the participant would stay and win, independently of all other factors (**Supplementary Table 1**).

Moreover, the cooperativity signature at t-1 predicted a higher probability for the children to stay with their choice. This indicates that children tracked whether the other agent was cooperating during the previous trial. More specifically, the marginal effects of the interaction between cooperativity signature and age (**Fig. 2b left**) indicated that older children performed significantly better after a positive (vs zero) cooperativity signature from the AA. In comparison, after controlling for other independent variables, younger children did not show such improvements in performance after trials with a positive cooperativity signature. This indicates that older children identified signatures of cooperativity better. In contrast, younger children also tended to stay less in the cooperative blocks, independently from other factors. However, taking into account the marginal effects of the interaction between the age of the child and the mode of the AA (cooperative vs. competitive) (**Fig. 2b right**), we observed that stay behavior was more frequent in cooperative blocks compared to the competitive blocks for older children, whereas, surprisingly, it decreased in cooperative blocks compared to competitive blocks for younger children. These results point towards an evolution of the learning process for intention adaptation and coordination strategy. We found no effect for sex or hierarchical condition. Our findings hold after a step-by-step drop-term procedure or addition of a random effect of time (see Supp. Mat., section GLM and GLME for further details).

Model-free analysis based on information theory reveals emergence of new strategies in older children. To examine whether and how children continuously adjust their behavior across cooperative and competitive blocks and reveal different choice strategies across age groups, we next computed

several metrics based on information theory to measure consistency in children's responses to the AA's choice and reward feedback. More specifically, we computed the entropy in participant's choice strategy in terms of stay or switch on the previous target (H(Str)), mutual information in reward-dependent strategy (MIRS), and conditional entropy of reward-dependent strategy (ERDS) over the time course of the experiment. We computed these quantities using a running average (with window size of seven trials) and separately for each of the three age groups (**Fig. 3**, see Methods for more details).

We found that during the cooperative blocks, better performance (**Fig. 3a**) in older children was accompanied by lower H(Str) (Spearman's correlation; r = -0.718, p = 7.15e-32), suggesting that older children were overall more consistent in their stay or switch between trials (**Fig. 3b**, Wilcoxon rank sum test on H(Str); KG vs. EPS: p = .0169; EPS vs. MPS: p = .0203). The lower H(str) observed in older children was mostly due to a decrease in ERDS (r = 0.852, p = 1.53e-55; **Fig. 3d**), corresponding to more consistent responses to reward feedback, and to a lesser extent, a decrease in MIRS (r = 0.438, p = 1.98e-10; **Fig. 3c**) corresponding to a weaker link between choice and previous reward outcome. These results indicate that the superior performance of older children during the cooperative blocks was due to two factors: (1) older children were able to choose more consistently across trials (less stochasticity in choice); (2) older children based their choices on reward outcome less strongly.

In contrast, the differences between the three age groups were smaller during the competitive blocks. More specifically, there were no significant differences in their consistency in strategy (H(Str)), even between the youngest (KG) and the oldest (MPS) age groups (Wilcoxon rank sum test, p = .582; **Fig. 3b**, inset). Yet, we observed that the oldest (MPS) age group responded more strongly to the outcome on the previous trials, as indicated by significantly larger MIRS than the younger EPS group (Wilcoxon rank sum test, p = .0285; **Fig. 3c** inset) or KG group (p = .038). Similarly, the randomness in strategy after accounting for the effect of reward outcome (ERDS) was significantly smaller for the oldest group (Wilcoxon rank sum test, p = .00205; **Fig. 3d** inset), but not significantly different between KG and EPS children (p = .311). The smaller ERDS for the oldest group compared to the younger group (EPS) was accompanied by a larger MIRS (**Fig. 3c**). This indicates that during the competitive blocks, only children in the oldest group were able to learn better from reward feedback and this made them more consistent in their responses to rewards.

Overall, these results suggest that only the oldest group of children (MPS) were able to adjust their strategy appropriately when the intention of AA changed between the cooperative and competitive blocks. Consistent with this interpretation, we observed that the MPS group showed the largest overall increase in MIRS from cooperative to competitive blocks (**Fig. 3c**; mean increase = 0.0788; Wilcoxon signed-rank test, p = .00127), whereas there were no significant differences for two younger age groups in MIRS between these two conditions (KG: mean increase = - 0.00286, p = .891; EPS: mean increase = -0.0438, p = .634). These results indicate the development of additional strategies in older children that allow them to learn from reward feedback differently depending on the nature of the interaction with the other agents.

Computational models track children's evaluation of the "other player's" intentions. To elucidate the computations underlying the children's strategic decision-making, we compared the ability of a variety of computational models in fitting choice data of individual children on a trial-by-trial basis. We sought to examine their ability not only to mentalize the other agent, but also to adapt to its intentions. Therefore, we fitted different Influence models (27,28), including the Mixed-Intentions Influence Learning (MIIL) and classical models (15). Classical influence models rely on mentalizing mechanisms that account for the influence of one's own choice on the other's choice. We tested a competitive as well as a cooperative version of the classical influence model. However, they did not eventually pass the model confusion analysis (see **Supplementary Materials**). In contrast, the MIIL model computes one decision value according to a competitive expert and another according to a cooperative expert. It then arbitrates between the two, based on the difference in their relative reliability, defined as the difference in unsigned value functions for the two choices determined by specific learning algorithms (see **Supplementary Materials**). We also compared these models to other models that did not rely on mentalizing mechanisms. Specifically, we fitted a Win-Stay/Lose-Switch (WS/LS) heuristic (32), a Q-Learning (QL) model based on reinforcement learning, and Bayesian models of different complexity, (Bayesian Sequence Learners of depth 0 and 1 (BSL0 and BSL1) (33) (see **Supplementary Materials** for more details on the models).

We found that the best fitting model for the majority of children in groups KG and EPS was the BSL0 model (**Fig.4a–b**), which is a Bayesian learning model that computes the mean and variance of the probability of the correct choice for the next trial (KG: Estimated Frequency Ef=0.605, protected Exceedance Probability pEP=0.999; EPS: Ef=0.530, pEP=0.986). This result shows that a majority of the children from 3 to 7 years old neither tracked the intentions of the other agent, nor relied on any mentalizing process for their decision in this task. They rather tracked with a simple Bayesian strategy which card was the most probable at each trial. In mid-primary schoolers (MPS), the MIIL model fitted the majority of the children best (MPS: Ef=0.483, pEP=0.956, **Fig.4a–b**). This indicates that the ability to mentalize and to track the intentions of others for strategic decision-making actually appears around the age of 8 years old, when children are in the third year of primary school (in China). Furthermore, this ability to mentalize and track intentions, with an un-signaled alternation between cooperative and competitive blocks, appears gradually among children starting from kindergarten (KG: Ef=0.122, pEP=10e-11; EPS: Ef=0.278, pEP=1.36e-2; MPS: Ef=0.483, pEP=0.956, **Fig.4a–b**). These suggest that children progressively develop the ability to mentalize and update their beliefs about future chosen targets, and arbitrate between the predicted intentions of another to compete or cooperate. This ability appears earlier in some children (at kindergarten) but the majority of mid-primary schoolers apply this decision process which remains the most common strategy in adults (15).

The sophistication of children's decision process increases with age. The BSL0 model best accounted for the behavior of the majority of younger children, but was almost entirely absent as the best model in the oldest group. Thus, the model frequencies of the oldest (MPS) vs youngest (KG) groups of children differed significantly. The posterior probability that the two groups had the same model frequencies was p = 6.7e-3 (uncorrected). Although there was no significant difference between the frequency of the different models in the KG and EPS groups (p=0.995) or in the EPS and MPS groups (p=0.990), there was a general trend towards an increase in the sophistication of the strategies such that the BSL1 model best accounted for the behavior of more EPS and MPS students than the BSL0. The WSLS heuristic remained a marginal strategy for some EPS and MPS participants, although its representation increased among the other models with age (see **Fig.4a–b**).
Finally, analysis of the arbitration parameters of the MIIL model for the children that were best fitted by this model, revealed how the model attributes the cooperative vs competitive tendency of the children (**Fig.4c**). In general, the children showed a low tendency to cooperate in the game. This was the case across the three age groups. Their tendency to cooperate increased during the cooperative blocks. In particular, in the older age group, children became more likely to cooperate than not during the cooperative blocks. Furthermore, for the MPS group, the tendency to cooperate in cooperative blocks increased progressively over consecutive cooperative blocks. This indicates that the older children improved their ability to decipher the cooperative intentions of the artificial agent during the cooperative blocks of the task. Thus, our results indicate differences in both model frequency between the age groups, but also that the children best fitted by the MIIL model were also more capable of recognizing cooperative intentions.

Discussion

Learning to adapt one's own decisions to those of another facilitates social interactions. This skill is required for both competitive or cooperative interactions that happens in the playground and classroom at school. Here, we examined how children across age groups adjust their choice behavior across cooperative and competitive blocks and moreover, we characterized the computations required to adapt to such changes in the intentions of others. The strength of our approach was to compare different computational models to assess how children learn to adapt their decisions to those of another. This allowed us to accurately captures, at a mechanistic level, the development of AToM abilities during childhood, avoiding issues of interpretation that can be encountered when relying only on behavior (13,14,34).

Overall, we found that by the age of 8 years old, most children used a computational mechanism relying on a mixed-intention influence learning model that arbitrates between cooperative and competitive intentions to adapt their decisions. Moreover, most children in mid-primary school exploit a cooperativity signature (i.e., intention to cooperate with their partner) better than younger children. Measures of entropy and mutual information based on information theory indicate that older (vs.

younger) children performed especially well in cooperative blocks due to two reasons: 1) higher choice consistency, and 2) weaker reliance on the previous outcome to decide. As a result, the oldest group of children discriminated between competitive vs cooperative blocks more efficiently and behaved more predictably in the cooperative rather than competitive blocks. This may reflect willingness to signal coordination to the other by taking the outcome of previous trials less into account.

From the computational point of view, we found evidence that most children in mid-primary school are able to track variations of another individual's hidden cooperative/competitive intentions. This was reflected in their behavior being best explained by a MIIL model (15) that exploits a control mechanism that gives a higher weight to the strategy with the most reliable prediction (35). In contrast, the best fitting model for the majority of children in KG and EPS groups (children from 3 to 7 years old) was a learning model that computes the mean and variance of the probability of the correct choice for the next trial. Children in this age group neither tracked the intentions of the other agent, nor relied on any mentalizing process for their decision in this task. They rather tracked which card was the most probable at each trial using a simple inference-based strategy.

We found that the tendency of older children (MPS group) to cooperate or compete can be captured by arbitration mechanism proposed by the MIIL model. Moreover, the relative weight of cooperation vs. competition was adjusted over time depending on the competitive vs cooperative blocks. More specifically, the intention to cooperate during cooperative blocks grows stronger with time over the course of consecutive blocks. This suggests that children that understand the intentions of the other player (in this case the AA) improved their performance in this task by realizing that the other player could both cooperate or compete. This indicates that older children (MPS group) can actively learn and adapt to the intentions of the other player. Most children in the younger (KS and EPS) groups did not show this ability except for a few children in the intermediate EPS group. However, the ability of such children to learn and adapt to the AA, as captured by the arbitration weight factor, was far lower than that of children in the oldest group. Note that in all three age groups, there was a strong bias towards competition. This might be because most videogames played on tablets by children at this age, and also in their daily experience, involve competition in one way or another. The competitive bias might also have been fostered by the experimental setup in which children were placed facing each other, which, by default, might encourage a competitive, rather than a cooperative or neutral context, for children.

Our computational modeling analysis indicates a shift in nature of the algorithms, from Bayesian inference model (BSL0) to MIIL, used by children between early primary school (EPS) and mid-primary school (MPS). This is consistent with behavioral reports that core aspects of AToM develop nonlinearly, with children reaching a milestone at the age of 7 years, at which time they attain the conceptual insight that mental states can be recursive (8). Indeed, it has been proposed that a distinct conceptual development occurring in middle childhood at around 7-8 years old could explain improved performance concerning social reasoning and reasoning about ambiguity (8,11). In contrast, a simple and general increase in information-processing capacities might predict a change from Bayesian Sequence Learners of depth 0 to depth 1 (i.e. BSL0 to BSL1) during development, and would not explain such conceptual development in reasoning. Our findings therefore confirm that a conceptual development occurs in children between early primary school (EPS) and mid-primary school (MPS) age. We characterize this change at the computational level, by showing that it is underlined by algorithms of distinct nature. According to this model, most children in the oldest group have achieved the insight that: a) their own previous choices can influence others' behavior and, b) that others can change their intentions during interactions and consequently switch strategies. We find that this conceptual insight occurs at around the same age, 8 years old, as reported in large-scale studies (11,19), that, however, lack characterization of computational mechanisms. This is an important period of cognitive development that might be linked to other complex forms of reasoning, such as scientific reasoning (9,36,37).

The MIIL model that fits most MPS children best is also the model that fitted most of adults best during the same task (15). Children above 8 years old are, in this sense, more similar to adults than younger children. A similar behavioral trend was observed for other repeated economic games (stickers game and sender-receiver game), where children older than 6.5 years of age become more similar to adults than children younger than 6.5 years old (19). Our results therefore indicate that most 8-year-old children have developed the computational mechanisms that they will probably maintain in adulthood to adapt to others' fluctuating intentions or to solve similar decision tasks. This corroborates the findings that rapid changes in decision process occur during childhood, plateau, then get optimized across adulthood (38).

In summary, our work provides evidence that important conceptual developments occur in children at around 8 years old that can be accounted for by the Mixed-Intentions Inference Learning model. We establish that, from mid-primary school onward, most children develop the abilities to adapt to the fluctuating intentions of others to cooperate or compete, and learn to adapt their own strategy as a consequence. They become able to mentalize the influence of their own choices on another's decisions and to arbitrate between cooperative and competitive strategies in a decision game that encourages adaptation to the hidden and fluctuating intentions of others. We identified the computational mechanisms involved and the period when specific aspects of AToM for social reasoning develop during childhood. These involve a conceptual change in the decision-learning process between the younger groups and the oldest group of children, suggesting that conceptual developments fuel some aspects of AToM development in children (7,8,11). Previous studies have addressed diverse consequences of AToM, e.g., children's everyday social experiences (39), their self-understanding (40,41), as well as their epistemological beliefs(42) or their academic abilities (43,44). These abilities are fundamental for children's psychological well-being and their success in life. Therefore, understanding the computational mechanisms underlying components of AToM during childhood has practical relevance to optimize educational and social environments, especially for e-learning applications that model children as computational agents.

We believe that our findings may help build a common computational framework which ties together specific psychological concepts of ToM and computational processes for cognitive science and artificial intelligence (45). Such computational framework should not only make specific, falsifiable predictions about the dynamic interplay between representations of self and other across age groups, but also account for group mentalizing i.e., how we reason about collective beliefs in large groups (46,47).

Materials and Methods

Participants

192 participants (aged 3.4 – 9 years old, M = 6.5, SD = 1.5, 100 females) were recruited from one primary school and one kindergarten in China. 114 participants (aged 6-9, M = 7.5, SD = 1.1, 60 females; 58 aged 6-7, M = 6.5, SD = 0.3, 30 females; 56 aged 8-9, M = 8.5, SD = 0.3, 28 females) were recruited from one primary school in Nanjing, Jiangsu province, China. These participants had normal or corrected-to-normal vision and were naive with regard to the purpose of the experiment. They and their parents gave written informed consent prior to participation. 78 participants (aged 3.4-6, M = 5.0, SD = 0.8, 40 females) were recruited from a kindergarten in Guangzhou city, Guangdong province, China. These participants had normal or corrected-to-normal vision and were naive with regard to the purpose of the experiment. Their parents gave written informed consent prior to participation. The study was approved by a local Ethical Review Board (NJUPSY202111001).

We categorized all above participants into three age groups: Kindergarten (KG) from 3 to 5 years old, Early Primary Schoolers (EPS), 6-7 years old, and Mid-Primary Schoolers (MPS), 8-9 years old.

Mixed intentions task and Artificial Agent

Participants performed a task comprising 154 trials on computer tablets (MS Surface Pro). They were led to believe that they were interacting with another child sitting in front of them, and who was playing the same game with them. In fact, both were playing against an artificial agent (AA) managed by a computer program. This simulated social interaction and allowed us to investigate the dynamics and mechanisms arbitrating between multiple learning algorithms. The experimenter conveyed a hierarchy to the 2 children when they first presented them to each other because, in addition to their names, the experimenter also clearly stated the school grade of each of them. Males played with males, and females with females. The tablet screens showed each participant four cards, two face-down (the other player's cards) and two face-up (their own cards). Participants were informed that to win, they had to choose the card of the same color as that chosen by the other person.

Critically, experimenters were careful not to specify whether the other player was an opponent or a partner i.e., whether the other player had the same victory condition (to match cards) or whether they would win if the cards did not match. Participants were also told that both they and the other player had to make their choices in two seconds (**Fig.1**). If the Artificial Agent (AA) played before the participant, one of the two face-down cards was removed from the playing field. If the participant chose first, only the selected card remained on the playing field. Then, when both had chosen, the cards were revealed. The participant scored points if the card colors matched, otherwise they received nothing. Participants were led to believe that their final payoff would be increased if they scored more points. No information about the other's payoff was given to the participants, they only knew that after an interaction, the other 'participant' would see the same screen but with their own outcomes which could be different for each of them.

Importantly, unbeknownst to the participants, the AA alternated between Competitive and Cooperative trial blocks. During this mixed-intentions task, the AA's strategy was determined by alternating between a Cooperative block of 20 trials that aimed to coordinate between the computer and the player, followed by a Competitive block of 10 trials that aimed to choose the opposite card as the player. The AA algorithm was designed to predict the color that would be chosen by the participant on the basis of a probabilistic analysis of the two previous choices and outcomes. More specifically, the AA calculated the probability p for the participant to select a particular target color based on the history of the two previous choices and their outcomes. It then chose the target that it deemed the most probable to be chosen by the participant in the cooperative mode, and the target that was the least probable in the competitive mode (see **Supplementary Materials** for more details about the AA's algorithm). We defined a competitive choice, made by the AA, as choosing the card of the color the participant was expected not to play, and a cooperative choice as choosing the card with the same color that the participant was expected to play. Thus, the AA exploited the bias of the participants, such that the more predictable the participant was, the more the algorithm made correct competitive or cooperative choices. Participants were not informed of the switches between cooperative and competitive interaction by the AA.

