



**HAL**  
open science

# Aspects neural et développemental de la perception de relations spatiales socialement pertinentes

Nicolas Goupil

► **To cite this version:**

Nicolas Goupil. Aspects neural et développemental de la perception de relations spatiales socialement pertinentes. Cognitive science. Université Claude Bernard - Lyon I, 2023. English. NNT : 2023LYO10068 . tel-04557983

**HAL Id: tel-04557983**

**<https://theses.hal.science/tel-04557983>**

Submitted on 24 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THESE de DOCTORAT DE  
L'UNIVERSITE CLAUDE BERNARD LYON 1**

**Ecole Doctorale N°476  
Neurosciences et Cognition**

**Discipline** : Sciences cognitives

Soutenue publiquement le 02/05/2023, par :  
**Nicolas Goupil**

---

**Neural and developmental aspects  
of the perception of socially-relevant  
spatial relations between people**

---

Devant le jury composé de :

Cristofori, Irene  
Southgate, Victoria  
Baudoin, Jean-Yves  
Leleu, Arnaud  
Papeo, Liuba  
Hochmann, Jean-Rémy

Maître de conférences  
Professeure  
Professeur des Universités  
Maître de Conférences  
Directrice de Recherche  
Directeur de Recherche

UCBL  
Université de Copenhague  
Université Lyon 2  
Université de Bourgogne  
UCBL  
UCBL

Examinatrice  
Présidente et rapporteure  
Rapporteur  
Examinateur  
Directrice de thèse  
Co-Directeur de thèse

## Acknowledgments

A few years ago, chances weren't so high that I would write a thesis. People made this possible, and for that, I am forever grateful.

To my supervisor, Liuba Papeo. Your scientific spirit inhabits every single word of this dissertation. I know it will continue to guide me in the future, and that I will discover, only there, how much you gave me. To my co-supervisor, Jean-Rémy Hochmann, whose vigilance built my trust in what I was doing, when I was failing. Liuba, Jean-Rémy, I am very proud that I have been part of your laboratory, where so many fantastic projects will be realized!

My second word goes to Victoria Southgate and Jean-Yves Baudouin, who accepted to review this thesis. My greatest hope is that you find some inspiration here, like I found inspiration in your works. I am also grateful to Irene Cristofori and Arnaud Leleu, who honored me with their presence at my defense, as well as Lara Bardi and Paul Downing, who shared their advice and care during my yearly guidance committee. I hope we all meet again!

In these years, my greatest luck has been to work at the beautiful Institut des Sciences Cognitives Marc Jeannerod. I am grateful to the successive directors, Angela Sirigu, who welcomed our team when we moved in, and Pier-Francesco Ferrari, who supported our investigation so closely. I also thank past and present workers of the institute: Annie Busson, Corinne Ruppé, Stéphane Duray, Emmanuelle Dubois-Trépas, May Cha Ly and Sophie Reysset for their support in all administrative matters, Sylvain Maurin, Johan Pacquit, Marco Bimbi and Thomas Perret for keeping safe and working the numeric space, Serge Pinède and Fabrice Herant for giving me a hand anytime I struggled with practical matters. Thank you all for your help and warmth.

Now to all the people of the institute, with whom I shared so many wonderful moments. S/o loud to my big brother in science, Étienne Abassi, we shared everything in science and life, to my sister in science, Céline Spriet, together we discovered the thrills of international scientific conferences (on TeamViewer), and to Émilie Serraille, the pillar of our team, with whom everything started. Love to Parvaneh Adibpour, who taught me the mysteries of EEG, Sofie Vettori, who could order my divagations into meaningful ideas, Mathilde Kremer, who taught me patience and care, and Jérémy Decroix, who shared my addiction to R. Many thanks to those who helped on anything: Alice, Sara, Marion, Marie H, Ilenia, Cristina, Bastien, Thibaut, Lisa, Elouan, and those who were just cool: Maëva & Mathilda, Antoine, Valentin, Hajer, Eloïse, Seb, Rémy 1&2, Toan, Felipe & Marie V, Jimmy & Holly, Jacopo & Giulia, etc. (if you don't find your name here, I just kept it in my heart <3). Final thanks to all the people who joined the studies (especially 001GB), for contributing and showing interest. The whole edifice is built on your participation.

À ma famille. Mes parents, Sophie et Louis, qui se sont parfois inquiétés. J'espère que vous êtes fiers ! Mémé et Juju, qui m'ont accueilli dans leur sanctuaire. Mon frère, Pierre, qui a toujours veillé. À Boris, Alexandre.

À ma sœur, Louise, qui m'a mis sur la voie.

## Table of content

Acknowledgments .....	2
Table of content .....	3
Résumé substantiel (français) .....	6
Abstract .....	9
General introduction .....	10
1. Why should we look at people face-to-face? .....	10
2. Perceiving <i>facingness</i> to represent social interaction .....	12
2.1. Representing social scenes from visual information .....	12
2.2. Perceiving social scenes in spatio-temporal visual information .....	13
2.3. Perceiving people facing as a whole .....	14
3. Development of social scene perception .....	16
3.1. Grounding infant cognition in perception .....	16
3.2. Representing third-party interactions in infancy .....	18
3.3. Perceiving social interactions: an effect of facingness? .....	19
4. Thesis overview .....	20
4.1. Current research .....	20
4.2. Overview of stimuli .....	20
4.3. Overview of chapters .....	21
Chapter 1 – Intermodulation responses show integration of interacting bodies in a new whole .....	23
1. Introduction .....	23
2. Materials & method .....	24
3. Results .....	31
4. Discussion .....	36
Chapter 2 – Visual perception grounding of social cognition in preverbal infants .....	39
1. Introduction .....	39
2. Experiment 1 – Facing vs. Non-facing .....	43
2.1. Materials & Method .....	43
2.2. Results .....	48
3. Experiment 2 – 1 vs. 2 .....	51
3.1. Materials & Method .....	51
3.2. Results .....	52
4. Experiment 3 – <i>Reciprocity</i> .....	53
4.1. Materials & Method .....	53