Behavioral analysis

Logistic regression. For the logistic regressions, we reported significant marginal effects of a given variable under the name "estimate" (for example: Cooperativity signature$_{t-1}$: estimate).

$$\ln\left(\frac{p}{1-p}\right) = x_0 + x_1X_1 + x_2X_2 + \cdots$$

X_i represents independent variables and x_i the associated weights in the logistic regression. P represents the probability of a given event. The marginal effect of the variable X_1 is defined as:

$$\widehat{y_1} = mean(logit^{-1}(x_1))$$

The mean is computed across all observed data. Thus, the marginal effect called "estimate" can easily be interpreted as the discreet change of the dependent variable given a unitary change of an independent variable. For the linear regressions, reported "estimate" represents x_i, i.e., the regression coefficient. Indeed, in a linear regression, the marginal effect of a variable is equal to the estimated coefficient.

Information theoretic metrics. To quantify the consistency of the adopted strategy in response to reward feedback, we utilized our previously developed metrics based on information theory (48,49). These include entropy of choice strategy (H(Str)), mutual information between reward outcome and strategy (MIRS), and conditional entropy of reward-dependent strategy (ERDS). Intuitively, the entropy of choice strategy (H(Str)) captures the randomness or uncertainty of the adopted strategy, trial-by-trial. Thus, it signals the overall stochasticity in the choice behavior. MIRS and ERDS values further aim to constrain the extent to which the agent's choice strategy can be predicted by the reward outcome prior

to that decision. Importantly, in our dyadic game settings, the reward feedback is directly predicated upon the other player's action and therefore could signal additional social value.

Specifically, the Shannon entropy of choice strategy, H(Str), is computed using the following equation:

$$H(Str) = -(P(stay) \cdot \log_2 P(stay) + P(switch) \cdot \log_2 P(switch)),$$

where S*tr* is the agent's adopted strategy, coded as *stay* (1), if the agent selects the same target as the previous trial, or *switch* (0) otherwise. Therefore H(Str) measures the level of uncertainty or surprise choice behavior in terms of stay or switch.

Mutual information between reward and strategy (MIRS) is calculated as follows:

$$MIRS = I(Str; Rew)$$
$$= -\sum_{Str\in\{stay,switch\}} \sum_{Rew\in\{win,lose\}} \left(P(Str, Rew) \cdot \log_2 \left(\frac{P(Str,Rew)}{P(Str)\cdot P(Rew)}\right)\right),$$

where *Rew* is the reward outcome (1= win, 0 = lose) on the previous trial.

Finally, the entropy of reward-dependent strategy (ERDS), is the remaining uncertainty in the strategy after accounting for the information given by previous reward outcome:

$$ERDS = H(Str|Rew) = H(Str) - I(Str; Rew) = -\sum_{Str\in\{stay,switch\}} \sum_{Rew\in\{win,lose\}} \left(P(Str, Rew) \cdot \log_2 \left(\frac{P(Str,Rew)}{P(Rew)}\right)\right).$$

These metrics are model-free in the sense that they do not assume any underlying structure or model in the learning and decision-making processes. As such, these measures can be computed directly from any segment of the task, making them especially useful for quantifying behavioral adjustment over time in dynamic, changing environments. For the running average plot, we computed the metrics for each individual using a moving window of 7 trials, such that the data point for trial *t* was computed from trials between *t*−6 and *t*. The values were then averaged for each age group. For calculating the bar plots in the inset, we calculated the metrics for each phase in each subject by compiling the last 10 trials of each Cooperation block and the last 5 trials of each Competition block (i.e., last half of each block) to reduce the contaminating effect of the preceding phase. These values were then averaged across subjects by age groups.

Learning models

To test for a dynamic tracking of implicit intention we compared 5 models: Mixed-Intention Influence Learning (MIIL) model, two Bayesian inference models (BSL), and two non-Bayesian models. The MIIL model involved theory of mind and has been found to best account for detecting mixed-intentions in adults(15).This MIIL model is constructed based on a controller that arbitrates between cooperative and competitive experts of the RL influence models, according to their reliability, before making the decision(15). These RL influence models rely on reinforcement learning to take into account the influence of one's own strategy on the strategy of the other (27,28). The other four models did not include the ability to adapt to changing intentions of others, and thus could not differentiate between cooperation and competition. The non-mentalizing BSL strategy relied on Bayesian inference given past sequences of choices. We used sequences of depths 0 and 1. The other two non-Mentalizing non-Bayesian models were a reinforcement learning model (QL) and a heuristic model 'Win/stay – Lose/Switch' (see **Supplementary Materials** for more details).

Bayesian Model Selection (BMS) was performed using the Variational Bayesian Analysis (VBA) toolbox in a random effect analysis relying on the free energy as the lower bound of model evidence. We used Exceedance Probability (EP)(50) to select the model used most frequently in our population.

Acknowledgements

This work was supported by the Agence Nationale pour la Recherche (ANR-21-CE37-0032) and by the Fondation de France (n°00089590) to J-C. D, by the National Science and Technology Innovation 2030 Major Program to Y.L. (N°2022ZD0205100) by a grant from National Natural Science Foundation of China to CQ (31970982; 32171019), and by a US National Institutes of Health grant to A.S. (R01 DA047870). This research has also benefited from the financial support of IDEXLYON from Université de Lyon (project INDEPTH) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005) and of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency.

Resources availability

Lead contact

Further information and requests for resources and inquiries should be directed to and will be fulfilled by the lead contact, Jean-Claude DREHER (dreher@isc.cnrs.fr). This study did not generate new unique reagents.

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data and codes supporting the main results are available at:
https://osf.io/nf2ma/?view_only=f90dac33122044d19d7e7dea1252c0a2

References

1. Premack D, Woodruff G. Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences [Internet]. 1978 [cited 2022 Jan 10];1(4):515–26. Available from: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/does-the-chimpanzee-have-a-theory-of-mind/1E96B02CD9850016B7C93BC6D2FEF1D0

2. Rakoczy H. Foundations of theory of mind and its development in early childhood. Nature Reviews Psychology 2022 1:4 [Internet]. 2022 Mar 7 [cited 2022 Nov 29];1(4):223–35. Available from: https://www.nature.com/articles/s44159-022-00037-z

3. Ho MK, Saxe R, Cushman F. Planning with Theory of Mind. Trends Cogn Sci. 2022 Nov 1;26(11):959–71.

4. Rakoczy H. Foundations of theory of mind and its development in early childhood. Nature Reviews Psychology [Internet]. 2022 Mar 7 [cited 2022 Aug 23];1(4):223–35. Available from: https://www-nature-com.insb.bib.cnrs.fr/articles/s44159-022-00037-z

5. Call J, Tomasello M. Does the chimpanzee have a theory of mind? 30 years later. Trends Cogn Sci. 2008;12(5):187–92.

6. Wellman HM. Developing a Theory of Mind. In: The Wiley-Blackwell Handbook of Childhood Cognitive Development, Second edition [Internet]. John Wiley & Sons, Ltd; 2010 [cited 2023 Jun 27]. p. 258–84. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/9781444325485.ch10

7. Miller SA. Children's Understanding of Second-Order Mental States. Psychol Bull. 2009;135(5):749–73.

8. Osterhaus C, Koerber S. The Development of Advanced Theory of Mind in Middle Childhood: A Longitudinal Study From Age 5 to 10 Years. Child Dev [Internet]. 2021 Sep 1 [cited 2022 May 13];92(5):1872–88. Available from: https://pubmed.ncbi.nlm.nih.gov/34223634/

9. Osterhaus C, Koerber S, Sodian B. Scientific thinking in elementary school: Children's social cognition and their epistemological understanding promote experimentation skills. Dev Psychol [Internet]. 2017 Mar 1 [cited 2022 Aug 26];53(3):450–62. Available from: /doiLanding?doi=10.1037%2Fdev0000260

10. Osterhaus C, Koerber S. Social cognition during and after kindergarten: The relations between first-order and advanced theories of mind. European Journal of Developmental Psychology [Internet]. 2021 [cited 2022 Jul 4];18(4):573–92. Available from: https://www.tandfonline.com/doi/abs/10.1080/17405629.2020.1820861

11. Osterhaus C, Koerber S, Sodian B. Scaling of Advanced Theory-of-Mind Tasks. Child Dev. 2016;87(6):1971–91.

12. Koerber S, Osterhaus C. Some but not all aspects of (advanced) theory of mind predict loneliness. British Journal of Developmental Psychology [Internet]. 2020 Mar 1 [cited 2022 Aug 22];38(1):144–8. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/bjdp.12302

13. Hayward EO, Homer BD. Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. British Journal of Developmental Psychology [Internet]. 2017 Sep

1 [cited 2022 Jul 4];35(3):454–62. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/bjdp.12186

14. Hughes C, Adlam A, Happe F, Jackson J, Taylor A, Caspi A. Good Test–Retest Reliability for Standard and Advanced False-Belief Tasks across a Wide Range of Abilities. J Child Psychol Psychiat. 2000;41(4):483–90.

15. Philippe R, Janet R, Khalvati K, Rajesh P.N. Rao, Lee D, Dreher JC. Neurocomputational mechanisms engaged in detecting cooperative and competitive intentions of others. Nat Commun [Internet]. [cited 2022 May 13];(in press):in press. Available from: https://doi.org/10.21203/rs.3.rs-1160167/v1

16. Chaminade T, Rosset D, Da Fonseca D, Nazarian B, Lutcher E, Cheng G, et al. How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. Front Hum Neurosci [Internet]. 2012 May 8 [cited 2022 Jan 10];6(MAY 2012). Available from: https://pubmed.ncbi.nlm.nih.gov/22586381/

17. Duéñez-Guzmán EA, Sadedin S, Wang JX, McKee KR, Leibo JZ. A social path to human-like artificial intelligence. Nat Mach Intell. 2023;

18. Bonnefon JF, Rahwan I, Shariff A. The Moral Psychology of Artificial Intelligence. Annu Rev Psychol [Internet]. 2024;75(1). Available from: https://hal.science/hal-04216056

19. Sher I, Koenig M, Rustichini A. Children's strategic theory of mind. Proc Natl Acad Sci U S A [Internet]. 2014 Sep 16 [cited 2023 Jun 23];111(37):13307–12. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.1403283111

20. McCormack T, O'Connor E, Beck S, Feeney A. The development of regret and relief about the outcomes of risky decisions. J Exp Child Psychol. 2016 Aug 1;148:1–19.

21. O'Connor E, McCormack T, Beck SR, Feeney A. Regret and adaptive decision making in young children. J Exp Child Psychol. 2015 Jul 1;135:86–92.

22. Guerini R, FitzGibbon L, Coricelli G. The role of agency in regret and relief in 3- to 10-year-old children. J Econ Behav Organ. 2020 Nov 1;179:797–806.

23. Klindt D, Devaine M, Daunizeau J. Does the way we read others' mind change over the lifespan? Insights from a massive web poll of cognitive skills from childhood to late adulthood. Cortex. 2017 Jan 1;86:205–15.

24. Devaine M, Hollard G, Daunizeau J. Theory of mind: Did evolution fool us? PLoS One. 2014;9(2).

25. Devaine M, Hollard G, Daunizeau J. The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? PLoS Comput Biol. 2014;10(12).

26. Rusch T, Steixner-Kumar S, Doshi P, Spezio M, Gläscher J. Theory of mind and decision science: Towards a typology of tasks and computational models. Neuropsychologia. 2020;146:1–44.

27. Hampton AN, Bossaerts P, O'Doherty JP. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U S A. 2008;105(18):6741–6.

28. Hill CA, Suzuki S, Polania R, Moisa M, O'doherty JP, Ruff CC. A causal account of the brain network computations underlying strategic social behavior. Nat Neurosci. 2017;20(8):1142–9.

29. Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat Hum Behav [Internet]. 2017 [cited 2023 Apr 28];1(4):64. Available from: www.nature.com/nathumbehav

30. Zhang L, Gläscher J. A brain network supporting social influences in human decision-making. Sci Adv. 2020;6(34).

31. Draperi M, Aïte A, Cassotti M, Le Stanc L, Houdé O, Borst G. Development of cool and hot theory of mind and cool and hot inhibitory control abilities from 3.5 to 6.5 years of age. PLoS One [Internet]. 2022 Jan 1 [cited 2023 Dec 14];17(1):e0262251. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0262251

32. Nowak M, Sigmund K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature [Internet]. 1993 [cited 2022 Jan 10];364(6432):56–8. Available from: https://www.nature.com/articles/364056a0

33. Daunizeau J, Rigoux L. Bayesian sequence learning [Internet]. 2014 [cited 2022 Jan 12]. Available from: https://mbb-team.github.io/VBA-toolbox/wiki/BSL/

34. Devine RT, Hughes C. Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. J Exp Child Psychol. 2016 Sep 1;149:23–40.

35. Charpentier CJ, Iigaya K, O'Doherty JP. A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. Neuron [Internet]. 2020 May 20 [cited 2022 Aug 29];106(4):687-699.e7. Available from: https://pubmed.ncbi.nlm.nih.gov/32187528/

36. Kyriakopoulou N, Vosniadou S. Theory of Mind, Personal Epistemology, and Science Learning: Exploring Common Conceptual Components. Front Psychol. 2020 Jun 12;11:1140.

37. Astington JW, Pelletier J, Homer B. Theory of mind and epistemological development: the relation between children's second-order false-belief understanding and their ability to reason about evidence. New Ideas Psychol. 2002 Aug 1;20(2–3):131–44.

38. Giron AP, Ciranka S, Schulz E, van den Bos W, Ruggeri A, Meder B, et al. Developmental changes in exploration resemble stochastic optimization. Nature Human Behaviour 2023 7:11 [Internet]. 2023 Aug 17 [cited 2023 Dec 14];7(11):1955–67. Available from: https://www.nature.com/articles/s41562-023-01662-1

39. Fink E, Begeer S, Peterson CC, Slaughter V, de Rosnay M. Friendlessness and theory of mind: A prospective longitudinal study. British Journal of Developmental Psychology [Internet]. 2015 Mar 1 [cited 2022 Aug 22];33(1):1–17. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/bjdp.12060

40. Bosacki SL. Children's theory of mind, self-perceptions, and peer relations: A longitudinal study. Infant Child Dev [Internet]. 2015 Mar 1 [cited 2022 Jul 4];24(2):175–88. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/icd.1878

41. Bosacki S, Moreira FP, Sitnik V, Andrews K, Talwar V. Theory of Mind, Self-Knowledge, and Perceptions of Loneliness in Emerging Adolescents. Journal of Genetic Psychology [Internet]. 2020 Jan 2 [cited 2022 Aug 22];181(1):14–31. Available from: https://www.tandfonline.com/doi/abs/10.1080/00221325.2019.1687418

42. Weinstock M, Israel V, Cohen HF, Tabak I, Harari Y. Young Schoolchildren's Epistemic Development: A Longitudinal Qualitative Study. Front Psychol. 2020 Jun 26;11:1475.

43. Lecce S, Bianco F, Devine RT, Hughes C. Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. J Exp Child Psychol. 2017 Nov 1;163:69–86.