4.2. Results .....	53
5. Experiment 4 – Body-object dyads .....	54
5.1. Materials & Method.....	55
5.2. Results .....	55
6. Discussion.....	56
Chapter 3 – A visual preference for face-to-face social relations in humans and macaque monkeys.....	60
1. Introduction .....	60
2. Experiment 1 – Human adults .....	62
2.1. Materials & Method.....	62
2.2. Results .....	64
3. Experiment 1b – Rating study on human adults.....	64
3.1. Materials & Method.....	64
3.2. Results .....	65
4. Experiment 2 – Human infants and children .....	66
5. Experiment 2a – Infants.....	66
5.1. Materials & Method.....	66
5.2. Results .....	67
6. Experiment 2b – Children.....	69
6.1. Materials & Method.....	69
6.2. Results .....	69
7. Experiment 3 – Macaques .....	71
7.1. Materials & Method.....	71
7.2. Results .....	72
8. Discussion.....	72
General discussion.....	76
1. Grouping bodies .....	76
2. Social evaluation .....	77
3. Outstanding questions and limitations.....	78
3.1. The “attentional hotspot” hypothesis.....	79
3.2. Integrating visual and social information .....	80
4. Conclusion.....	80
Bibliography .....	81
Appendices .....	98
1. Appendix A – Supplementary material of Chapter 1 .....	98
1.1. Alpha suppression analysis.....	98

2.	Appendix B – Supplementary materials of Chapter 2 .....	101
3.	Appendix C – Supplementary material of Chapter 3.....	104
3.1.	Supplementary material of Experiment 1b .....	104
3.2.	Memory task with human adults.....	105
3.3.	Macaques rearing.....	106
4.	Appendix D – Published version of Visual perception grounding of social cognition in preverbal infants (Goupil, Papeo & Hochmann, 2022).....	107
5.	Appendix E – Preprint of Category-specific effects of high-level relations in visual search (Goupil, Kaiser & Papeo, 2023).....	107

## Résumé substantiel (français)

De nombreux aspects de la vie en société tels que les normes culturelles ou les relations sociales ne s'observent que dans les interactions d'autres personnes. Reconnaître et comprendre ces interactions, dites sociales, est de première importance pour la vie en groupe, donc la survie d'espèces sociales. Pour détecter, reconnaître et identifier les agents sociaux, la vision humaine a développé des mécanismes perceptifs spéciaux. Ces mécanismes semblent émerger précocement dans la vie, afin d'orienter l'attention visuelle du bébé le plus tôt possible vers les aspects de l'environnement les plus pertinents socialement, et ainsi poser les bases de la cognition sociale. Les présents travaux abordent l'hypothèse selon laquelle la perception visuelle développe également des mécanismes pour détecter, reconnaître et identifier précocement les interactions sociales elles-mêmes, par-delà l'identification d'agents individuels. Son focus porte sur l'un de ces mécanismes, le groupement automatique d'agents sociaux (i.e. deux corps) dans une configuration spatiale suggérant qu'ils entretiennent une forme très particulière d'interaction sociale : l'interaction face-à-face. Cette recherche se déroule en trois chapitres, traitant des questions suivantes : Les personnes se faisant face sont-elles perçues différemment des personnes ne se faisant pas face (Chapitre 1) ? À quel âge les bébés deviennent-ils sensibles aux relations spatiales d'autres personnes, dont la pertinence sociale varie (Chapitre 2) ? Quelle est la relation entre représenter deux personnes face-à-face (c'est-à-dire une relation spatiale) et représenter une interaction sociale (Chapitre 3) ?

Sur la base de recherches précédentes (Adibpour et al., 2021 ; Papeo et al., 2017 ; 2019), le premier chapitre étudie l'hypothèse selon laquelle le système visuel groupe automatiquement les corps de face (mais pas de dos), en une seule unité perceptive. Ceci fut mesuré chez l'humain adulte par électroencéphalographie (EEG), dans un paradigme de « dual frequency-tagging » (marquage fréquentiel double). Dans ce paradigme, des dyades de corps furent présentées de face ou de dos, chaque individu clignotant à une vitesse (fréquence) différente. Afin de contrôler la spécificité des effets évoqués à la perception de corps, des paires d'objets (chaises, machines) furent présentés de la même manière, de face ou de dos, avec le même clignotement. L'activité neurale reflétant l'intégration des corps en une unité visuelle (signal d'intermodulation) fut ensuite dissociée de l'activité neurale évoquée par le clignotement individuel des corps (signal individuel) par analyse fréquentielle. Cette analyse a révélé un signal d'intermodulation pour les corps de face mais pas les corps de dos, ainsi qu'un signal individuel plus faible pour les corps de face que les corps de dos. Aucun de ces effets ne fut observé avec des objets. Ainsi, le chapitre 1 suggère que le système visuel groupe les personnes se faisant face en une seule unité perceptive, une représentation unitaire capable d'affecter le traitement visuel des corps individuels.

Le second chapitre étudie l'hypothèse selon laquelle les relations spatiales entre les personnes sont discriminées dès la petite enfance, avant même le développement de capacités cognitives sociales sophistiquées. Dans cette étude, le regard de bébés de 6 mois a été mesuré par « eye-tracking » (suivi oculaire), dans un paradigme de temps de

regard différentiel. L'expérience 1 montre que les bébés de 6 mois discriminent les corps qui se font face des corps qui ne se font pas face, et regardent plus longtemps les corps qui ne se font pas face. L'expérience 2 montre ensuite que les bébés regardent aussi longtemps des corps de face qu'un corps isolé (mais moins longtemps que deux corps qui ne se font pas face), comme s'ils traitaient les corps de face ainsi qu'une unité visuelle (e.g. un seul corps). L'expérience 3 montre que les bébés discriminent globalement les relations spatiales qui signalent une interaction, qu'elle soit réciproque (les deux corps sont de face) ou non réciproque (seulement un corps fait face à l'autre), de la relation spatiale qui indique l'absence d'interaction (les deux corps ne se font pas face). Enfin, l'expérience 4 montre que les effets ci-dessus sont spécifiques aux dyades humain-humain (sociales) et ne se généralisent pas aux paires humain-objet (non sociales). En résumé, le chapitre 2 indique que les bébés sont sensibles aux relations spatiales entre les personnes dès l'âge de 6 mois, et suggère que comme les adultes, les bébés traitent les corps de face plus efficacement que les corps de dos, comme une unité visuelle.

Le troisième chapitre teste l'hypothèse qu'une scène dans laquelle deux personnes se tiennent de face est spontanément interprétée comme une interaction sociale. Ceci fut mesuré dans un paradigme de temps de regard différentiel similaire à celui du chapitre 2, présentant deux corps de face simultanément avec deux corps de dos, cette fois-ci à des participants adultes. Dans ce paradigme, l'attribution d'une valeur sociale plus élevée à un type de stimulus (e.g. des corps de face) se manifeste par une préférence visuelle (un temps de regard plus long) par rapport à un type de stimulus dont la valeur sociale est inférieure (e.g. des corps de dos). Indépendamment de cette mesure (Expérience 1a), une expérience a mesuré la valeur des stimuli sur des dimensions sémantiques sociales (Expérience 1b). La préférence visuelle a ensuite été testée tout au long du développement humain (Expérience 2a : bébés de 7, 10, 15, 18 mois ; Expérience 2b : enfants de 3, 5 ans), puis dans une espèce de primates sociaux non humains (Expérience 3 : macaques). L'expérience 1 révèle que les adultes regardent plus longtemps les corps de face (expérience 1a), qu'ils jugent supérieurs que les corps de dos sur les dimensions sémantiques sociales (expérience 1b), montrant ainsi une préférence visuelle et une évaluation sociale plus élevée pour les corps de face. L'expérience 2 révèle que cette préférence apparaît à 5 ans, passé le stade précoce de développement découvert dans le chapitre 2 durant lequel bien que les bébés discriminent les corps de face et les corps de dos, ils ne montrent pas de préférence pour les corps de face. L'expérience 3 montre que les macaques, comme les humains, préfèrent visuellement des congénères de face plutôt que de dos. En somme, le chapitre 3 indique que la sensibilité perceptive qu'ont les bébés vis-à-vis du positionnement relatif des corps se développe en une préférence visuelle pour le face-à-face durant l'enfance, pouvant refléter l'attribution d'une signification sociale particulière. Ce processus semble également se développer chez des primates sociaux non humains.

Cette thèse propose que la perception visuelle représente automatiquement les personnes face-à-face, une relation spatiale suggérant l'interaction, comme une unité (Chapitre 1). Ce groupement semble faciliter le traitement des scènes sociales dès 6 mois de vie, un âge durant lequel les bébés sont particulièrement sensibles aux relations spatiales

d'autres personnes (Chapitre 2). Cette sensibilité pourrait servir de base au développement de capacités cognitives sociales plus avancées durant l'enfance, non seulement chez l'homme mais aussi d'autres espèces primates non humaines vivant en société (Chapitre 3).

## Abstract

Recognizing and understanding social interactions is particularly important for the survival of group-living species. Human vision has developed highly sensitive perceptual mechanisms to detect, recognize and identify social agents. These mechanisms, emerging early on in life, may serve to guide visual attention toward socially relevant aspects of the environment, and support the development of social cognition. The present thesis addresses the hypothesis that visual perception has also developed efficient mechanisms to detect, recognize and identify social *interaction*. I focus on one such mechanism, proposed to account for the efficient visual processing of social interaction: the automatic grouping of social agents (i.e., bodies) seen face-to-face, a particularly important (and common) spatial positioning for social interaction. In Chapter 1, using frequency-tagging EEG in human adults, I provide evidence for the automatic integration of two people facing each other. In Chapter 2, using eye-tracking in a differential looking time paradigm, I show that infants as young as 6 months already show a sensitivity to differences between social scenes based on the spatial positioning of bodies (facing or non-facing), and that, like adults, they may process facing bodies more efficiently than non-facing bodies. Finally, in Chapter 3, using the same paradigm as Chapter 2, I describe the transition from the perceptual discrimination of spatial relations between people in infants, to understanding the social relevance of facing (*vs.* non-facing) bodies in childhood. In this chapter, I also present evidence suggesting that, like children and adults, monkeys (*macaca mulatta*) assign an exceptional social significance to the face-to-face body configuration. Overall, results show that visual perception represents seemingly interacting (face-to-face) bodies as a group, or a single perceptual unit, that may facilitate the processing of facing bodies already in infancy. I propose that this perceptual sensitivity to spatial relations between people anticipates and sets the ground for the development of social cognitive abilities.