44. Ebert S. Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. J Exp Child Psychol. 2020 Mar 1;191:104739.

45. Barnby JM, Dayan P, Bell V. Formalising social representation to explain psychiatric symptoms [Internet]. Vol. 27, Trends in cognitive sciences. The Authors; 2023. p. 317–32. Available from: https://doi.org/10.1016/j.tics.2022.12.004

46. Khalvati K, Mirbagheri S, Park SA, Dreher JC, Rao RPN. A Bayesian theory of conformity in collective decision making. In: Advances in Neural Information Processing Systems [Internet]. 2019 [cited 2023 Jan 31]. Available from: https://dl.acm.org/doi/10.5555/3454287.3455157

47. Khalvati K, Park SA, Mirbagheri S, Philippe R, Sestito M, Dreher JC, et al. Modeling other minds: Bayesian inference explains human choices in group decision-making. Sci Adv [Internet]. 2019 Nov 27 [cited 2023 Jan 31];5(11). Available from: https://pubmed.ncbi.nlm.nih.gov/31807706/

48. Trepka E, Spitmaan M, Bari BA, Costa VD, Cohen JY, Soltani A. Entropy-based metrics for predicting choice behavior based on local response to reward. Nat Commun [Internet]. 2021 Nov 12 [cited 2023 Mar 21];12(1):1–16. Available from: https://www.nature.com/articles/s41467-021-26784-w

49. Woo JH, Aguirre CG, Bari BA, Tsutsui KI, Grabenhorst F, Cohen JY, et al. Mechanisms of adjustments to different types of uncertainty in the reward environment across mice and monkeys. Cogn Affect Behav Neurosci [Internet]. 2023 Jun 1 [cited 2023 Aug 17];23(3):600–19. Available from: https://link.springer.com/article/10.3758/s13415-022-01059-z

50. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - Revisited. Neuroimage [Internet]. 2014 Jan 1 [cited 2022 Jan 12];84:971–85. Available from: https://pubmed.ncbi.nlm.nih.gov/24018303/

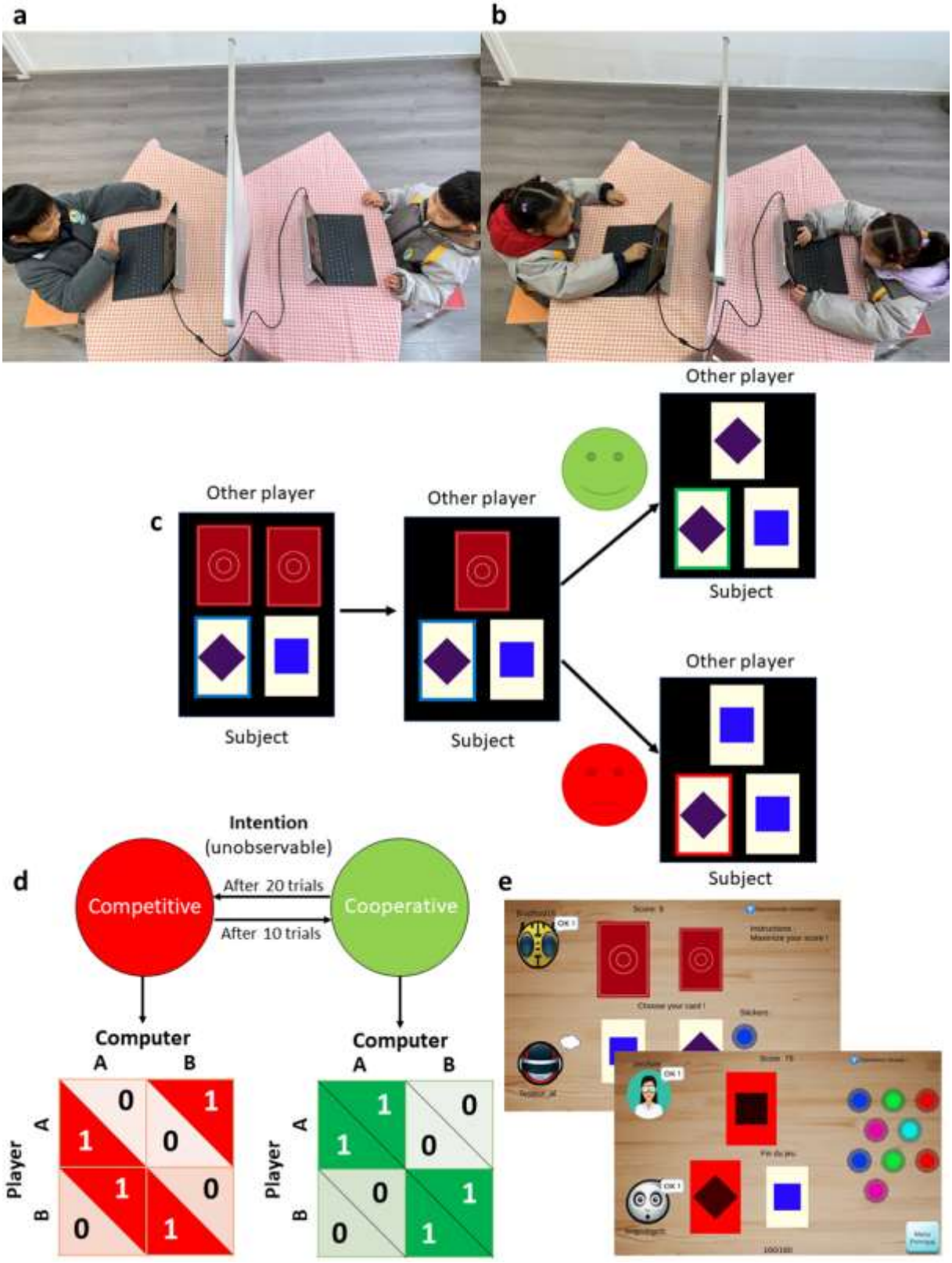


Fig. 1. Description of the procedure to study adaptation to fluctuating competitive and cooperative intentions of others in children. (a)-(b) Experimental setup. Two children were tested on individual tablets. They played facing each other, so each could see the other, but not the other's tablet. They were told that their tablets were connected, that the other player represented on their tablet was the child sitting opposite, and that each of them had to maximize their own score to maximize their final reward. In reality both children were actually playing with an Artificial Agent (AA). Males were paired with males and females with females. **(c) Experimental protocol.** The two players both chose one card

among two cards presented. When both children had chosen, the chosen cards were revealed, and if the card they had chosen was the same as that chosen by the AA, they scored points. If their card was not the same as that indicated for them by the AA, that child was not rewarded. **(d) Payoff matrix for the two possible modes of the Artificial Agent (AA).** The AA that decided which choice was rewarded for the participant operated in either a competitive or a cooperative mode, based on the payoff matrices depicted above. **(e) Snapshots of the game on tablets.** The player chose one of the two cards presented on the lower part of the screen. The jewels on the right represent the score of the child. The goal of the game was to score as many points as possible, by selecting the correct card among the two cards of the player (displayed at the bottom of the screen), knowing that the other player's cards were displayed on top of the screen.

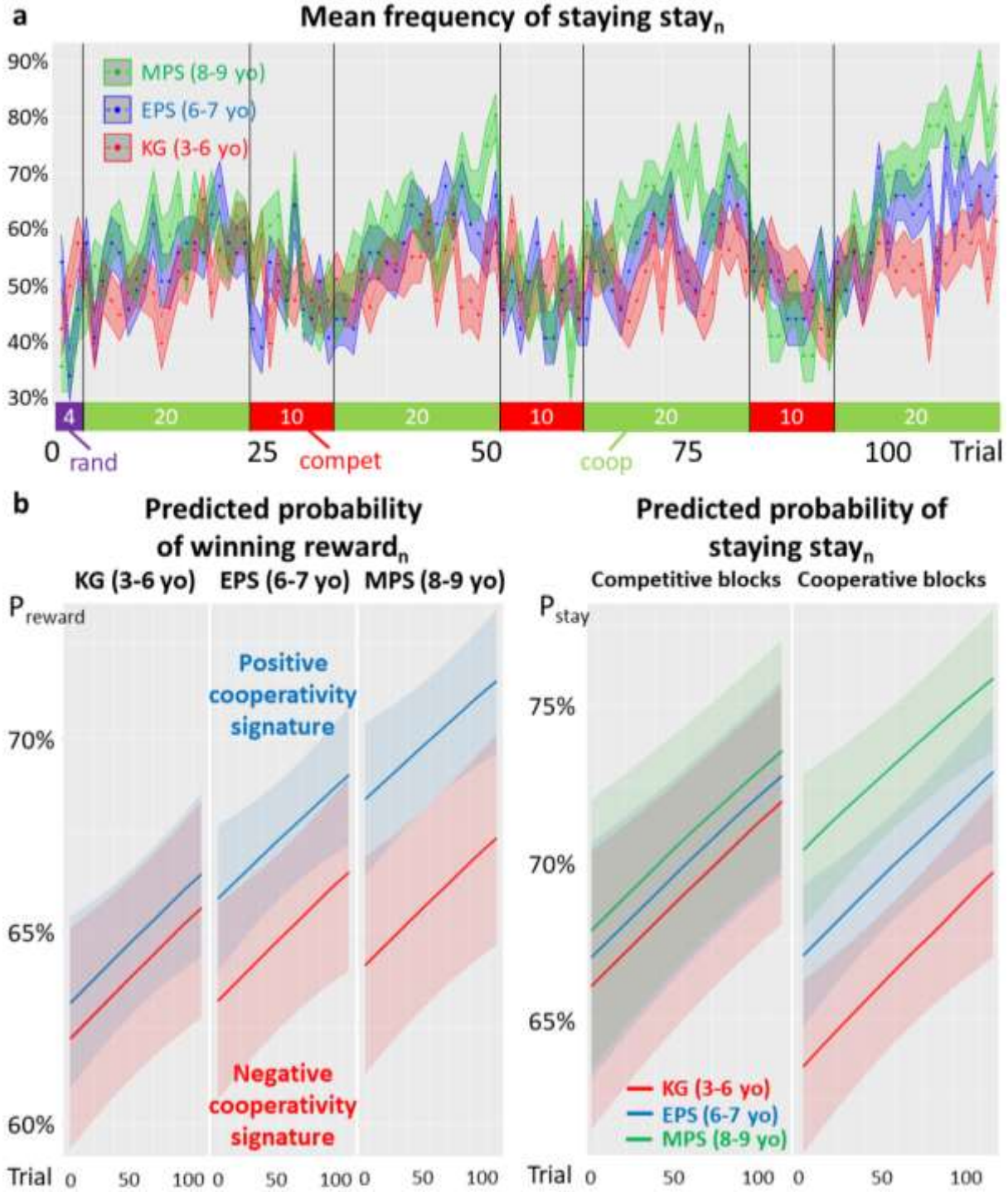


Fig. 2. Behavioral results. (a) Mean probability for the children to choose the same target as the one they chose in the previous trial ("stay"). The mean frequency to "stay" is represented in green, blue and red for children in mid-primary school (MPS), early primary school (EPS), and kindergarten (KG) respectively, with their respective standard errors (ribbons). The purple bar represents the first 4 initial trials in which the Artificial Agent (AA) played randomly for initialization purposes. The green bars represent the cooperative blocks and the red bars the competitive blocks. **(b) Generalized Mixed Effects (GLME) models, marginal effects. Left**: predicted probability for the subjects to choose the correct target in trial n, reward$_n$, plotted against the trial number (x-axis), after a positive cooperativity signature (blue) vs after a negative cooperativity signature (red) for different ages. **Right**: predicted probability for the subjects to choose the same target at trial n as previously (at trial n-1), stay$_n$ (right). The older the child, the better they succeed in coordinating with the AA after a positive cooperativity signature and the more stable they behave in the cooperative blocks.

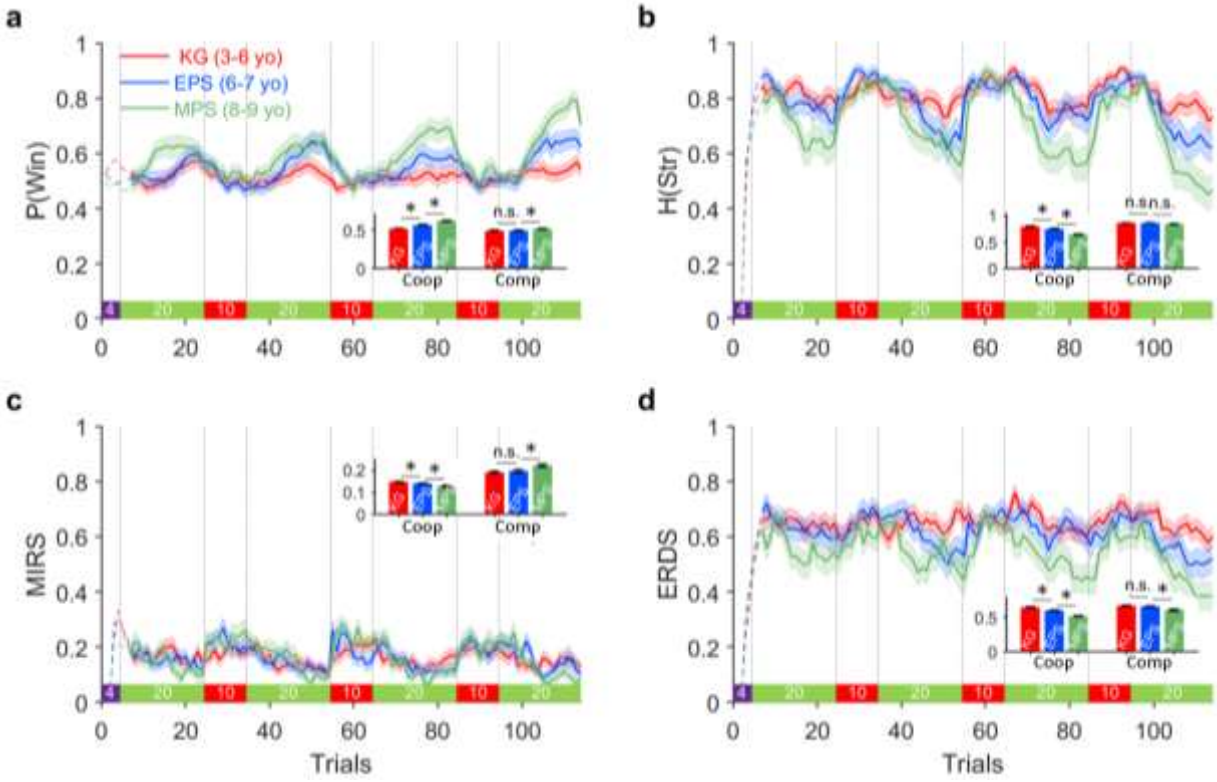


Fig. 3. Model-free analysis based on information theory. (a) Probability of winning reward over trials. Running averages of the probability of reward are computed using a moving window of 7 trials. Performance steadily increased with age during the cooperative blocks, but only slightly increased in older children in the competitive blocks. Inset shows the mean value for each group during each phase (Cooperative and Competitive). **(b) Shannon entropy for stay vs. switch over trials H(Str).** During the cooperative phase, improvement in performance with age was accompanied by decrease in H(str) as older children were better at choosing less randomly. No overall effect was observed during the competitive blocks. **(c–d) Mutual Information between reward outcome on the previous trial and the adopted choice strategy (MIRS) and entropy of reward-dependent strategy (ERDS)**(48). During the cooperative blocks, both MIRS and ERDS decreased in older children, showing that older children chose less randomly, and that they were also able to base their choice strategy less strongly on reward outcome. In contrast, during the competitive blocks, larger MIRS in 8-9 yo compared with 6-7 yo was accompanied by smaller ERDS in 8-9 yo compared with 6-7 yo. (Shaded error bar = S.E.M.)

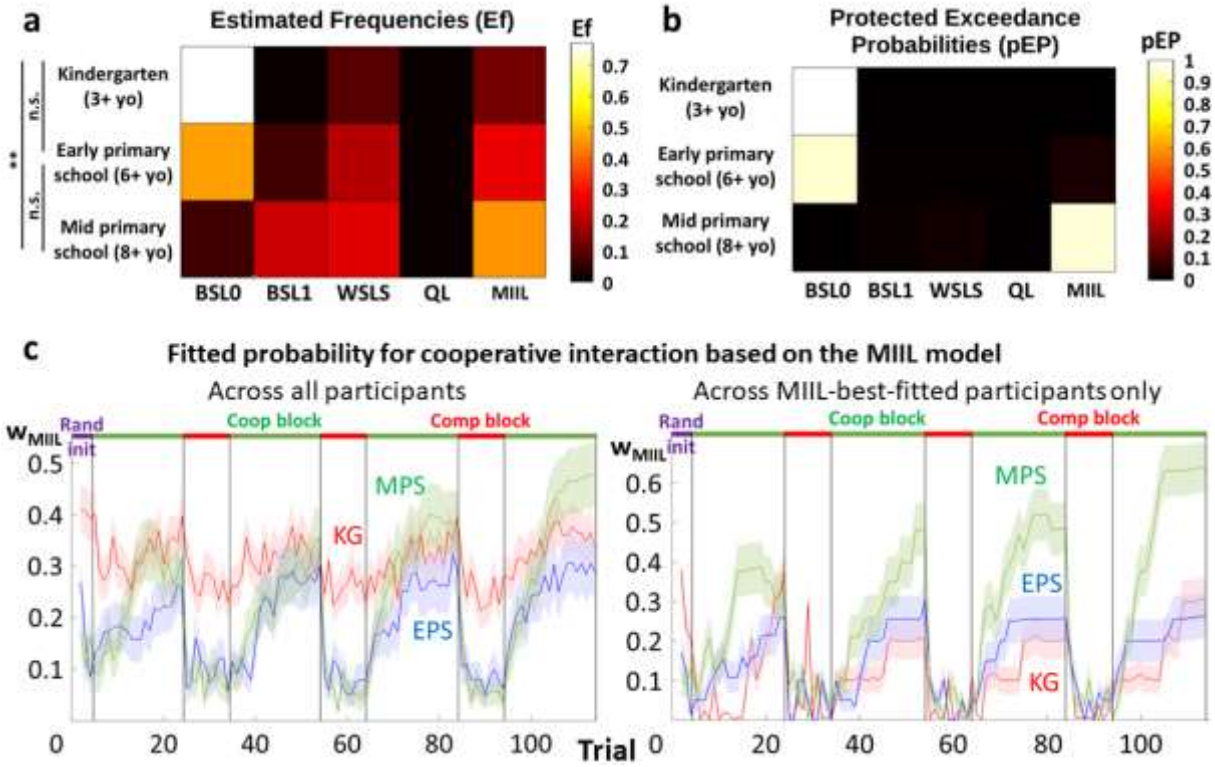


Fig. 4. Computational learning models. (a)-(b) Bayesian model comparison. x-axis: model names; y-axis: age group; Color-scale: (a) estimated frequencies (Ef), i.e., the frequency at which one model fitted one subject best in the group population compared to the other tested models; (b) protected exceedance probabilities (pEP), i.e., the probability that measures how likely it is that any given model is more frequent than all other models in the comparison set, corrected for the possibility that observed differences in model evidence (over subjects) are due to chance (50). In the Kindergarten (KG) and Early Primary School (EPS) groups, the model that best fitted most of our group of participants is the BSL0. In Mid-Primary School, the MIIL model best fitted most of the participants. **(c) Mean estimated cooperation probability computed from the MIIL model.** x-axis: trial number. y-axis: w_{MIIL}, the mean probability, according to the MIIL model, that the participant attempts to cooperate (left) across all participants (group by group) and (right) across only those participants that were best fitted by the MIIL model for the 114 trials (group by group). This indicates that children in kindergarten fail to use sophisticated MIIL computations. The initial purple area shows the 4 random initializing trials, green areas are the Cooperative blocks, and red areas are the Competitive blocks.