# General introduction

## 1. Why should we look at people face-to-face?

Living in social group has compelled the development of complex cognitive skills, from learning from others to managing resource exchanges, while avoiding deception and misinformation (Baumard et al., 2013; Byrne, 1996; Cheney et al., 1986; Csibra & Gergely, 2011; Dunbar & Schultz, 2007; Gergely & Csibra, 2013; Robson & Kaplan, 2003). For these skills to be realized, however, individuals first need to detect, recognize and identify conspecifics fast and accurately. Primate vision is remarkably prepared for this task, showing high sensitivity to visual patterns characterizing bodies and body movements of conspecifics and other animals (Adolphs, 2001; Matsuno, 2021; Nakayama, 2010).

Visual information provides, on its own, important cues to understanding others. Faces, gaze direction, body postures or movements provide information about other agents' identity, personality traits, emotional state, focus of attention and intentions (Aviezer et al., 2012; Macrae & Quadflieg, 2010; Poyo Solanas et al., 2020; Oosterhof & Todorov, 2008). Ubiquitous to human visual environments since birth (Jayaraman et al., 2015; Fausey et al., 2016), these stimuli trigger perceptual mechanisms that support important socio-cognitive abilities (Baron-Cohen, 1997; Morton & Johnson, 1991). Yet looking at others individually overlooks essential aspects of the social world, only expressed in collective behaviors and inter-individual interactions (Quadflieg & Koldewyn, 2017). Viewing exchanges between others gives access to key information, without the risks that come from interacting oneself (McGregor et al., 1993; Lloyd & Morrison, 2008). By observing others' interactions, one can learn information such as the value of goods or the use of tools (Paradise & Rogoff, 2009; Tatone et al., 2021; Thiele et al., 2021b), the individuals' morality, trustworthiness or aggressivity (Decety & Cacioppo, 2012; Heider & Simmel, 1944; Hesse et al., 2016), the implicit norms that rule social interactions (e.g., social distance; Remland et al., 1995), and the networks of hierarchies and alliances in which agents and observers are embedded (Cloutier et al., 2016; Farrow et al., 2011; Lau et al., 2020).

Early research on social scene perception has highlighted the fact that, like other stimuli, social interactions possess typical visual structures (Heider & Simmel, 1944; McArthur & Baron, 1983). For example, people perceive and act upon each other better when they stand close and facing toward each other (Jiang et al., 2012; Lahnakoski et al., 2020). This spatial relation scaffolds the first social experiences infants have with their caregivers (Meltzoff & Moore, 1977; Slaughter, 2021; Tomasello & Farrar, 1986; Trevarthen, 1974). Viewing other persons face-to-face gives full access to their facial expression, gaze direction, body posture, and gestures, supporting many forms of social interactions, like shared attention, joint action, imitation, communication, cooperation, interaction, social learning and affective bonding (Baron-Cohen, 1997; Carpenter et al., 1998; Chartrand & Bargh, 1999; Csibra & Gergely, 2011; Ferrari & Gallese, 2007; Gergely & Csibra, 2013; Mundy & Newell, 2007; Sebanz et al., 2006).

Dyads of people interacting face-to-face are ubiquitous in human social environments (Dunbar et al., 1995; James, 1951).

*Facingness*, the mutual perceptual accessibility of two social agents, may provide one of the most reliable visual cues of social interactions (Papeo et al., 2017; Papeo, 2020). Recent research has shown that vision is particularly sensitive to this visuo-spatial cue (Papeo, 2020). Viewing people face-to-face increases their visual saliency (Papeo et al. 2017; Papeo et al., 2019; Vestner et al., 2019), and changes, by enhancing, how their bodies are represented in visual brain areas (Abassi & Papeo, 2020; 2022; Bellot et al., 2021). Hence, visual perception seems to possess a mechanism able to use spatial relations between people that are socially-relevant, i.e. that suggest social interaction, to represent them (i.e., bodies and their actions) more efficiently (Papeo, 2020). Like other perceptual mechanisms designed to process important biosocial signals (e.g., faces and biological motion), the perceptual mechanisms underlying the efficient processing of social interaction may be in place early on in life. The current thesis examines the following questions: what mechanism can account for the particularly efficient processing of facing bodies? And how early does this mechanism emerge in life?

To provide a background to these questions, the next sections will discuss research emphasizing the effect of *facingness* in social scene perception, and motivate the hypothesis of perceptual grouping as a key mechanism for the efficient processing of people in social interaction. Then, I will present research suggesting a close link between social scene perception and the development of social expectations in infancy. This overview will highlight the possible contribution of the present investigation on visual perception to understanding the groundings of social cognition.