Supporting Information for

Computational mechanisms underlying the emergence of theory of mind in children

Toan Nong[1,2], Yansong Li[3,4,*], Chen Qu[6,*], Jae Hyung Woo[7,**], Yuqian Wang[3,**], Chengguo Miao[3,**], Xiaoyi Liu[3,**], Qingxia Liu[6,**], Jean-Baptiste Van der Henst[5], Rémi Philippe[1,2],Edmund Derrington[1,2], Alireza Soltani[7], Jean-Claude Dreher[1,2]

1- *ISC-MJ, CNRS UMR5229, Neuroeconomics Lab, Bron, France*
2- *University of Lyon 1, UCBL1, Lyon, France*
3- *Reward, Competition and Social Neuroscience Lab, Department of Psychology, School of Social and Behavioral Sciences, Nanjing University, Nanjing, China*
4- *Institute for Brain Sciences, Nanjing University, Nanjing, China*
5- *CRNL-CNRS-UCBL1, Bron, France*
6- *Center for Studies of Psychological Application, School of Psychology,South China Normal University, Guangzhou, China*
7- *Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA.*
8- **,**: equal contributions*

This PDF file includes:

Learning Models

All of these models were initially fitted separately to each participant and in each condition. We applied a Variational Bayes Approximation to fit and optimize all of the tested models based on the Free Energy criterion(1) with the VBA-toolbox(2) in Matlab. All priors were set to default value except for the sigma priors (covariance matrix of priors) that was set to 10 times the identity matrix. This ensured we had good fits of the tested models to the data and accounted for variability in behavior (see **Supplementary Materials**). We then compared all the tested models with a group-level random-effect Bayesian Model Selection (BMS)(3,4) based on the Free Energy criterion with the VBA toolbox.

Bayesian Sequence Learning (BSL) model. The BSL model(5) tracks the (log) odds of $P(o_t=1|o_{t-K})$, where o is the partner's choice (binary outcome). This variable is updated according to a Laplace-Kalman filter, yielding 2 sufficient statistics (mean and variance) per combination of past outcomes. BSL can learn sequences of arbitrary depth K. For example:

- if K=1, then BSL tracks 2 probabilities, namely: $P(u_t=1|u_{t-1}=1)$ and $P(u_t=1|u_{t-1}=0)$. In this case, BSL needs to know about the partner's previous move u_{t-1}.
- if K=2, then BSL tracks 4 probabilities, namely: $P(u_t=1|u_{t-1}=1, u_{t-2}=1)$, $P(u_t=1|u_{t-1}=0,u_{t-2}=1)$, $P(u_t=1|u_{t-1}=1,u_{t-2}=0)$ and $P(u_t=1|u_{t-1}=0,u_{t-2}=0)$. In this case, BSL needs to know about the partner's two previous moves u_{t-1} and u_{t-2}. etc.

More generally, the BSL model tracks 2^K probabilities.

Win-Stay Lose-Switch (WSLS) model. This heuristic(6) keeps the same target as the previous one if previous reward was positive, and changes targets otherwise.

Q-Learning (QL) model. The constituents of this model in its simplest form are:

- a set of (action) hidden value states. In two-armed bandit problems, there are two of these (n in a n-armed). In general, there will be as many values as there are available actions. Some behavioral biases can be captured by the initial conditions on these hidden value states.

- a learning rate. This parameter controls the impact of reward prediction error on the value update. Note that one may want to ask whether the learning rate depends upon experimental factors (pathological condition, gain/loss domains, etc…)

- a behavioral temperature and a bias. These parameters control the exploitation vs exploration ratio of the agent, and capture choice randomness.

Q-learning models simply assume that subjects update the value of possible actions. In its simplest form, the Q-learning algorithm expresses the change in value Q(t+1)-Q(t) from trial t to trial t+1 as being linearly proportional to the prediction error. This yields the following learning rule:

$$Q_{t+1} = Q_t + \alpha * (r_{t+1} - Q_t)$$

Where rt is the reward delivered to the subject at trial t, and α is the (unknown) "learning rate" of the subject.

The Q-learning evolution function thus takes the agent's previous action and the feedback received for the previous action as data inputs. We complement with a softmax decision rule in the Q-learning observation function. As a reminder, this is an equation that expresses the probability Pt(ai) of the subject to choose action ai at trial t:

$$P_t(a_i) = \frac{\exp(\beta * Q_t(a_i))}{\sum_j \beta * Q_t(a_j)}$$

Where β is the (unknown) temperature.

Fitting the above Q-learning model to behavioral data means finding estimates of the learning rate α, the inverse temperature β, and the initial values Q0 that best explains the observed choices.

Influence Learning model. This mentalizing model was described by Hampton et al(7), who also found neural correlates with this model in humans(7). The influence model relies on a Taylor expended reinforcement learning. This model describes computations underlying the capacity to mentalize (in the context of a strategic game). Let $p_n = P_n(o = 1)$ be the agent's prediction of the other's next move, i.e., the probability that the other will pick the first alternative option. The "influence learning" rule can be written as follows:

$$p_{n+1} = p_n + \eta * (o - p_n) - \lambda * k_1 * p_n * (1 - p_n) * (a - \frac{\beta * x_n - k_2}{k_1}) \qquad (Eq.\ 0)$$

where o is the other's last move, a is the agent's last move, η is the weight of the agent's prediction error, λ is the weight of the other's prediction error, k_1 & k_2 are derived from the game's payoff table, x_n is the tracked log-odds of p_n, β is the other's temperature, i.e. the tendency of the other to explore vs exploit information (the closer it is to 0, the more it explores by choosing randomly without using previous information, and the closer it is 1, the more it exploits previous information and chooses deterministically).
Note that for all subjects, the tested models fitted better than chance.

Mixed-Intentions Influence Learning (MIIL) model.

The Mixed-Intentions Influence Learning model extends the Influence model described by Hampton et al (7). The original influence model considers how one's own actions influence the other's future actions. Thus, to compute the probability of updating of the other's strategy, we replaced update of opponent strategy ($Eq.1$) in the player decision rule ($Eq.2$).

$$p_{t+1}^{*} = p_t^{*} + \eta * \delta_t^{p} \qquad (Eq.1)$$

$$\begin{array}{ll} q = & s(2p^{*} - 1) \quad in\ cooperative\ mode \\ q = & s(1 - 2p^{*}) \quad in\ competitive\ mode \end{array} \qquad (Eq.2)$$

Then, with a Taylor expansion taking η close to 0, we added the influence terms (Δp : influence update signal of the participant, Δq : influence update signal of the other):

$$\Delta q \approx +\eta 2\beta q_t(1 - q_t)(P_t - p_t^{*}) \qquad (Eq.3)$$

$$\begin{array}{ll} \Delta p \approx & +\eta 2\beta p_t(1 - p_t)(Q_t - Q_t^{*}) \quad in\ cooperative \\ \Delta p \approx & -\eta 2\beta p_t(1 - p_t)(Q_t - Q_t^{*}) \quad in\ competitive \end{array} \qquad (Eq.4)$$

Thus, in the competitive mode, there is only a sign difference between the term of influence of the two players which is not the case in the cooperative setting. A player can thus incorporate the influence of his/her action on the strategy of the other player:

$$\begin{array}{l} p_{t+1}^{*} = p_t^{*} + \eta_1(P_t - p_t^{*}) + \eta_2 2\beta p_t^{*}(1 - p_t^{*})(Q_t - q_t^{**}) \quad in\ cooperative \\ p_{t+1}^{*} = p_t^{*} + \eta_1(P_t - p_t^{*}) - \eta_2 2\beta p_t^{*}(1 - p_t^{*})(Q_t - q_t^{**}) \quad in\ competitive \end{array} \qquad (Eq.5)$$

$$q_{t+1}^{*} = q_t^{*} + \eta_1(Q_t - q_t^{*}) + \eta_2 k_1 2\beta q_t^{*}(1 - q_t^{*})(P_t - p_t^{**}) \qquad (Eq.6)$$

The Influence model update rules. p_t^{*} Is the predicted opponent strategy. P_t is the opponent choice and then $(P_t - p_t^{*})$ is the action prediction error. The influence update is due to the $(Q_t - q_t^{**})$ term. Thus Q_t is the player's own action and q_t^{**} the inferred probabilities that the opponent has of the player themself (second-order beliefs). Thus, in the cooperative and competitive modes the influence will occur in the opposite directions. In the mixed-intention influence model, we decline p_{t+1}^{*} in $p_{t+1}^{coop\ *}$ and $p_{t+1}^{comp\ *}$. To compute the p_t^{**} and q_t^{**}, we invert the decision function ($Eq.2$):

$$\begin{array}{ll} q_t^{**} = & \frac{1}{2} - \frac{1}{2\beta} ln\left(\frac{1 - p^{*}}{p^{*}}\right) \\ p_t^{**} = & \frac{1}{2} + \frac{1}{2\beta} ln\left(\frac{1 - p^{*}}{p^{*}}\right) \quad in\ competitive \\ p_t^{**} = & \frac{1}{2} - \frac{1}{2\beta} ln\left(\frac{1 - q^{*}}{q^{*}}\right) \quad in\ cooperative \end{array} \qquad (Eq.7)$$

The influence learner uses the inferred probability that the other chooses *a*, the payoff matrix and the other temperature to compute a decision value:

$$DV = \frac{p^{*}}{Temperature} * \left(Payoff(\ self = a, other = a) - Payoff(self = a, other = \bar{a})\right) +$$

$$\frac{(1 - p^{*})}{Temperature} * \left(Payoff(self = \bar{a}, other = a) - Payoff(self = \bar{a}, other = \bar{a})\right) \qquad (Eq.8)$$

The Mixed-Influence model makes a difference between the Cooperative and Competitive modes. The mixed-intentions model makes the assumption that the cooperative expert and the competitive expert come from the same influence model, because, from the point of view of the participants, there is no indication that there are two modes of interaction. Therefore, it is more parsimonious to assume that a single process (i.e. same computational model for both experts) is engaged along the task.

For the mixed-intentions setting, we ran the competitive and cooperative models in parallel, avoiding the need for the payoff matrix to be learnt. On the first trial, each expert gives a prior probability that the other would choose the "*action a*" ($p^{*}_{coop,0} = p^{*}_{comp,0} = 0.5$), then each expert follows its own walk in generating on each trial the probability that the other will choose option a for each possible mode of interactions, P^{a}_{comp} and P^{a}_{coop}. We then transformed the probability with the inverse sigmoid function to have values ranging from $-\infty$ to $+\infty$. We have a binomial choice configuration thus $V^{a} = -V^{b}$ in both competitive and cooperative settings. Thus, as $V_i^a$and V_i^b get close to zero, uncertainty for *i*, the other's intention, increases. We defined the reliability of the intention *i* as the absolute value of V_i^a and the probability that the other intention is cooperative as the sigmoid function of the difference in reliability between the two modes:

$$P_{coop}^{t} = \frac{1}{1 + e^{\beta(|V_{coop}^{t}| - |V_{comp}^{t}| - \delta)}} \qquad (Eq. 9)$$

where β is the inverse temperature controlling for the stochasticity of the mode of interaction and δ is the bias towards cooperative mode.
Then, with $P_{comp}^{a,t}$ and $P_{coop}^{a,t}$ we computed the decision value given the Competitive and Cooperative payoff matrix, DV_{comp}^{a} and DV_{coop}^{a} respectively and weighted them by the probability of the corresponding mode of interaction to compute the total decision value:

$$DV^{t} = P_{coop}^{t} * DV_{coop}^{a} + \left(1 - P_{coop}^{t}\right) * DV_{comp}^{a} \qquad (Eq. 10)$$

We call $P_{coop}^{t} * DV_{coop}^{a}$ the cooperative component of the model and $\left(1 - P_{coop}^{t}\right) * DV_{comp}^{a}$the competitive component of the model. The sigmoid function s generated the probability of selecting choice *a* at trial t:

$$p^{a,t} = s(DV^{t}) \qquad (Eq. 11)$$

Finally, the reward prediction error was defined as the reward at trial t for action a:

$$PE = R^{a,t} - p^{a,t} \qquad (Eq. 12)$$

In summary, the 'mixed intentions' version of the influence model computes one decision value according to a competitive expert and another according to a cooperative expert and then arbitrates between the two, based on the difference in their respective reliability (see **Supp. Fig. 1**). We defined reliability as the difference in unsigned value functions for two choices given by particular learning algorithms.

Artificial Agent

The AA calculated the probability p for the participant to select a particular card color based on the history of the two previous choices and their outcomes. It then chose the card that it deemed the most probable to be chosen by the participant in the cooperative mode, and the card that was the least probable in the competitive mode. The artificial agent (AA) selected its card according to the probability for the player to choose a specific card after a given sequence of events. By sequence of events, we mean a specific set of two consecutive choices and outcomes. If we consider two choices, A and B, and two possible outcomes, win (W) and lose (L), an example for a sequence of events could be: player chose A and won at trial n-2, then chose A and lost at n-1. We will sum up this sequence as: A_WA_L.
To determine its choice at trial n, the AA will check in its memory table what choice the participant usually followed with after such sequence of events, based on the previous history of actions of the player. For example, let's assume that the sequence of events A_WA_L just occurred. The AA must now decide its move at trial n. Let's also assume that this sequence of events already happened four times in the past, and that the player chose to play card A 75% of the time and card B 25% of the time. Then, the algorithm in the cooperative mode will decide to play card A with a 75% probability for its next move, and card B with a 25% probability. Vice-versa in the competitive mode, i.e. card A with 25% of probability and card B with 75% of probability. If we assume that the AA as well as the player choose card A for trial n, then

the updated history of choices for the specific sequence of events A_WA_L would become: card A chosen at 80%, and card B at 20%. Also, the current sequence of events will become A_LA_W.
Let's assume that the current memory table is the one shown in **Supp. Tab. 7**. If we take the same example as described in the previous paragraph, the **current history of events** is "…**A_WA_L**...". After the AA and participant choose A, the history of events becomes "…**A_WA_L***A_W*" and the memory table becomes **Supp. Tab. 8**. Now let's assume that the *sequence* history is "…**A_WA_L***$A_WB_LB_LA_WB_W$*…". The sequence of the subject's choices in this example would be "…**AA***ABBAB*…". Consequently, in this example, the AA's choices would be "…**AB***AAAAB*…". Following the same reasoning as presented in the previous paragraph, we can update the memory table. It will become at the end of this sequence of event the memory table shown in **Supp. Tab. 9**.
For the AA, in the cooperative mode, the sequence of choices and probabilities would be "…**A(/)B(/)***A(75%)A(100%)A(100%)A(100%)B(50%)*…", where the percentage of probability of the choice of the algorithm is given in parentheses and where "/" are undefined probabilities in this example because they depend on previous events before this sequence occurred and that we did not define. For this study in children, we used a greedy version of this algorithm. It means that probabilities strictly greater than 50% were rounded at 100% while any probability strictly lower than 50% was rounded at 0%. This choice of implementation was made to adapt and simplify the difficulty of the task to children.

GLMs & GLMEs.

We first modeled our GLMs based on the following formulas:

$$\begin{aligned} Reward_n \sim 1 &+ Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-2} + Stay_{n-3} \\ &+ AA\_Stay_{n-1} + AA\_Stay_{n-2} + AA\_Stay_{n-3} + Coop\_sig_{n-1} + Coop\_sig_{n-2} + Trial \\ &+ AA\_mode + Hierar\_cond + Sex + Age + DT + (Age : Coop\_sig_{n-1}) \\ &+ (Age : Coop\_sig_{n-2}) + (Age : AA\_mode) \end{aligned}$$

$$\begin{aligned} Stay_n \sim 1 &+ Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-2} + Stay_{n-3} \\ &+ AA\_Stay_{n-1} + AA\_Stay_{n-2} + AA\_Stay_{n-3} + Coop\_sig_{n-1} + Coop\_sig_{n-2} + Trial \\ &+ AA\_mode + Hierar\_cond + Sex + Age + DT + (Age : Coop\_sig_{n-1}) \\ &+ (Age : Coop\_sig_{n-2}) + (Age : AA\_mode) \end{aligned}$$

After applying a dropterm() process based on the AICs of the models successively, the final models were:

$$\begin{aligned} Reward_n \sim 1 &+ Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-3} + AA\_Stay_{n-1} \\ &+ AA\_Stay_{n-2} + AA\_Stay_{n-3} + Coop\_sig_{n-1} + Coop\_sig_{n-2} + Trial + Age \\ &+ (Age : Coop\_sig_{n-1}) \end{aligned}$$

$$\begin{aligned} Stay_n \sim 1 &+ Reward_{n-1} + Reward_{n-3} + Stay_{n-1} + Stay_{n-2} + Stay_{n-3} + AA\_Stay_{n-1} \\ &+ AA\_Stay_{n-2} + AA\_Stay_{n-3} + coop\_sig_{n-1} + coop\_sig_{n-2} + Trial + Age \\ &+ AA\_mode + (AA\_mode : Age) + (Age : coop\_sig_{n-2}) + Sex + Condition \end{aligned}$$

To control for random effects and the interaction and autocorrelation between effects of each variable over time, we used Generalized Linear Mixed-Effects (GLME) models implemented with the glmer() function of lme4 package in R with a binomial family. We ensured that we had good estimates of random effects and accounted for variability in behavior using Bound Optimization by Quadratic Approximation. GLME1 assessed the children's tendency to select the same target as that in the previous trial ("stay") and GLME2 assessed their performance in successfully finding the same card as the AA ("reward"). We proceeded to further simplifications through another dropterm() process which led to the following formula:

$$\begin{aligned} Reward_n \sim 1 &+ Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-3} + AA\_Stay_{n-1} \\ &+ AA\_Stay_{n-2} + AA\_Stay_{n-3} + Coop\_sig_{n-1} + Coop\_sig_{n-2} + Trial + Age \\ &+ (Age : Coop\_sig_{n-1}) + (1 + Trial \,|\, ID) \end{aligned}$$

$$\begin{aligned} Stay_n \sim 1 &+ Reward_{n-1} + Reward_{n-3} + Stay_{n-2} + AA\_Stay_{n-1} + AA\_Stay_{n-2} + AA\_Stay_{n-3} \\ &+ Coop\_sig_{n-1} + Coop\_sig_{n-2} + Trial + Age + AA\_mode + (AA\_mode : Age) \\ &+ (Age : Coop\_sig_{n-2}) + (1 + Trial \,|\, ID) \end{aligned}$$

We then proceeded to an addterm() process with the regressors that were previously removed from the GLMs. This led to the following formula:

$$Reward_n \sim 1 + Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-3} + AA\_Stay_{n-1} + AA\_Stay_{n-2} + AA\_Stay_{n-3} + Coop\_sig_{n-1} + Coop\_sig_{n-2} + Trial + Age + (Age: Coop\_sig_{n-1}) + (1 + Trial\,|ID)$$

$$Stay_n \sim 1 + Reward_{n-1} + Reward_{n-3} + Stay_{n-2} + AA\_Stay_{n-1} + AA\_Stay_{n-2} + AA\_Stay_{n-3} + Coop\_sig_{n-1} + Coop\_sig_{n-2} + Trial + Age + AA\_mode + DT + (AA\_mode: Age) + (Age: Coop\_sig_{n-2}) + (1 + Trial\,|ID)$$

The final results and outputs of the GLMs and GLMEs are summed up in **Supp. Tab. 10**.