## 2. Perceiving *facingness* to represent social interaction

*This section reviews the effects of facingness on social scene perception, and how these effects suggest that facing bodies are perceptually grouped.*

### 2.1. Representing social scenes from visual information

Research over decades has extended our understanding of visual functions, from the mere encoding of basic features (motion, orientation, color, etc.) to representing objects (Gibson, 1961; Marr, 1982), and the tuning of highly specialized mechanisms dedicated to detect, identify and recognize social entities (Nakayama, 2010). Research showed that vision encodes with particular efficiency the complex structure of social entities, promptly integrating multiple parts into coherent *wholes*. This attunement was highlighted in the most automatic visual processes, in studies reporting that information about the structure of faces, bodies or biological movement, enhances the discrimination and representation of social entities (Downing et al., 2004; Johansson, 1973; Neri et al., 1998; Purcell & Stewart, 1988). For instance, research on face perception highlighted the key role of the spatial configuration of face parts for processing faces, by showing how breaking such configuration by inverting faces, while preserving low-level visual information such as luminance or contrast, drastically reduces performance in the detection or identification of people (Bruyer, 2011; Maurer, 2002; Yin, 1969).

Parallel research has investigated whether similar visual efficiency could support the detection of social interactions. This started in 1944, with Heider and Simmel's hypothesis that minimal spatio-temporal patterns could spontaneously evoke rich representations of social interactions (Heider and Simmel, 1944). Showing that participants spontaneously reported intuitions about animacy, personality, intentions and social relationship such as "chasing", "marrying", "fighting", from the movements of geometrical shapes, this study first demonstrated that social interaction can be inferred from the mere spatiotemporal features of an event. Years later, neuroimaging research used Heider and Simmel's kind of animations to study the brain networks responding to perception of intentional movements (Castelli et al., 2000), and isolated the representation of motion related to social events *versus* mechanical events (Martin & Weisberg, 2003). Using comic strips or videos, subsequent works isolated the processing of social interactions from the processing of individual actions (Centelles et al., 2011; Iacoboni et al., 2004; Kujala et al., 2012; Quadflieg et al., 2015; Walter et al. (2004). showing different brain activations associated with viewing interacting agents *vs.* viewing agents acting independently.

Overall, neuroimaging studies have unraveled three distinct networks that are critical to support the representation of social scenes. First, there is a network of visual brain structures specialized in perceiving social stimuli such as faces, gaze, bodies and biological motion (Downing et al., 2001; Grossman et al., 2000; Haxby et al., 2000; Hoffman & Haxby, 2000; Kanwisher et al., 1997; Macrae & Quadflieg, 2010; Peelen & Downing, 2007). Then, there are a second network driving visual information towards internal motor representations (Jeannerod,

2001; Rizzolatti et al., 2001), and a third network supporting inferences about the mental states of others (Frith et al., 2003; Gobbini et al., 2007; Schurz et al., 2014). Summarizing the evidence, some researchers have proposed a tripartite model of social scene representation, emphasizing the core role of visual processes specific to perceiving social scenes (Quadflieg & Penton-Voak, 2017; Quadflieg et al., 2015; Yang et al., 2015). Along these lines, more recently, researchers have highlighted a third visual pathway specialized in visual processing of social information (Grossmann, 2021; Pitcher & Ungerleider, 2021; Weiner & Gomez, 2021). Based on neuroanatomical and functional neuroimaging evidence, there would be a distinctive stream of visual brain areas dedicated to processing social stimuli, flowing from the early visual cortex through motion selective areas to body selective areas up to the posterior superior temporal sulcus. Converging with the tripartite model described above, this model roots cognitive process representing social scenes in the visual brain, as it proposes that regions tuned to socially relevant visual stimuli (face, body) and biological motion perception provide key input-information to the neural networks underlying social cognition.

## **2.2. Perceiving social scenes in spatio-temporal visual information**

Following Heider & Simmel (1944), researcher have defined the spatiotemporal visual features driving the representation of social interaction, such as the agents' converging direction, the contingency of their movements, or their synchronization (Bassili et al., 1976; Scholl & Tremoulet, 2000). This research showed that manipulating spatiotemporal features may not only change the interpretation of social events, but also trigger fast and automatic biases in the efficiency with which observers of a social scene perform a concurrent task. For instance, deviating how directly a shape moves toward another shape, which creates the illusion of a “chase”, decreases how efficiently participants can detect the chasing shape, in a task that demands to avoid it (Gao & Scholl, 2011).

One line of work has investigated how the presence of one agent influences the perception of another agent, and how this may be mediated by visual features cueing an interaction between both agents. Using point-light displays of human figures depicting moving human bodies through a few dots (Johansson, 1973), a study showed that the movement of one person was discriminated better in the presence of a second person involved in a synchronized interaction, even when this information was irrelevant to the task (Neri et al., 2006). Following up on this result, Hirai & Kakigi (2009) investigated the extent to which the interactivity of walkers could influence their neural representation. They showed that walkers moving toward each other induced a stronger impression of social interaction than walkers moving away from each other. This difference was absent when stimuli were inverted, indicating that it depended on processing walkers as bodies, and not just their local motion cues. Then, measuring neural activity with magneto-encephalography, Hirai & Kakigi (2009) found that directing walkers toward each other affected their neural representation, increasing the peak amplitude of neural activity in the left hemisphere, 300 milliseconds (ms) after stimulus onset. Again, this difference was evidenced with upright stimuli but not inverted stimuli. Using the same manipulation, a recent fMRI study further showed that point-light displays of

people moving and acting toward each other elicit a stronger neural activity and connectivity in visual areas dedicated to the perception of bodies and biological motion (Bellot et al., 2021). Critically, this study found that the neural representation of the whole scene, as well as of the individual bodies, was more accurate when the two people moved toward (versus away from) each other. Consistent with these neural effects, a match-to-sample task in Bellot et al. (2021) showed better performance with bodies moving toward (vs. away from) each other when they were upright, but not inverted.