Control analyses and model checks

We ran several control analyses. We first conducted a Confusion Analysis (CA) as a model check to ensure that the compared models that we tested, given our data, were not confused between each other and that our model selection was indeed reliable(8). We therefore generated data multiple times with Monte-Carlo simulations (approx. 2800 in total) for each of our tested models with different sets of parameters, and conducted the same analysis as presented previously on the simulated data. This meant fitting the data with all of the tested models, then performing a BMS on the models to select the one that fitted best. If this analysis is reliable and the models are distinguishable, then the model that we simulated at the beginning should also be that selected by our analysis. The confusion matrix (**Supp. Fig.2.a**) showed that the final selected model was the one which most of the time initially generated the data.

Next, we checked if the learning parameters of the MIIL model could be identified from observed data. To do so, we generated simulated data with multiple Monte-Carlo simulations (approx. 9800 in total) from the influence model using different sets of learning parameters sampled from a uniform distribution. The sampling was done (uniformly) between the most extreme empirical values found in our data analysis for each of our seven learning parameters. We then fitted the generated data with the influence model with the same method as previously. We then regressed the estimated parameters on the simulated parameters. We found for each learning parameter a positive correlation between estimated and simulated parameters (**Supp. Fig. 2.b**). The regression indicates that the estimated parameters were actually underestimated compared to those used to generate simulated data.

Supplementary analyses

We initially fitted and compared other models than those presented in the main results: 2 influence models(7), one in a purely competitive setup and one in a purely cooperative setup (see **Supp. Fig.3,** top). However, they did not pass the confusion analysis (see **Supp. Fig.3,** down), i.e., they were confused with the BSL0 model when simulated, then tested through our analysis. This implies that, for our main results, the overrepresentation of the BSL0 model in the younger groups might actually hide the presence of cooperative and competitive Influence models among them. However, this does not change the main result that a transition in model representation occurs around age 8 in MPS when the MIIL model becomes preponderant. Still, the results of the comparison between all tested models is given here for readers' information (**see Supp. Fig.3,** top).

References

1. Daunizeau J, Rigoux L. The variational Bayes approach [Internet]. 2014 [cited 2022 Jan 12]. Available from: https://mbb-team.github.io/VBA-toolbox/wiki/The-variational-Bayesian-approach/

2. Daunizeau J, Adam V, Rigoux L. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. PLoS Comput Biol [Internet]. 2014 [cited 2022 Jan 10];10(1):1003441. Available from: www.ploscompbiol.org

3. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. Neuroimage [Internet]. 2009 Jul 15 [cited 2022 Jan 12];46(4):1004–17. Available from: https://pubmed.ncbi.nlm.nih.gov/19306932/

4. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - Revisited. Neuroimage [Internet]. 2014 Jan 1 [cited 2022 Jan 12];84:971–85. Available from: https://pubmed.ncbi.nlm.nih.gov/24018303/

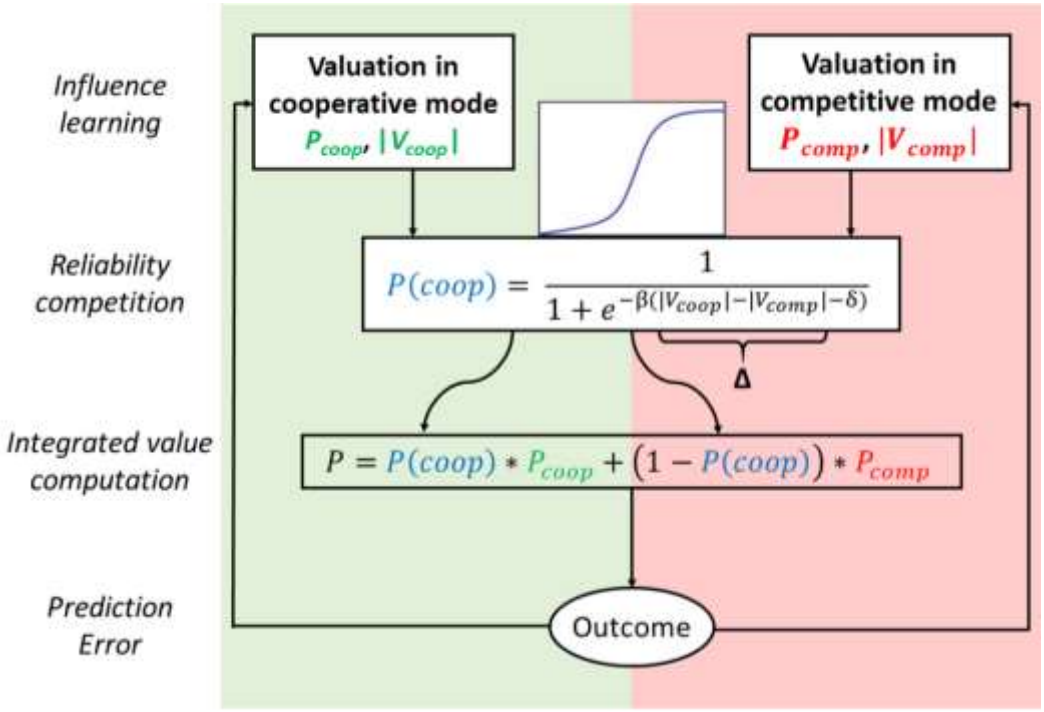
5. Daunizeau J, Rigoux L. Bayesian sequence learning [Internet]. 2014 [cited 2022 Jan 12]. Available from: https://mbb-team.github.io/VBA-toolbox/wiki/BSL/

6. Nowak M, Sigmund K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature [Internet]. 1993 [cited 2022 Jan 10];364(6432):56–8. Available from: https://www.nature.com/articles/364056a0

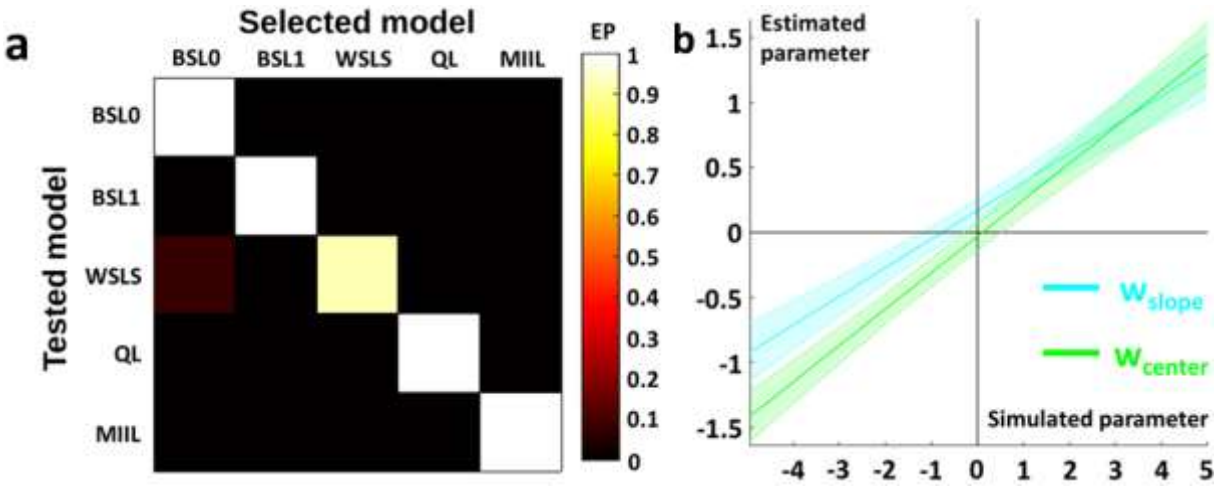
7. Hampton AN, Bossaerts P, O'Doherty JP. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U S A. 2008;105(18):6741–6.

8. Palminteri S, Wyart V, Koechlin E. The Importance of Falsification in Computational Cognitive Modeling [Internet]. Vol. 21, Trends in Cognitive Sciences. Trends Cogn Sci; 2017 [cited 2022 Jan 13]. p. 425–33. Available from: https://pubmed.ncbi.nlm.nih.gov/28476348/

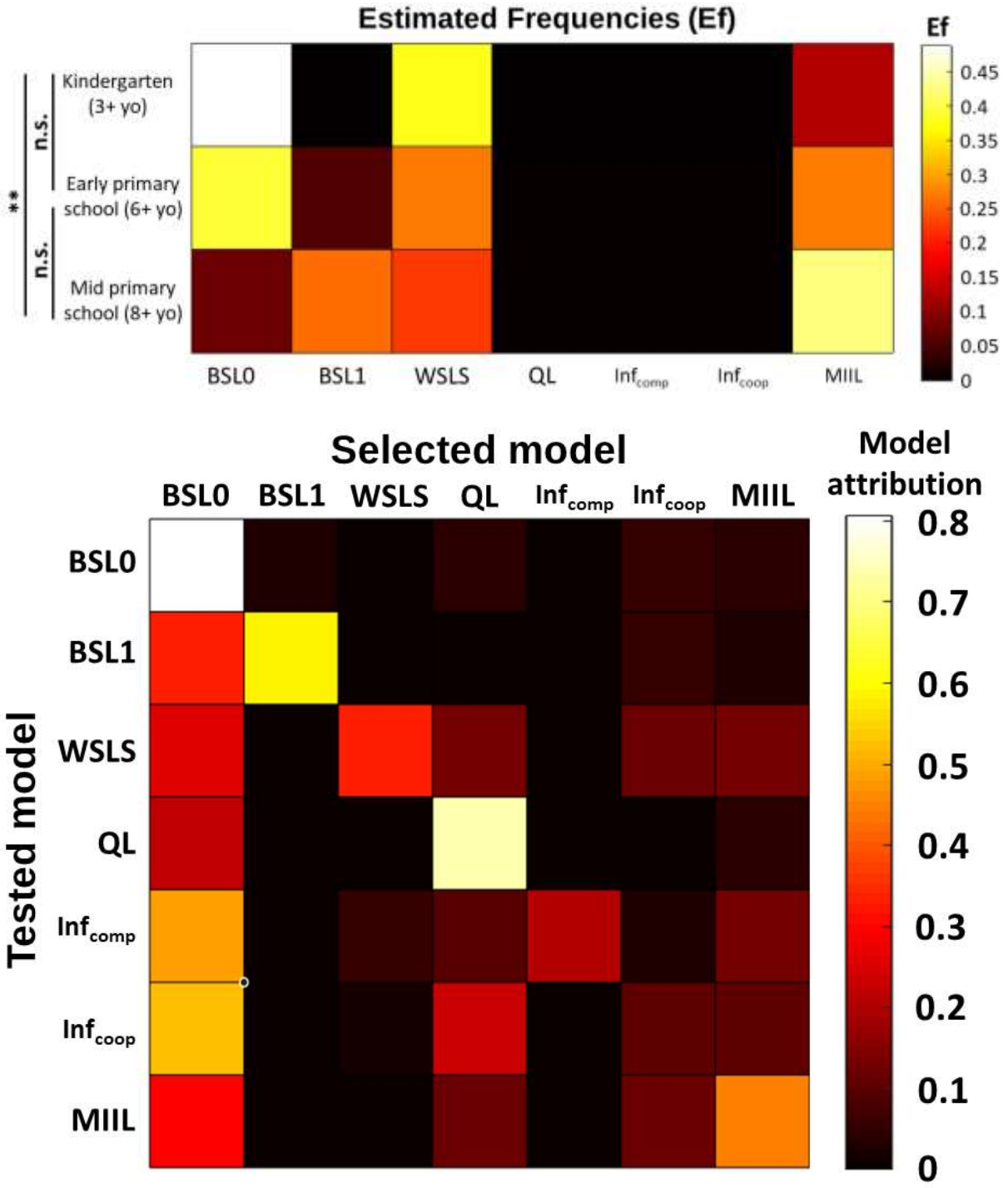
9. Philippe R, Janet R, Khalvati K, Rajesh P.N. Rao, Lee D, Dreher JC. Neurocomputational mechanisms engaged in detecting cooperative and competitive intentions of others. Nat Commun [Internet]. [cited 2022 May 13];(in press):in press. Available from: https://doi.org/10.21203/rs.3.rs-1160167/v1



Supp. Fig. 4. Scheme of the Mixed-Intention Influence model (adapted from Philippe et al(9)**).** Two influence models (one cooperative and the other competitive, resp. P_{coop} in green and P_{comp} in red) compute a value for choosing one specific target (P in black). A controller uses the difference between the absolute value of the value of each expert (called reliability) to compute a probability that the other is cooperating (blue). Then, the model weights the value of each expert according to the probability of being in cooperative or in competitive modes to produce a final decision value. Then it compares its predictions to the actual reward and computes again a new value for each expert.



Supp. Fig. 5. Model checks. (a) Confusion matrix. x-axis: model selected as the best fitting one after model fitting and Bayesian Model Selection (BMS); y-axis: initial simulated model that generated the data; Color-scale: model attribution ratio. **(b) Simulation-recovery analysis.** X-axis: value of the simulated parameters of the influence model that generated the simulated data; Y-axis: value of the estimated parameter after model fitting and parameter estimation on the simulated data. We simulated then recovered the two parameters constituting the arbitration weight that estimates the cooperation tendency of the subject in the MIIL model. BSL0&1: Bayesian Sequence Learning of depth 0 and 1. WSLS: Win-Stay/Lose-Switch. QL: Q-Learning.



Supp. Fig. 6. Analysis with inclusion of cooperative and competitive Influence models (top) Bayesian model comparison. X-axis: model names; Y-axis: age group; Color-scale: Estimated frequencies (Ef), i.e., the frequency at which one model fitted one subject best in the group population compared to the other tested models. In the Kindergarten (KG) and Early Primary School (EPS) groups, the model that best fitted most of our group of participants is the BSL0. In Mid-Primary School, the MIIL model best fitted most of the participants. Difference between the group profiles of children in KG and MPS is significant (P=0.0067, uncorrected). **(down) Confusion matrix.** X-axis: model selected as the best fitting one after model fitting and Bayesian Model Selection (BMS); Y-axis: initial simulated model that generated the data; Color-scale: model attribution ratio. BSL0&1: Bayesian Sequence Learning of depth 0 and 1. WSLS: Win-Stay/Lose-Switch. QL: Q-Learning. Inf_{coop}: Cooperative Influence model. Inf_{comp} : Competitive Influence model

A_WA_W A: 1 ; B: 0	**A_WA_L** A: 3 ; B: 1	**A_LA_W** A: 1 ; B: 0	**A_LA_L** A: 1 ; B: 2
B_WB_W A: 0 ; B: 0	**B_WB_L** A: 0 ; B: 0	**B_LB_W** A: 0 ; B: 2	**B_LB_L** A: 1 ; B: 0
B_WA_W A: 0 ; B: 0	**B_WA_L** A: 10 ; B: 2	**B_LA_W** A: 0 ; B: 0	**B_LA_L** A: 0 ; B: 0
A_WB_W A: 5 ; B: 2	**A_WB_L** A: 3 ; B: 0	**A_LB_W** A: 0 ; B: 0	**A_LB_L** A: 2 ; B: 3

Supp. Tab. 7: Example of a memory table of the history of the sequences of events considered by the algorithm. The algorithm (Artificial Agent) keeps a history of the choices of the player after a given sequence of events that occurred during a session. In each cell, the sequence of events is **in bold**. An example of a history of the player's following choice is given below the sequence of events. "A: 1 ; B: 0" means "After the sequence of events in bold, the player chose card A once and never card B until now". The algorithm updates this table at each trial and chooses its card based on it. **W = Win ; L = Loss ; A = Card A chosen ; B = Card B chosen.**

A_WA_W A: 1 ; B: 0	**A_WA_L** A: 4 ; B: 1	**A_LA_W** A: 1 ; B: 0	**A_LA_L** A: 1 ; B: 2
B_WB_W A: 0 ; B: 0	**B_WB_L** A: 0 ; B: 0	**B_LB_W** A: 0 ; B: 2	**B_LB_L** A: 1 ; B: 0
B_WA_W A: 0 ; B: 0	**B_WA_L** A: 10 ; B: 2	**B_LA_W** A: 0 ; B: 0	**B_LA_L** A: 0 ; B: 0
A_WB_W A: 5 ; B: 2	**A_WB_L** A: 3 ; B: 0	**A_LB_W** A: 0 ; B: 0	**A_LB_L** A: 2 ; B: 3

Supp. Tab. 8: First updated memory table of the history of the sequences of events. The algorithm (Artificial Agent) keeps a history of the choices of the player after a given sequence of events that occurred during a session. In each cell, the sequence of events is **in bold**. If we consider the sequence of events "…$A_WA_LA_W$…", this is the updated table from **Supp. Tab. 7**. **W = Win ; L = Loss ; A = Card A chosen ; B = Card B chosen.**

A_WA_W A: 1 ; B: 0	**A_WA_L** A: 4 ; B: 1	**A_LA_W** A: 1 ; B: 1	**A_LA_L** A: 1 ; B: 2
B_WB_W A: 0 ; B: 0	**B_WB_L** A: 0 ; B: 0	**B_LB_W** A: 0 ; B: 2	**B_LB_L** A: 2 ; B: 0
B_WA_W A: 0 ; B: 0	**B_WA_L** A: 10 ; B: 2	**B_LA_W** A: 0 ; B: 1	**B_LA_L** A: 0 ; B: 0
A_WB_W A: 5 ; B: 2	**A_WB_L** A: 3 ; B: 1	**A_LB_W** A: 0 ; B: 0	**A_LB_L** A: 2 ; B: 3

Supp. Tab. 9: Updated memory table of the history of the sequences of events. The algorithm (Artificial Agent) keeps a history of the choices of the player after a given sequence of events that occurred during a session. In each cell, the sequence of events is **in bold**. If we consider the sequence of events "…$A_WA_LA_WB_LB_LA_WB_W$…", this is the updated table from **Supp. Tab. 7**. **W = Win ; L = Loss ; A = Card A chosen ; B = Card B chosen.**

Models	GLM		GLME	
Variables	$Reward_n$: Estimate (Std. Error)	$Stay_n$: Estimate (Std. Error)	$Reward_n$: Estimate (Std. Error)	$Stay_n$: Estimate (Std. Error)
Intercept	-0.79419 (0.11351) ***	-1.01799 (0.12891) ***	-0.71931 (0.12678) ***	-0.78958 (0. 17080) ***
$Reward_{n-1}$	0.13777 (0.03067) ***	0.42592 (0.03092) ***	0.11649 (0.03120) ***	0.42246 (0.02944) ***
$Reward_{n-2}$	0.22214 (0.03086) ***	/	0.20311 (0.03132) ***	/
$Reward_{n-3}$	0.09386 (0.03117) **	0.08087 (0.03148) *	0.07641 (0.03158) *	0.08593 (0.03164) **
$Stay_{n-1}$	0.31885 (0.03078) ***	0.12632 (0.03096) ***	0.30152 (0.03122) ***	/
$Stay_{n-2}$	/	0.18913 (0.03966) ***	/	0.09100 (0.04080) *
$Stay_{n-3}$	0.08633 (0.03907) *	0.12491 (0.03939) **	0.07109 (0.03943) .	/
AA_Stay_{n-1}	0.11296 (0.03089) ***	0.19588 (0.03974) ***	0.09678 (0.03128) **	0.17040 (0.04047) ***
AA_Stay_{n-2}	0.07616 (0.03940) .	0.17140 (0.03960) ***	0.06986 (0.03958) .	0.16351 (0.03174) ***
AA_Stay_{n-3}	0.10179 (0.02862) ***	0.25365 (0.02896) ***	0.07905 (0.02918) **	0.16725 (0.02984) ***
$Coop\_sig_{n-1}$	-0.19871 (0.13376)	0.20653 (0.05734) ***	-0.21677 (0.13427)	0.22357 (0.05851) ***
$Coop\_sig_{n-2}$	0.13516 (0.03963) ***	-0.11472 (0.18523)	0.13944 (0.03975) ***	-0.12383 (0.18964)
Trial	0.04219 (0.01474) **	0.07281 (0.01528) ***	0.04419 (0.01506) **	0.08420 (0.02062) ***
Age	0.02797 (0.01644) .	0.02112 (0.01835)	0.02725 (0.01838)	0.01987 (0.02493)
AA_mode	/	-0.20953 (0.09820) *	/	-0.24844 (0.10025) *
Sex (Male)	/	-0.04353 (0.02845)	/	/

Hierar_cond	/	-0.03278 (0.02158)	/	/
DT	/	/	/	0.02368 (0.01623)
$Coop\_sig_{n-1}$:Age	0.04749 (0.01978) *	/	0.05122 (0.01987) **	/
$Coop\_sig_{n-2}$:Age	/	0.04109 (0.02781)	/	0.04668 (0.02852)
AA_mode:Age	/	0.03259 (0.01429) *	/	0.03831 (0.01461) **

Supp. Tab. 10: Comparison of GLMs and GLMEs for behavioral factors. Hierar_cond, AA_mode, Coop_sig, Reward, AA_Stay and Stay are coded effects (Hierar_cond: different = 1, equal = 0 ; AA_mode: competitive = -1, random = 0, cooperative = 1; Coop_sig, Reward, AA_Stay & Stay: True = 1, False = 0). Trial & DT were normalized. Signif. codes: '***': $p < 0.001$; '**': $p < 0.01$; '*': $p < 0.05$.

CHAPTER 3: General Discussion

This dissertation aimed to investigate ToM abilities across different species of primates with a computational modelling approach. To this end, I explored 3 projects that systematically applied the same computational approach to study diverse ToM components in various populations and contexts. In this General Discussion, I will first summarize the principal results of our 3 research projects. I will then take some perspective to analyze from a higher standpoint the strengths, limitations and prospects for each project. I will finally propose an interpretation of our findings and conclude with recommendations on future research directions.

A. Summary

1) Project 1: Computational modelling of coordination in a transparent Bach-or-Stravinsky Game: humans' vs macaques' observation learning process

This project focused on the emergence of cooperation in iterated decision games. We had many questions in mind regarding the data analysis. First, the cross-species nature of the experiment raises questions about the behavioral differences between humans and rhesus macaques. Second, with our computational approach in mind, we asked which computational mechanisms could help explain data generation. Third, we wanted to know how information would shape the decision process and influence the emergence of a consensus (or not). More specifically, we wanted to establish how transparency would affect participants' tendency to coordinate with each other.

To do so, Möller & Kagan exploited the possibilities offered by newly acquired technology. A large transparent tactile screen on which information can also be displayed is connected to a computer. Coupled with hand captors that were used to measure decision times, this very promising experimental setup enabled them to implement simultaneous iterative games and test them under transparency. This led to the concept of transparent 2x2 (two players, two options) games in which players can observe each other's actions [1,2]. In contrast to classic simultaneous or sequential games, in this game-theoretic setting, the access to the information about the current choice of the other agent is probabilistic. Indeed, in each round, there are three possible cases:

1. Player 1 knows the choice of Player 2 before making its own choice.
2. Player 2 knows the choice of Player 1 before making its own choice.
3. Neither player knows the choice of the partner.

In this particular context, an interesting game to study is the Battle of the Sexes, or Bach-or-Stravinsky game (BoS) [3]. In this game, each agent has an individually "preferred" option, in the sense that they are more rewarded if they choose their individual "preferred" option. However, coordination on the same option by both agents is also rewarded with a bonus reward. The bonus is distributed equally to both players so that any coordinated choice is rationally better than no coordination for both agents. This creates an inherent conflict about who profits the most, which offers interesting opportunities to study social interactions in primates, including their decision learning process. In the context of strategic dyadic

decision games, different computational mechanisms may indeed account for the emergence of coordination in primates.

We therefore analyzed the behavioral data of adult humans and rhesus macaques that played the transparent, iterative BoS game. We applied a computational modelling approach to study the learning process of the two groups of Primates under different conditions. We showed that the difference in choice behaviors between adult humans and rhesus macaques derived from a difference in decision processes. We identified a target-based POMDP decision process in humans that fully exploited the information of the visual cue to coordinate. In contrast, macaques were mainly fitted best by either a target-based influence model that did not exploit the current visible choice, or a side-based POMDP model that exploited the visible choice of the other agent to coordinate on the same side.

The nature of the computational models that were found to fit macaques' behavior better can be explained to a certain extent by both current literature and previous behavioral observations. The target-based influence model is held in common with other species of monkey, that were studied when playing similar strategic decision games in a non-transparent setup [4]. However, the transparent nature of our setting allowed another type of strategy, that led to high rewards with low efforts, the side-based POMDP. Coordinating based on the side of the other's decision, rather than on the nature of the cue, requires less effort and attention while also leading to high rewards.

Humans were found to use more sophisticated, ToM-based, Bayesian decision processes in non-transparent setups [5]. In our transparent setup, the visual access to the other's current action hijacks the usefulness of using such sophisticated computational models. Therefore, the use of a target-based POMDP model makes more sense as it spares the computational complexity of high-order mentalizing to focus on the exploitation of the current visual cue to coordinate.

2) Project 2: Mentalizing mechanisms underly strategic coordination in Guinea baboons (Papio papio)

Most studies that report evidence of ToM in non-human primates (NHPs) use false-belief tests, based on anticipatory looking of another agent's action [6–9]. However, such studies suffer from three main limitations. First, it has been difficult to prove that tests of ToM can only be explained by mentalistic accounts rather than simpler non-mentalistic accounts (construct validity) [10–18]. Second, false-belief tests can only provide a limited insight into ToM. They are perhaps sufficient to identify ToM in the case of success only, but are however uninformative in case of failure [11,19,20]. Furthermore, such tests are unable to measure the level of sophistication of social reasoning [11,21]. Third, yet another limit of this type of experiment is that the participants are passively looking at scenes of interacting individuals [6–9], rather than actively engaging in the social interaction [4,22]. In addition, many ToM studies in NHPs resort to direct interactions with humans [4], thereby limiting their motivational and ecological validity [12,13].

In our second project, we developed a very different approach. Pairs of Guinea baboons, who were part of a colony of 18 individuals, had free access to automated devices

to play together in a strategic coordination game, matching pairs, on computer touch screens. Strategic coordination is indeed thought to constitute a key component of ToM that allows animals to coordinate group behavior in the wild [23–28]. Therefore, we combined this experiment with formal computational models of this economic game to investigate whether the model that explains coordination behavior best actually includes ToM mechanisms.

This approach has several advantages over previous paradigms. Our computational approach provides a mechanistic understanding of observed behavior that can specify mentalistic and non-mentalistic processes, and may also differentiate different levels of sophistication. Moreover, our experimental set-up directly addresses motivational and ecological concerns by actively engaging baboons who freely and willingly interact with a conspecific, without any direct intervention by the experimenters. This allows us to maximize the ecological validity of our study without losing experimental control.

Our findings provide evidence that baboons mentalize their influence on the behavior of other baboons. Baboons were allowed to learn to play a simple game (matching pairs) in which any two baboons, playing in adjacent work devices, learned to coordinate by choosing the same icon as each other from two presented on the screens. Having first learned to coordinate with their conspecifics, for the experimental task, the baboons were required to play the same game (and coordinate) without being able to see their partners' choices.

In the control condition, the baboons played the same coordination game without a partner. In fact, in both the control and test conditions, the baboon played against the same computer algorithm that decided which choices would be rewarded. This control condition allowed us to attribute differences in behavior between the social *versus* non-social condition strictly to the baboons' beliefs concerning the social aspect of the task, and not to the strategy of the other baboon, present in the adjacent box during the social condition.

We modelled their choice behavior with a variety of computational models including a simple heuristic guessing strategy, reinforcement learning models with and without theory of mind mechanisms, and Bayesian models of different depth with and without theory of mind mechanisms.

Our computational results demonstrate that, when baboons were led to believe they played with a partner, the best model of their behavior was a reinforcement learning model that included a Theory of Mind mechanism (the Influence model). However, when they played alone, they used a win-stay/lose-switch heuristic. Furthermore, because we observed that relative social dominance played an important role in influencing decision making, we developed a new reinforcement learning model, the RL-ELO-Influence model, that takes relative social rank into account.

Our findings show that baboons take the influence of their previous choices on their partners' behavior as well as their own relative social dominance into account when they predict the best choice for coordinating with that partner. We show that a ToM-based strategy is the best explanation from among many possible options to explain strategic coordination behavior in Guinea baboons. This approach was only possible because of the unique experimental set up that allowed us to study the behavior of the tested baboons with many other conspecifics and limiting any confounds that might result from humans interacting with them or teaching them how to play.

At a mechanistic level, this project characterizes some computational mechanisms that give rise to coordination learning, which constitutes a major advance in our understanding of the computational mechanisms engaged in ToM. At the psychological level, our results indicate that Guinea baboons resort to a basic ToM process to learn to coordinate with their conspecifics. Thus, ToM processes used for coordination learning may have evolved in the common ancestor of the Old-World monkey and *Hominoidea*, that diverged in evolution 30 million years ago.

3) Project 3: Computational Modelling of the ability to read changes of cooperative vs competitive intentions: developmental changes in children

The true intentions of others are fundamentally concealed and must be indirectly inferred from their observable behaviors. Nevertheless, humans often interact without knowing the intentions of others in scenarios that may involve either competition or cooperation. This makes social situations difficult to predict, especially when the others' intentions fluctuate over time between goals that coincide or compete with our own. Such situations require a relatively advanced capacity to identify and adapt to the intentions of others.

Advanced Theory of Mind (AToM) has been developed to describe multiple aspects of an advanced understanding of the minds of others [29–32]. AToM factors include social reasoning, reasoning about ambiguities, and recognizing transgressions of social norms [30,31]. Previous studies have shown that children begin to understand more complex forms of mental-state reasoning around 6 years old [30,33–35]. Children are thought to reach a milestone at 7 years old when they attain the conceptual insight that mental states can be recursive [30]. For example, the realization that someone can hold a false belief about a belief of someone else. Thus, AToM has been proposed to rely on recursive reasoning [36], i.e., the ability to reason about second- and higher-order beliefs [37,38], and on the ability to interpret social actions [35]. However, other important cognitive mechanisms that underlie distinct aspects of AToM may emerge at different times during childhood. In particular, adapting to fluctuating cooperative versus competitive intentions of others may be an important cognitive component required for social reasoning, a core component of AToM.

Recently, Philippe et al. (in press) built a computational account of how human adults discriminate the intentions of others when playing an economic game in which the partner, an artificial agent, switched between competitive and cooperative strategies without signaling this to the participants. They found that adults not only make inferences about the likely decision of others and update beliefs about the current level of cooperativity of others, but also arbitrate between the likely intentions of others based on their respective reliability. This so-called "mixed-intention model", that arbitrates between competitive and cooperative experts, outperforms other learning models in predicting adults' choice behavior. They also found that specific brain regions tracked the reliability of this arbitration process.

This sophisticated neurocomputational mechanism may be absent or only partially developed in younger children. Therefore, in the third project, we have sought to identify its appearance during development to discover what computational mechanisms are used by children before they develop a deeper theory of mind, and at what age the MIIL develops during childhood. To identify milestones in the acquisition of AToM-based strategy during normal human development we used the same economic game to explore the decision

behavior of children in three different age groups (3-5, 6-7 and 8-9 years old). To this end, children completed a task in which they played a card-matching game on computer tablets. They played in pairs, facing a same-gender child. Both were told that they were playing with each other, but they were not allowed to communicate. Unbeknownst to them, they were actually playing with an Artificial Agent (AA) that switched between cooperative and competitive trial blocks (two modes). The goal of the participating children was to score as many points as possible in the game by guessing the correct color. The more points they scored the more reward they would obtain at the end of the experiment.

We compared a large battery of computational models at each age to identify which models fitted each child best in each age group. Children in the youngest age group favored a Bayesian decision process to track the most probable strategy of the other player, taking into account the other player's (i.e., the artificial agent's) behavior in the previous trial. At 8 years old, the children began to conceptualize not only the effect of their own choices on their opponent's, but also the importance of tracking the other's intentions in volatile situations. The model accounting best for behavior around this age was, similar to that found by Philippe et al., in adults, i.e., a controller that arbitrates between competitive and cooperative expert influence models according to the probability of being in a competitive of cooperative interaction. Thus, from this age children began to play according to the same Advanced Theory of Mind strategy as adults.

Our results provide a mechanistic understanding of key processes required in the development of (advanced) Theory of Mind and characterize the development of the computations required to adapt to the fluctuating cooperative or competitive intentions of others during childhood. We believe these results will have broad appeal to those interested in child psychology and computational neuroscience. Our study combines a computational modelling approach with behavioral data acquired in different age-range in children to identify Theory of Mind mechanisms during human development. Hence, this study might represent a milestone in the elucidation of the ontogeny of theory of mind mechanisms in humankind.

B. Perspectives

In this part, I want to provide more context to the reader about the projects that were presented in this dissertation. For each project, I will describe its origins, the main difficulties that I faced, and how I overcame them. The next step will lead us to analyze the strengths and weaknesses of each project. Finally, I will take a step back to present the bigger picture, and propose future research directions for each of them.

1) Project 1: Tribulations, The Origins

a- Retrospective

Everything started with this first project. It initially consisted in applying computational modelling to some data that were shared by Sebastian Möller and Igor Kagan. These data had already been collected by them in Germany, and they were already in the process of writing and reviewing a research article with these data [39]. They designed and implemented the experiment themselves, and, furthermore, eventually published their research article [1]. My

role at the beginning, was only to implement and test algorithmic computational models of the data of humans and naïve macaques. What was supposed to be a simple Master project at the beginning, turned out to be a whole PhD project, the origin of my PhD thesis.

In collaboration with their team we decided to collect and analyze additional data in more conditions to gain more statistical power and explore certain ideas more profoundly. They therefore collected supplementary data and shared them with us, so as to allow us to conduct more data analyses. This led me to apply more diverse methods to analyze the data, such as Exploratory Data Analysis (EDA, see Chapter 1, section C.4) b- ii.), correlational computational modelling (see Chapter 1, section C.4) b- ii.), and, of course, algorithmic computational modelling (see Chapter 1, section C.4) b- iii. & C.4) b- iv.).

Multiple reasons motivated us to pursue these data analyses further. Besides the obvious, requirement for me to obtain data on which to polish my modelling skills, this experiment had some very interesting features that I should like to highlight here. First, the experimental setup is novel, original, and promising. The transparent, tactile screen allows many possibilities. Secondly, the transparent game theory framework used to study the emergence of a compromise was very appealing to me. Third, the data are of very high quality and very rich. Many aspects of the behavior were captured by this setup, with various conditions (naïve, in training, and trained) and species (macaques and humans).