Most research on social interaction perception and understanding has employed dynamic stimuli, and rightly so, as social interactions are dynamic events unfolding over time. However, all the key visual features of a social interaction may not involve motion. *Facingness*, a spatial positioning that sets people in a configuration in which they can perceive each other, may trigger the representation of social interaction even in the absence of motion. In the next section, I will introduce the effects of this static visual feature on visual perception.

### **2.3. Perceiving people facing as a whole**

Papeo, Stein and Soto-Faraco (2017) investigated how *facingness* influences the perception of social scenes, that is, visual scenes encompassing two human bodies. To study this, the authors presented static images of body dyads, either facing toward or facing away from each other, very briefly and before a mask appeared to block the visual processing of stimuli. Body stimuli were interspersed with non-social stimuli (chairs and plants), and participants were instructed to recognize the visual category of each image (body, chair, or plant). In half trials, the stimuli were presented inverted. In this backward-masking paradigm, authors showed an advantageous recognition of facing (vs. non-facing) bodies. Moreover, measuring the difference in accuracy between upright and inverted stimuli, the so-called inversion effect, authors showed that the cost of inversion was larger for facing bodies than for non-facing bodies. No such effect emerged with chairs, indicating a process specific to perceiving relations between bodies: the two-body inversion effect. Since the extensive face perception literature has shown that the magnitude of the inversion effect is an index of visual sensitivity (i.e., the inversion effect is the largest for stimuli such as faces and bodies, to which the visual system is maximally sensitivity, than for any other familiar object) the larger inversion effect for facing (vs. facing) bodies first suggested that human vision is particularly sensitive to facing bodies. Later experiments showed that the inversion effect for dyadic stimuli was maximal for bodies face-to-face, relative to bodies in non-reciprocal relationship (one bodies faces the back of another body), human-object pairs or object-object pairs (Papeo & Abassi, 2019). Moreover, Papeo et al. (2017) showed that participants were faster at judging facing (vs. non-facing) bodies along an arbitrary dimension, a speed as fast as judging single bodies. This suggested that facing bodies, in contrast to non-facing bodies, could be processed as fast as one body; that is, as one single visual unit.

Visual search experiments provided further evidence that facing bodies may be perceptually grouped into one unit. Papeo, Goupil & Soto-Faraco (2019) compared how efficiently participants could find a target dyad of facing bodies (facing targets) among a set of distracting dyads of non-facing bodies, with how efficiently participants could find a target dyad of non-facing bodies (non-facing target) among a set of distracting dyads of facing bodies (for illustration of the paradigm, see Figure 1 of Appendix E). This study showed that when the task parameters required participants to search serially through arrays, they detected non-facing targets faster than facing targets. As this effect increased with the number of distractors, it reflected how efficiently participants could go through the distractors (i.e., could check and reject them) before finding the target. This pattern was analogous to previous effects showing that the search for a target is facilitated when distracting stimuli can be grouped in larger perceptual units: by reducing the number of distractors, grouping would facilitate the detection of targets (Kaiser et al., 2014; 2019). In this perspective, the results of Papeo et al. (2019) suggested that facing dyads are automatically grouped. Authors directly addressed this hypothesis in a second experiment, by testing whether participants could access an individual body as efficiently when it composed a facing dyad, or a non-facing dyad. This could relate to known object-inferiority effect, that is, a cost on accessing individual features composing a grouped visual unit, for instance an individual trait drawing the eye or the mouth of a face-like pattern (Suzuki & Cavanagh, 1995). Confirming perceptual grouping, participants detected individual bodies less efficiently in facing dyads than non-facing dyads. (follow-up experiments on visual search are reported in Appendix E).

Converging with behavioral experiments, neuroimaging research provided evidence that facing dyads are processed as a group. Abassi & Papeo (2020) tested the neural processing of facingness with fMRI, by presenting participants with single bodies, facing bodies, and non-facing bodies. This study evidenced that facing dyads elicit stronger neural responses than non-facing dyads in body-selective visual brain areas. More direct evidence for grouping came from multivariate pattern combination (MVPC) analysis. In MVPC, the neural response to a pair of objects is modeled by the responses to the individual constituent objects: If the pair response is accurately modeled by the individual responses, it can be assumed that the objects are processed independently and in parallel. Using MVPC, Abassi & Papeo (2020) found that for non-facing bodies, the pair response was modeled by an average of the individual-body responses, suggesting an independent processing of unrelated bodies. By contrast, when bodies were face-to-face, the pair response was less well predicted by the individual responses in an anterior part of the cortex, suggesting the involvement of additional integrative processes.

In all the above studies, scene perception was consistently advantaged when people were face-to-face relative to when they were back-to-back, and this sensitivity related to processing facing bodies as a single perceptual unit, or a whole. To recall, in Papeo et al. (2017), the two-body inversion effect indexed a stronger cost of inversion with facing dyads, associated to higher visual sensitivity mediated by configural processing. Moreover, participants processed facing bodies as fast a single body, as if facing bodies were processed as a unit. In Papeo et

al. (2019), participants searched through facing bodies more efficiently, but accessed individual bodies in facing dyads less efficiently, as if the primary representation of facing dyads is a whole that needs to be *disassembled* in order to access its individual components (the single bodies). In Abassi & Papeo (2020), the neural response to body dyads was stronger with facing bodies, which the brain processed non-linearly, suggesting a holistic process. Together, existing evidence converge in showing processing differences between facing and non-facing bodies, to the advantage of facing bodies, possibly resulting from processing individual bodies as a single perceptual unit, as opposed to two units.