The quantity of data is also not to be underestimated. However, the data were tricky to analyze, especially as a greenhorn modeler. They had to be re-analyzed multiple times, for various reasons (apart from programming errors…). The most interesting causes were, first, that, in the first attempts to "blindly" analyze the data (i.e., without any feedback or suggestions from the other team, to avoid any bias), I initially ignored the importance of considering the side of the choice of the participant, rather than only the ID of the chosen target. This was then suggested by Kagan & Möller, who had noticed in their data and in the behavior of their macaques, a choice bias towards a side. This highlights the importance of exchanging with the experimenters during the whole modelling process, to incorporate their feedback, observations, and intuitions. The data therefore had to be re-analyzed and re-modelled from a "side" based standpoint, too.

Another reason we had to restart the data analysis from (almost) scratch, is the model confusion analysis. This step is central for any robust computational modelling approach, and at that time, was unknown to me (it was also not standard for the field at the time). Very few computational modelling research articles actually presented clear sanity checks, which we eventually corrected in this project, and all those that followed, too. This moment was also the occasion for us to also complement our algorithmic computational models with EDA and correlational modelling (GLMEs). This project provided a rich environment and many occasions to step up my technical skills.

b- Limitations and open questions

Just like any research experiment, this one also suffers from weaknesses and limitations. One of them is the low number of monkey participants, as often the case in primate laboratory research. There also is the unequal number of sessions and trials between different macaques, with some of them playing with multiple different partners, while others only played with one partner. There also is a lack of controls for the behavior of the partner, which makes it difficult

to determine whether the behavior is driven purely by the feedback, or unrelated social cues from the partner, e.g. social dominance, social bonds, etc. Furthermore, as the attentive reader might have already noticed, the initial goal of this experiment was not to study ToM per se, but rather to study the emergence and the learning process behind coordination. It can however be linked to observational learning, which can arguably be considered as a precursor of ToM abilities [40]. This is also why I chose to include this project in my PhD thesis. Interesting open questions would be to compare, for the same subjects, transparent and opaque settings. Decorrelating the positions of targets for coordination between the 2 subjects would also be necessary to ensure that participants do not automatically coordinate by simply coordinating on the same side.

Finally, the experimental protocol makes it difficult to determine the depth of understanding of the task by the monkeys. More specifically, it is difficult to distinguish what drives their coordination intent. Indeed, the correlation between the sides of the targets of the two players on the transparent screen prevents the possibility to differentiate between basic behaviors and sophisticated decision processes and therefore any deeper interpretations of the arising of a strategy based on the side of the target. Nevertheless, this project is currently still in preparation. It was held on pause for a certain amount of time, due to a prioritization of the two following projects. Indeed, the limitations that we faced in this project made us realize the experimental constraints that we needed to care of in the design of future projects. This helped us to know what conditions to look for in the design and setup of future experimental protocols.

2) Project 2: An Unexpected Panacea

a- Retrospective

The idea of this project sprouted at the end of 2019, based on some of the previous realizations that were still evolving. The need for a similar experimental setup, but with more programmability and controls of the task, led us to contact a team in Aix-Marseille that were acquaintances of my supervisor: Claidière & Fagot's lab. Their unique experimental setup, with a colony of willing baboon participants playing video games on connected tactile tablets with potentially versatile (social) environmental conditions conquered our hearts (and brains). We therefore designed, with the help of Claidière et al., the second project presented in this dissertation.

In this project, I participated in the elaboration of the experimental protocol, analyzed the data, and wrote the draft of the article. The design of the experimental protocol and its implementation were completed at the beginning of 2020, just before the first French lockdown. Luckily for us, data collection could be automated and performed remotely, from March to May 2020. So, data collection could be performed during the lockdown. This project also turned out to be very enriching in terms of scientific approach and technical skills. An unexpected panacea.

The main difficulties that I faced were with the implementation and the programming of the data analysis to reach robust results. Under limited guidance, this demanded a lot of reading and learning by myself, as did the writing of the article. Writing a research article accurately and concisely is very difficult. The phrasing of some sentences needs to be polished

multiple times. Also, after I finished writing the article in a certain format, it was eventually decided to reformat for another journal.

b- Limitations and open questions

This promising experiment, as we have insisted a lot already, has many strengths. The novelty of our experimental approach, the rigor with which data were analyzed, as well as the importance of our findings, which provide powerful new evidence that Theory of Mind might not be unique to humans. It however has weaknesses that we have to concede. We for example suffer from a low number of subjects, which limits the statistical power of our study. This is a common downside in the field of primate research. We could have extended the training phase of the baboons longer, perhaps to obtain more participants that passed the coordination criterion. However, we had to find a compromise between number of included subjects and potential overlearning.

Also, just like with many computational approaches, we only tested a finite set of computational models. This means that our model selection only holds for this set of models, and that we did not find the best computational model ever (this would be difficult to prove anyways…). There definitely is a better model that could explain our data. However, what we argue is that the better model has to include some common computational mechanisms with our best computational model. Firstly, it must account of the effect of one's actions on the other's outcome. Secondly, the decision process must be modulated by the relative difference in dominance. Should some alternative, simpler explanations than ToM exist, they would be directly testable against the proposed computational mechanisms in this manuscript. We also want to underline that the ToM mechanism that is captured by the Influence Learning model might not be at the level of, or as deep as, what psychologists and ethologists usually call (full-blown) ToM.

Furthermore, we specifically designed the algorithm behind the other's decisions to adaptively encourage stability and predictability in the subjects' decisions (see Algorithm Section in Supp.Mat.&Meth.). For a given specific trial, the algorithm chose the rewarded target stochastically based on the statistics derived from the history of the two previous choices and outcomes of the subject. While this allowed us greater control over participants' decision cues without suppressing the contingency of the rewards to the subjects' decisions, we cannot exclude the possibility that a different algorithm could have led to different behavioral results and that our findings that more dominant individuals tend to impose their choices more by maintaining that choice, were possibly induced by the cooperative nature of the rewarding algorithm. However, we argue that the existence of within-subject differences in behaviors, depending on the relative rank of the baboons in the dyad, would still hold true even with different rewarding algorithms. This result indeed did not depend on the nature of the rewarded strategy.

Also, we cannot guarantee that the tested baboons really understood the rules of the Coordination Game, nor that they actually believed that they were interacting with the baboon in the adjacent box during the whole testing process. Even if the training phase was designed to maximize this context, these are the limitations of working with animals with which one cannot communicate. There is also risk that the subjects failed due to the cognitive load already required by the task and learning about the contingencies in the games used.

However, on a side note, even if the tested baboons did not fully understand the rules of the game, the within-subject change in behavior that depended on their relative dominance in each dyad would still indicate a clear effect of social dominance on social decision learning. If they did not believe that they were interacting with the baboon in the adjacent box, this would mean that the mere presence of another baboon, which can see and is visible to the tested baboon influenced their decision learning process.

In order to maximize the ecological validity of our study without losing too much experimental control [41], we tested freely-ranging baboons, that willingly entered the experimental workstations, with (or without, depending on conditions) another conspecific to participate in this study. We therefore had no control over which conspecific the tested subjects partnered, and could not test unwilling individuals. Our group of tested baboons is thus of limited number. Future studies might investigate larger groups of individuals, and also other social species to check to what extend such results might generalize. We also acknowledge that the naturalistic validity of our setup is not perfect. Some might question the ecological validity of experiments tested through touchscreens in ADLM boxes. Earlier studies have shown that this system is an efficient tool for the assessment of cognitive functions in experimental tasks e.g., memory [42], reasoning [43] or perception [44], social interactions in networks [45] and has a positive impact on animal welfare [46]. We therefore believe that our results would also generalize to coordinating behaviors in more natural setups.

Some interesting future directions that come to mind would be to combine eye-tracking and camera recordings with the current setup. Access to gaze-direction and fixation-time durations would provide better insight to model the mind of the baboons. It would also enlarge the range of experimental protocols to study ToM. Other video recordings would also allow to better study how the relationship between baboons that just played next to each other evolves, depending on how their playing session went. We also encourage the testing of different species of monkeys or great apes with similar experimental setups. This would provide a cross-species viewpoint which would be more than welcome in ToM research.

Finally, if the simplicity of the coordination game that was used is an advantage when it comes to interpreting behavioral results, it might not be sufficiently sophisticated to unleash the full potential of our participants' meta-representational abilities. Indeed, such a simple game might not require enough depth of reasoning and ToM for the participants to display the full extent of their ToM ability. This is why we recommend the testing of more sophisticated dyadic social games, for example that used by Philippe et al., (in press) and in our third project.

3) Project 3: Improvise, Adapt, Overcome

a- Retrospective

The last project that we presented in this dissertation was actually co-designed with the 2nd one. Indeed, the 2nd experiment on baboons was initially supposed to constitute a cross-species comparison with human adults and children. However, we decided, in conjunction with Claidière's team, to first test the baboons with a simpler task, and to check whether the baboons could at least pass the coordination criterion and produce meaningful

results. Now that everything is almost finished and that we have more experience, conducting the experiment with baboons as initially intended will probably occur at some point in the future.

The whole experiment was mainly inspired by the main PhD project of our colleague, Rémi Philippe, recently published [47] and was an extension of his project to younger populations, to determine at which age of childhood we would observe results similar to those he found in adults. At the beginning, we also wanted to do a cross-cultural study and include French primary schoolers. We had a contact, thanks to Jean-Baptiste Van der Henst, with a primary school in Lyon that was ready to help us conduct the experiment, however, COVID-19 broke out before we were able to fully implement the task on tablet to test the children. When the first lockdown was over, sanitary restrictions prevented us from experimenting in France with children, at their school or at the lab. Our contact also stopped answering our messages. Consequently, we piloted this experiment in China during the period of May-June 2020, because the team that collaborated with us had no lockdown or restrictions (yet).

The final data collection was performed between March and June 2021 in China by collaborating teams, (Yansong Li and Chen Qu). We could not collect any data in France in 2020 or 2021, because of the successive lockdowns and the unwillingness of the schools to risk inviting external experimenters and spreading COVID-19. Also, sanitary restrictions would have prevented the experiment from being run in good conditions. We could have tried again in 2022, but other projects piled up, and we wanted to focus on the analyses of the data collected in China, and those from the other projects.

My involvement in this project consisted of participation in the design of the experiment, the elaboration of instructions (written materials and explanation videos), the task programming, the data analysis, and the redaction of the first draft. The task was programmed in C# with Unity, a software to create cross-platform applications. It was very interesting and challenging to learn to program and implement a game on Unity. A valuable skill that will hopefully prove to be useful in the future. The programmed task was used for a valuation program as a proof-of-concept (PoC). Even if this project's timeline was a bit more chaotic, it made me evolve positively by forcing me to improvise, adapt, and overcome.

b- Limitations and open questions

This study still has some downsides. Just like in the previous studies, the same computational approach was applied. So, it suffers from the same issues, for example, the best model that we found that constitutes the best among those tested, and not in general. Still, the relative consistency with current literature is reassuring. Furthermore, we did not collect any supplementary cognitive measures such as IQ, or performance at school. We did collect socio-economic measures such as wealth status, or the number of siblings, all reported by the parents, however, we did not find any significant effect of these.

A weakness that we have to signal is that we actually do not explicitly ask participants any feedback on their thoughts about the other's intentions. This is mainly because the younger children actually cannot explicitly describe it, and because this experiment also aims to test non-human primates (NHPs), that will not be able to answer such questions in any way. Furthermore, there was a risk of biasing other children that were going to be tested on the experiment by asking too explicit questions.

Another question that we did not address in our current experiment is about the flexibility of learning in children. Indeed, it might be that children become more flexible with age, so that they would need less time to adapt to the different intentions of our AA. In our experiment, we used blocks of 20 cooperative then 10 competitive trials. With longer blocks of cooperative trials, a majority of the younger groups of children might have been able to adapt to the changes of intentions. This would need to be tested with a higher ratio of cooperative (vs competitive) trials of the AA. For example, with blocks of 30 cooperative then 10 competitive trials. We also think that setting up the experiment as, ideally, perfectly neutral would incentivize children more to actively track the intentions of the other player. For example by making them play on the same tablet next to each other.

We also asked a second question during this experiment about the effect of social hierarchy on the decision process of children. Indeed, social hierarchy and dominance might also have an important influence on the interactions between two children playing a strategic decision game. It was shown that children can in fact understand social hierarchies from a very young age, even when they are not directly involved in it. They observe different social cues such as, age, wealth and physical postures, to predict who has the highest status [48–51]. If we define power as the ability to get one's own way and control the actions of others, then children are also sensitive to asymmetrical power relationships. For example, they associate more power to individuals who impose their choices, who are imitated and who set norms [50,52,53]. They tend to prefer higher-status characters in terms of physical dominance, decision power, prestige, and wealth [54] and show a tendency to align their choices with them [55]. Depending on their culture, children can also favor objective information provided by dominant characters more than subordinate ones (e.g., in France [56]), or by subordinate characters more than dominant ones (e.g., in Japan [57]).

We therefore also tested the influence of social hierarchy on the children's decision strategies. To do so, the school grade of the children was used as a proxy measure for social hierarchy: a 3rd year primary schooler was considered to be higher in the social hierarchy than a 1st year primary schooler, but two children in the same grade were considered to be of similar status. Thus, for the "Symmetric" condition children were paired with others of the same grade and in the "Asymmetric" condition the children were paired with another from a different grade. We expected pairs of children who played with partners of different grades to pay more attention to the intentions of their partner, compared to pairs of children who played with a child in the same grade as them. This would translate into a better performance at the decision game for children in asymmetrical pairs than for children in symmetrical pairs. In terms of model, we expected to find the same type of learning model for both conditions, but with a difference in learning parameters between the two conditions. However, we did not observe any effect in our data. This could be due to a negligible effect of social hierarchy in the context of video games, as the children indeed played on a tablet. Another possibility is that the effect might rapidly decrease with time while the children are playing, up to the point that it is not detectable anymore on the whole session. The mixed-effect modelling should have detected such an effect, though. The salience of the difference in hierarchy might alternatively not have been sufficient to observe any effect. Making the children play on the same tablet next to each other might be more efficient to detect this effect.

A pending question would still be to know whether the same brain regions are activated for the older children, that were best approached by the MIIL model, as for adults (Philippe et al., in press). Also, it would be interesting to know how the patterns of activations are distinct/similar between children and adults, and how connectivity might evolve with age up to adulthood. Similarly, the brain activations for younger children that were not best approached by the MIIL model, and the correlations between the activation patterns and the hidden states and parameters of the BSL0 model remain unknown. Testing other populations, for example deaf or autistic children or adults could also be interesting, to evaluate any differences with neurotypical populations. We could imagine, if differences were found, to try using an improved version of our PoC to diagnose, or even better, reduce impairments on the cognitive level. This remains very speculative, so this will definitely need more investigations.

C. Final Conclusion

In this final conclusion, we want to insist on some last few points. We introduced at the beginning of this dissertation the current state-of-the-art in research about Theory-of-Mind in primates, humans and non-humans. We insisted on the challenges faced in the field, and proposed a computational approach that, hopefully, will convince the reader about its relevance and usefulness to study social behavior, and ToM abilities in particular.

If we refer to Marr's framework [58], to understand an information-processing system, we have to consider an embedded hierarchy of levels of explanation. These different levels of explanation explore at different scales the mechanistic processes that underly social or non-social behaviors [59]. On the highest level, we need to understand the goal of the system under study, sometimes also called the "computational" level. The 2nd level, a.k.a. the "algorithmic" level, asks what algorithms are applied behind the computational level and requires to specify the computational model that underlies the system's process. The last "implementational" level asks how the "algorithmic" level is physically encoded and requires to study the brain networks, areas, cells, etc. that lead to the emergence of the 2 first levels. This framework can be applied to study ToM abilities (see Figure 31). Each ToM ability, or subcomponent, that one wants to study can be dissected under Marr's framework to provide better insights on the mentalizing abilities of individuals or populations. However, to do so, the first step is to define the context and to specify the ToM subcomponents under study as clearly as possible.

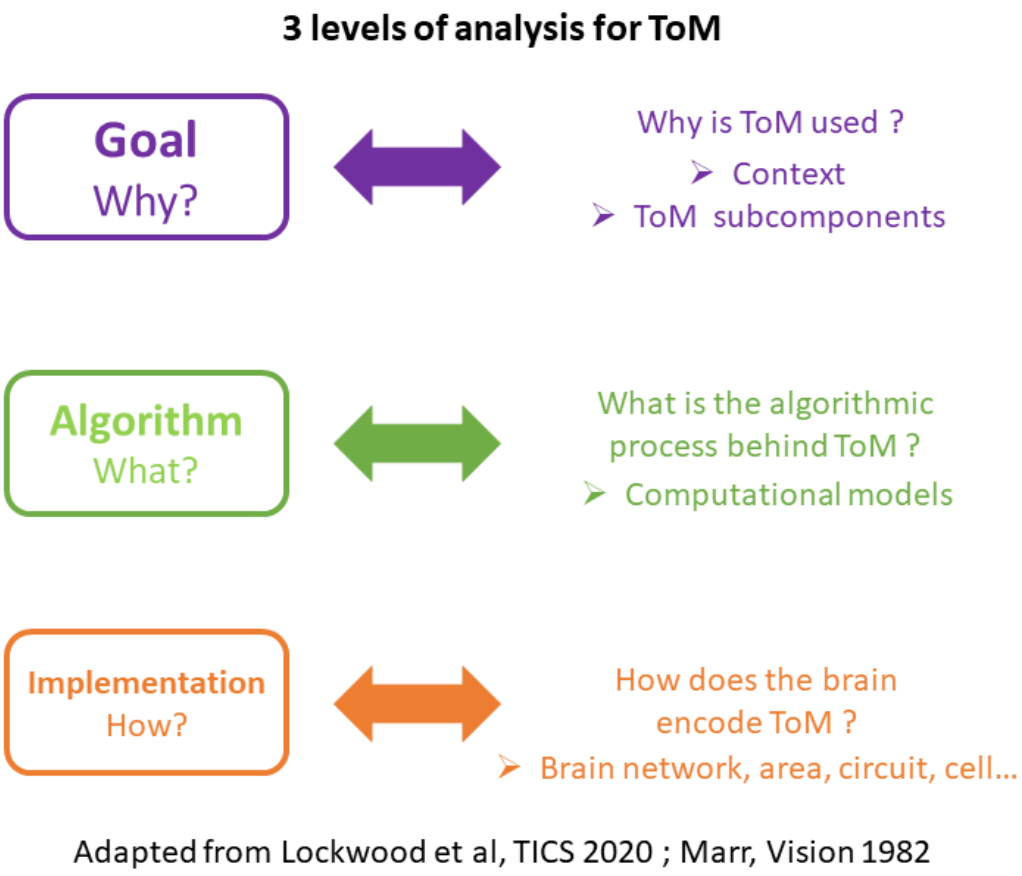


Figure 31: Marr's framework [58] adapted to the study of Theory-of-Mind. To understand how specific Theory-of-Mind is, as compared to other 'non-social' processes such as bird flight, we need to understand the social goal (are we cooperating, learning from, or helping the other person or group). Next, we need to understand the algorithm by which we achieve this. The relatively recent use of computational models such as reinforcement learning, cost–benefit trade-off, meta-Bayesian models, and cognitive maps are some examples of algorithms that could be used. Finally, we need to know how the meta-representational process of ToM is implemented, and in which brain areas, circuit, or cell it is realized. Adapted from Lockwood et al [59].

We focused in this thesis on a specific context in which ToM could be used: strategic coordination games. For this specific context, we established ToM subcomponents that were likely involved in the decision process of our populations, given the best models that fitted their behaviors: representing the influence of one's action on the other's decision and adapting to non-signaled changes of intentions (see Figure 32).

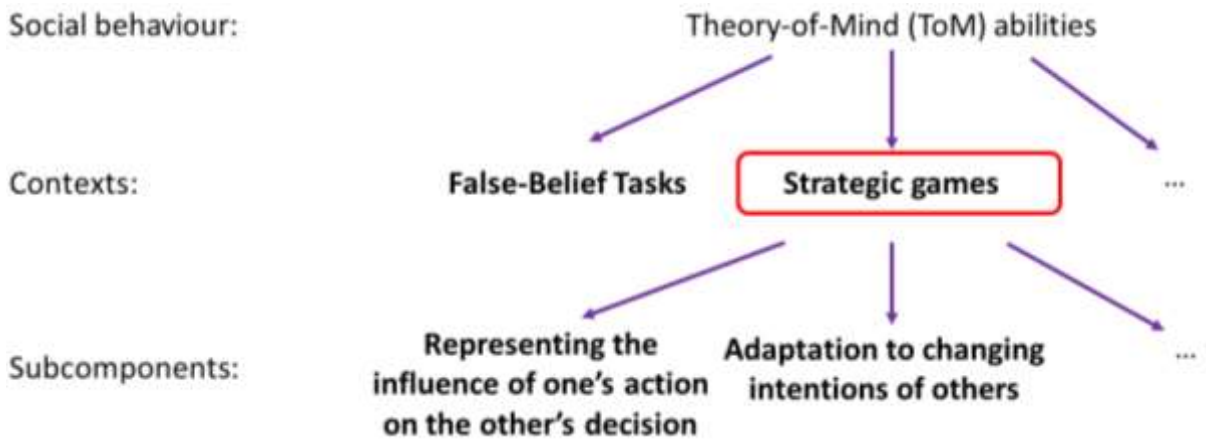


Figure 32 : Context and subcomponents of ToM that are under study in the projects presented in this manuscript.

The projects conducted in this dissertation tested different populations of Old-World monkeys, more precisely baboons (*Papio papio*) in project 2 and rhesus macaques (*Macaca mulatta*) in project 1, and humans of different ages, more precisely children from 3-9 years old in project 3 and adults in project 1 (see Figure 33). Our results indicate the presence of some meta-representational abilities in some species of monkeys (baboons) under relevant conditions, and for the development of an advanced component of Theory-of-Mind in most children starting mid-primary school. A shortcoming of our projects is however that we did not apply the exact same task, or at least an adapted version, across species and development. It would have been interesting to adapt to all of the populations the same task testing a clearly defined subcomponent of ToM.

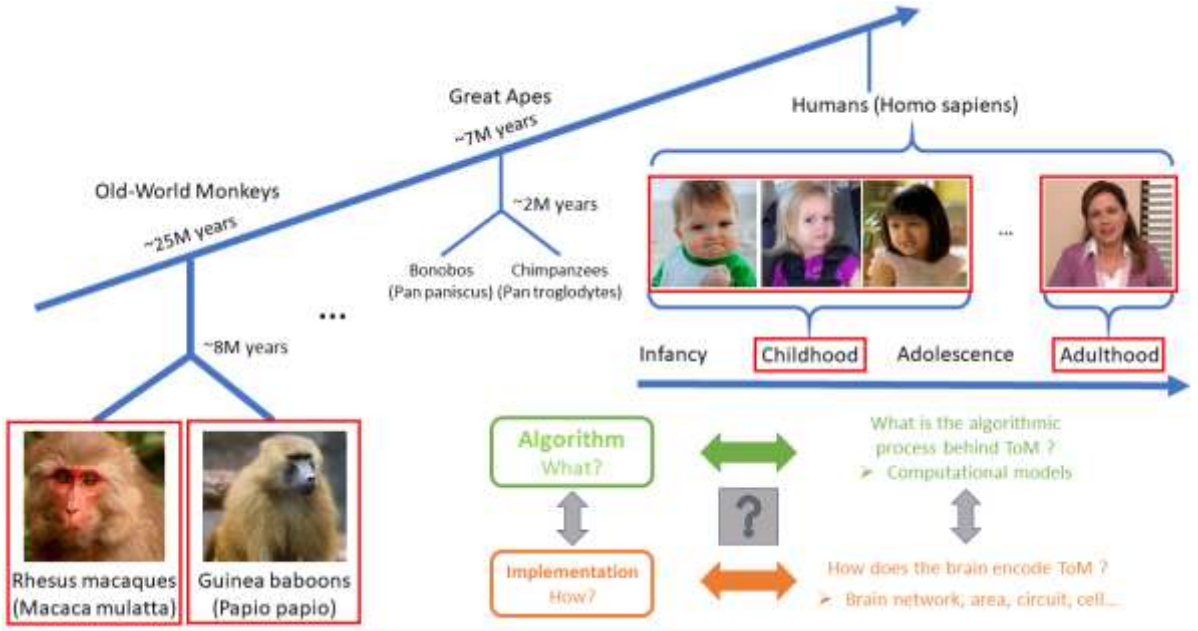


Figure 33: Summary of the populations studied in the projects of this PhD thesis. The ToM processes were studied in this thesis at the algorithmic level of Marr's framework, which still leaves open the questions about the implementational level of ToM processes.

We found that even though humans and monkeys might at first appear to differ a great deal in terms of decision process and observational learning abilities. In certain lab conditions,

a focus on designing more ecological and engaging experiments can reveal very promising results, although this will need further investigations to be confirmed. In particular, baboons seem to be able to represent the influence of their action on the outcome of their partner, and very interestingly, also adapt this representation depending on their relative dominance compared to their partner. As for the development of advanced ToM abilities, such as adaptation to non-signaled fluctuations of intentions, a majority of children appear to acquire a decision process similar to that of adults' at about 8 to 9 years of age. Similar non-invasive approaches might be applied to study ToM in neurodiverse populations and eventually form the basis of diagnostic tests or the evaluation of treatment protocols.

To conclude, the series of studies reported in this dissertation examine various components of ToM abilities across different species of primates and different stages of human development. Through a common framework and methodology, i.e. strategic decision-making games and a computational modelling approach, combined with promising experimental setups, we investigated various aspects of ToM, and could characterize some of its components. This thesis may have important implications for further research investigating ToM abilities in primates, may it be monkeys, great apes, or humans at different stages of their development.

D. References

1. Möller, S. *et al.* Human and macaque pairs employ different coordination strategies in a transparent decision game. *Elife* **12**, (2023).
2. Unakafov, A. M. *et al.* Emergence and suppression of cooperation by action visibility in transparent games. *PLoS Comput Biol* **16**, e1007588 (2020).
3. Kilgour, D. M. & Fraser, N. M. A taxonomy of all ordinal 2 × 2 games. *Theory Decis* **24**, 99–117 (1988).
4. Devaine, M. *et al.* Reading wild minds: A computational assay of Theory of Mind sophistication across seven primate species. *PLoS Comput Biol* **13**, 1–25 (2017).
5. Devaine, M., Hollard, G. & Daunizeau, J. The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? *PLoS Comput Biol* **10**, (2014).
6. Hayashi, T. *et al.* Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex. *Cell Rep* **30**, 4433-4444.e5 (2020).
7. Kano, F., Call, J. & Krupenye, C. Primates Pass Dynamically Social Anticipatory-Looking False-Belief Tests. *Trends in Cognitive Sciences* vol. 24 777–778 Preprint at https://doi.org/10.1016/j.tics.2020.07.003 (2020).
8. Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. Great apes anticipate that other individuals will act according to false beliefs. *Science (1979)* **354**, 110–114 (2016).
9. Kano, F., Krupenye, C., Hirata, S., Tomonaga, M. & Call, J. Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proc Natl Acad Sci U S A* **116**, 20904–20909 (2019).
10. Meunier, H. Do monkeys have a theory of mind? How to answer the question? *Neurosci Biobehav Rev* **82**, 110–123 (2017).

11. van der Vaart, E. & Hemelrijk, C. K. 'Theory of mind' in animals: ways to make progress. *Synthese* **191**, 335–354 (2014).

12. Krupenye, C. & Call, J. Theory of mind in animals : Current and future directions. *WIREs Cognitive Science* **10**, (2019).

13. Lewis, L. S. & Krupenye, C. Theory of Mind in Nonhuman Primates. in *Primate Cognitive Studies* (eds. Schwartz, B. L. & Beran, M. J.) 439–482 (Cambridge University Press, Cambridge, 2022). doi:10.1017/9781108955836.018.

14. Heyes, C. Animal mindreading: what's the problem? *Psychon Bull Rev* **22**, 313–327 (2015).

15. Heyes, C. M. Theory of mind in nonhuman primates. *Behavioral and Brain Sciences* **21**, 101–148 (1998).

16. Call, J. & Tomasello, M. Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn Sci* **12**, 187–192 (2008).

17. Martin, A. & Santos, L. R. What Cognitive Representations Support Primate Theory of Mind? *Trends Cogn Sci* **20**, 375–382 (2016).

18. Horschler, D. J., MacLean, E. L. & Santos, L. R. Do Non-Human Primates Really Represent Others' Beliefs? *Trends in Cognitive Sciences* vol. 24 594–605 Preprint at https://doi.org/10.1016/j.tics.2020.05.009 (2020).

19. Penn, D. C. & Povinelli, D. J. On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. in *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 362 731–744 (2007).

20. Lurz, R. If chimpanzees are mindreaders, could behavioral science tell? Toward a solution of the logical problem. *Philos Psychol* **22**, 305–328 (2009).

21. Rusch, T., Steixner-Kumar, S., Doshi, P., Spezio, M. & Gläscher, J. Theory of mind and decision science: Towards a typology of tasks and computational models. *Neuropsychologia* **146**, 1–44 (2020).

22. Canteloup, C., Piraux, E., Poulin, N. & Meunier, H. Do Tonkean macaques (Macaca tonkeana) perceive what conspecifics do and do not see? *PeerJ* **2016**, e1693 (2016).

23. Boesch, C. Cooperative hunting roles among Taï chimpanzees. *Human Nature* **13**, 27–46 (2002).

24. Bullinger, A. F., Wyman, E., Melis, A. P. & Tomasello, M. Coordination of Chimpanzees (Pan troglodytes) in a Stag Hunt Game. *Int J Primatol* **32**, 1296–1310 (2011).

25. Samuni, L., Preis, A., Deschner, T., Crockford, C. & Wittig, R. M. Reward of labor coordination and hunting success in wild chimpanzees. *Commun Biol* **1**, 1–9 (2018).

26. Surbeck, M. & Hohmann, G. Primate hunting by bonobos at LuiKotale, Salonga National Park. *Current Biology* vol. 18 R906–R907 Preprint at https://doi.org/10.1016/j.cub.2008.08.040 (2008).

27. Strandburg-Peshkin, A., Papageorgiou, D., Crofoot, M. C. & Farine, D. R. Inferring influence and leadership in moving animal groups. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 373 Preprint at https://doi.org/10.1098/rstb.2017.0006 (2018).

28. Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D. & Crofoot, M. C. Shared decision-making drives collective movement in wild baboons. *Science (1979)* **348**, 1358–1361 (2015).

29. Koerber, S. & Osterhaus, C. Some but not all aspects of (advanced) theory of mind predict loneliness. *British Journal of Developmental Psychology* **38**, 144–148 (2020).

30. Osterhaus, C. & Koerber, S. The Development of Advanced Theory of Mind in Middle Childhood: A Longitudinal Study From Age 5 to 10 Years. *Child Dev* **92**, 1872–1888 (2021).

31. Osterhaus, C., Koerber, S. & Sodian, B. Scaling of Advanced Theory-of-Mind Tasks. *Child Dev* **87**, 1971–1991 (2016).

32. Hayward, E. O. & Homer, B. D. Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology* **35**, 454–462 (2017).

33. Miller, S. A. Children's Understanding of Second-Order Mental States. *Psychol Bull* **135**, 749–773 (2009).

34. Osterhaus, C., Koerber, S. & Sodian, B. Scientific thinking in elementary school: Children's social cognition and their epistemological understanding promote experimentation skills. *Dev Psychol* **53**, 450–462 (2017).

35. Osterhaus, C. & Koerber, S. Social cognition during and after kindergarten: The relations between first-order and advanced theories of mind. *European Journal of Developmental Psychology* **18**, 573–592 (2021).

36. Perner, J. Higher-order beliefs and intentions in children's understanding of social interaction. in *Developing Theories of Mind* 271–294 (1988).

37. Miller, S. A. Children's Understanding of Second-Order Mental States. *Psychol Bull* **135**, 749–773 (2009).

38. Miller, S. G. Clinical reasoning and decision-making in cases of child alignment: Diagnostic and therapeutic issues. in *Working with alienated children and families: A clinical guidebook* (eds. A. J. L. Baker & S. R. Sauber) 8–46 (2013).

39. Möller, S. *et al.* Human and macaque pairs employ different coordination strategies in a dyadic decision game with face-to-face action visibility. *bioRxiv* 2020.03.13.983551 (2021).

40. Charpentier, C. J., Iigaya, K. & O'Doherty, J. P. A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. *Neuron* **106**, 687-699.e7 (2020).

41. Fan, S., Dal Monte, O. & Chang, S. W. C. Levels of naturalism in social neuroscience research. *iScience* **24**, 102702 (2021).

42. Fagot, J. & de Lillo, C. A comparative study of working memory: Immediate serial spatial recall in baboons (Papio papio) and humans. *Neuropsychologia* **49**, 3870–3880 (2011).

43. Thompson, R. K. R. & Oden, D. L. Categorical Perception and Conceptual Judgments by Nonhuman Primates: The Paleological Monkey and the Analogical Ape. *Cogn Sci* **24**, 363–396 (2000).

44. Parron, C. & Fagot, J. Comparison of grouping abilities in humans (Homo sapiens) and baboons (Papio papio) with the Ebbinghaus illusion. *J Comp Psychol* **121**, 405–411 (2007).

45. Claidière, N., Gullstrand, J., Latouche, A. & Fagot, J. Using Automated Learning Devices for Monkeys (ALDM) to study social networks. *Behav Res Methods* **49**, 24–34 (2017).

46. Fagot, J., Gullstrand, J., Kemp, C., Defilles, C. & Mekaouche, M. Effects of freely accessible computerized test systems on the spontaneous behaviors and stress level of Guinea baboons (Papio papio). *Am J Primatol* **76**, 56–64 (2014).

47. Philippe, R. *et al.* Neurocomputational mechanisms engaged in detecting cooperative and competitive intentions of others. *Nat Commun* in press doi:10.21203/rs.3.rs-1160167/v1.

48. Keating, C. F. & Bai, D. L. Children's Attributions of Social Dominance from Facial Cues. *Child Dev* **57**, 1269 (1986).

49. Brey, E. & Shutts, K. Children Use Nonverbal Cues to Make Inferences About Social Power. *Child Dev* **86**, 276–286 (2015).

50. Charafeddine, R. *et al.* How Preschoolers Use Cues of Dominance to Make Sense of Their Social Environment. *http://dx.doi.org/10.1080/15248372.2014.926269* **16**, 587–607 (2015).

51. Terrizzi, B. F., Brey, E., Shutts, K. & Beier, J. S. Children's developing judgments about the physical manifestations of power. *Dev Psychol* **55**, 793–808 (2019).

52. Over, H. & Carpenter, M. The Social Side of Imitation. *Child Dev Perspect* **7**, 6–11 (2013).

53. Gülgöz, S. & Gelman, S. A. Who's the Boss? Concepts of Social Power Across Development. *Child Dev* **88**, 946–963 (2017).

54. Enright, E. A., Alonso, D. J., Lee, B. M. & Olson, K. R. Children's Understanding and Use of Four Dimensions of Social Status. *https://doi.org/10.1080/15248372.2020.1797745* **21**, 573–602 (2020).

55. Charafeddine, R. *et al.* Do Preschoolers Align Their Preferences With Those of a Powerful Individual? *Front Educ (Lausanne)* **5**, 293 (2021).

56. Bernard, S. *et al.* The boss is always right: Preschoolers endorse the testimony of a dominant over that of a subordinate. *J Exp Child Psychol* **152**, 307–317 (2016).

57. Charafeddine, R. *et al.* Cross-Cultural Differences in the Valuing of Dominance by Young Children. *J Cogn Cult* **19**, 256–272 (2019).

58. Marr, D. *Vision*. (W. H. Freeman, 1982).

59. Lockwood, P. L., Apps, M. A. J. & Chang, S. W. C. Is There a 'Social' Brain? Implementations and Algorithms. *Trends Cogn Sci* **24**, 802–813 (2020).