### **3. Development of social scene perception**

*This section addresses the development of social scene perception in infancy. It highlights the contribution of developmental research to understanding the emergence of core social cognitive skills, mechanisms of social scene perception, and some evidence suggesting a precocious influence of facingness.*

#### **3.1. Grounding infant cognition in perception**

The visual mechanisms of social perception (e.g., perception of faces, bodies and biological motion) emerge early in life and possibly contribute to developing a mature social cognitive system. Developmental studies have shown a tight relationship between early developing perceptual mechanisms and cognitive abilities in several domains of cognition (Carey, 2009). Infants' representation of objects lays on the perception of cohesion, boundedness and movement on a continuous path (Spelke, 1990). Perception of causality correlates with the spatiotemporal continuity in the motion of objects (Leslie & Keeble, 1987; Michotte, 1963; Muentener & Carey, 2010; Scholl & Termoulet, 2000), and infants attribute intentionality and goal-directedness to self-propelled objects (Csibra, 2003; Gergely et al., 1995; Premack, 1990), and objects with eyes (Gergely & Csibra, 2003; Newman et al., 2010).

Very early on, infants discriminate social entities from visually-matched stimuli, relying on the global structure of stimuli. For example, studying face perception, many studies (e.g., Morton & Johnson, 1991; Farroni et al., 2005) presented newborns with schematic stimuli composed of two dots above a third dot (like eyes above a mouth), or irregular stimuli having the inverted pattern (one dot above two dots). Adopting a preferential looking paradigm, researchers presented both types of stimuli simultaneously, and recorded which type of stimuli neonates spontaneously preferred to look at (i.e., looked longer). They found that neonates preferred to look at face-like schemata over irregular stimuli. A study by Bardi et al. (2011) unraveled a similar effect in the case of biological motion. Newborns were simultaneously presented with two sorts of point-light displays: one moved according to the principles of biological motion (i.e., a hen walking, exemplifying semi-rigid motion), the other with rigid patterns of motion. Infants preferred point-lights displays obeying semi-rigid biological-like motion. Using a similar paradigm, Zieber et al. (2015) presented 3.5-months-old infants with displays showing a regular body next to a body

with limbs rearranged in an irregular configuration (e.g., the arms stemming from the knees). Infants discriminated the two stimuli, but instead of preferring the prototypical body configuration (like neonates prefer face-like patterns), they looked longer at irregular body configurations, which was interpreted as surprise (similar effects have also been reported with “uncanny” faces; Lewkowicz & Ghazanfar, 2011). This effect was found when bodies were upright, but not inverted, suggesting that infants processed the structure of bodies rather than just discriminating images based on the low-level visual differences.

As infants grow up, they extract and process more complex information about social entities. Early studies found that neonates’ preference for faces disappears after one month to reappear after two months. Morton & Johnson (1991) interpreted this U-shaped developmental trajectory as a shift from a first process enabling infants to detect conspecifics, to a second process enabling infants to learn about them. Investigating the development of holistic face perception over the first semester of life, Cashon & Cohen (2004) found a dissociation between the development of infants’ discrimination of inverted faces, following an inverted U-shaped trajectory (i.e. absent at 3 months, present at 4-6.25 months, absent at 7 months), and their processing of upright faces, following an N-shaped trajectory (i.e. absent at 3 months, present at 4-5.75 months, absent at 6.25 months, present at 7 months). Authors interpreted this dissociation as reflecting a shift in infants’ processing of faces from a local to a global level, that is, from processing features independently to integrating them into a whole. Research on body perception has evidenced a similar development, from encoding the structure of bodies to extracting more complex information. Infants as young as 3.5 months can discriminate bodies in their regular configuration, from bodies in irregular configuration, for instance with legs in place of arms (Zieber et al., 2015). However, infants may not recognize information about the sex of a person before 5 months, and recognize emotional body movements before 6.5 months (Bhatt et al., 2016).

This development may not only help infants to extract subtler information about people, but also understand how to interact with them. Recording preferential looking times, Farroni et al. (2004) found that newborns, like adults, prefer direct gaze than averted gaze in schematic faces. Adapting the stimuli to induce an eye-movement from direct gaze to averted gaze, authors further induced shifts in infants’ attention, an effect which they interpreted as a primitive form of gaze following. Investigating infants’ following of gaze, head, or body cues such as pointing, Butterworth & Itakura (2000) showed that infants not only extrapolate the directed acts of people toward a particular zone in space, but integrate these acts with their targets. Senju et al. (2008) further argued that infants understand actions directed toward objects as communicative acts. Finding that infants look longer at (i.e. visually prefer) faces gazing toward (vs. away from) objects only following direct eye-contact, authors proposed that infants’ processing of the relation between a directed act and its target object depends on the setting of a communicative context, by the ostentation of a communicative intention. These works show how basic visual processes enable infants to join attention with caregivers upon specific aspects of the world, not just by looking at the same objects by coincidence,

but by integrating the observed object and the caregivers' act. Doing so, these mechanisms may ground the development of core social abilities, such as word learning or mentalizing (Baron-Cohen, 1997; Carpenter et al., 1998; Slaughter & McConnell, 2003).

### **3.2. Representing third-party interactions in infancy**

Around the time Castelli et al. (2000) started investigating the neural representation of social interactions, as depicted in Heider & Simmel's (1944) animations, developmental researchers began studying infants' understanding of social relations between third-parties with similar displays. Premack & Premack (1997) first adapted a visual habituation paradigm to test 12-months-old infants with animations of geometrical shapes, engaging in either positive or negative interactions. In the habituation paradigm, infants are repeatedly presented with a type of stimulus until they habituate, as shown by a decrease in looking time. Next, they are presented with either the same or a new type of stimulus. Dishabituation to the new stimuli, as indicated by a significant increase in looking time after habituation, show that they can discriminate between the two types of stimuli. Premack & Premack (1997) habituated infants with interactions that were positive (caressing, helping) or negative (hitting, hindering), and then presented negative hitting events. They found that infants dishabituated more (looked longer at negative interactions) after habituating to positive interactions than after habituating to negative interactions, indicating that they interpreted the valence of the interactions. Later research replicated this result, further showing that young infants have a preference for prosocial individuals (Hamlin et al., 2007). Using such stimuli, research has progressively uncovered the sophisticated social inferences young infants make from observing social interactions, about thematic roles (Rochat et al., 2004; Tatone et al., 2015; 2021), affiliative relations (Fawcett & Tunçgenç, 2017; Liberman et al., 2014, 2021; Powell, 2021; Powell & Spelke, 2013, 2018a, 2018b; Rhodes et al., 2015; Spokes & Spelke, 2017), or hierarchies (Gazes et al., 2017; Mascaro et al., 2012; 2014; Pun et al., 2016; Thomsen et al., 2011).

These studies manipulated the same visual features that induce impressions of social intentions and relations in adults. Motion synchrony (Fawcett & Tunçgenç, 2017), for instance, which infants' as young as 4-month-old use to draw expectations about the affiliation of imitators with their models (Powell & Spelke, 2018a, 2018b), or the velocity with which shapes contact each other, that may induce different impressions of "gentleness" (Premack & Premack, 1997). Two studies illustrate how infants' expectations about social intentions may follow their ability to discriminate such perceptual cues (Rochat et al., 1997; 2004). In a first study, Rochat et al. (1997) tested 3- to 6-months-old infants' discrimination of a chasing event. In a preferential looking paradigm, infants viewed an animation depicting geometrical shapes moving as if one shape chased the other, simultaneously with another animation showing the two shapes moving independently. Analyzing looking times, they found that even the youngest infants could discriminate random movement from a chase. Interestingly, older infants (6 months) behaved

similarly to adults, looking longer at the non-social motion than the chase. In a second study, Rochat et al. (2004) questioned the extent to which this visual sensitivity reflected the representation of thematic roles, that is, whether infants represent a structured event where a chaser is put in relation to a chase. Using similar displays as Rochat et al. (1997) in a habituation-dishabituation paradigm, this study tested whether 3- to 10-months-old infants could dishabituate to role reversal, by exchanging the color of agents. Infants were habituated with animations depicting a blue disc as a chaser and a red disc as a chasee (or vice-versa), and dishabituated with test animation showing the ex-chaser as a chasee and the ex-chasee as a chaser. Eight- to 10-month-olds discriminated role reversal, indicating that they inferred roles in the animations. Both studies show how discriminating a spatio-temporal of social interaction such as the contingent movement of two shapes, may develop into representing social intentions in agents, suggesting how visual sensitivities to the structure of social interactions, like sensitivities to the structure of social entities, may help developing core socio-cognitive skills.

### **3.3. Perceiving social interactions: an effect of facingness?**

Infants' ability to discriminate the structure of faces, bodies, or biological motion can support core processes of social cognition during first-person interaction. Similarly, infants' understanding of third-party social relations may emerge from discriminating specific visual features. Do infants use facingness, the visual cue discussed above? One study by Augusti et al. (2010) suggests that young infants process conversations depending on the spatial relation of people. Authors presented 4-, 6- and 11-months-old infants with videos of people conversing face-to-face or back-to-back, while recording infants' gaze with eye-tracker. Measuring infants' gaze switches between speakers, authors found that infants as young as 6 months alternate more between people interacting face-to-face than people back-to-back. As this behavior did not occur when heads had the eyes closed, authors concluded that it did not result from mere gaze following. Consistent with this, in a habituation study, Beier & Spelke (2012) next showed that 10-months-old infants discriminate people orienting toward each other from people orienting away from each other, during a conversation. Thus, research suggests that infants are sensitive to the mutual positioning of agents in the rich context of a real conversation.

Galazka et al. (2014) investigated this sensitivity with a more controlled stimulation, recording 14-months-old infants' preferential looking times between point-light walkers moving toward each other, in meaningful interactions, or away from each other. Authors reported a visual preference for walkers moving toward each other when stimuli were upright, but not inverted, which ensured that infants did not discriminate stimuli simply by tracking differences in low-level visual features (e.g., points moving toward the center or the periphery of displays). This study found that the presence of an additional congruent cue, 2 voices talking to each other, strongly enhanced the visual preference for interacting stimuli. Interested in the emergence for this preference, Thiele et al. (2021a) tested 7- to 13.5-months-old infants' in a preferential looking times paradigm, showing people interacting face-to-face (videos displayed fully-fleshed actors) and people interacting back-to-back (actors were vertically swapped).

The authors observed a shift from no discrimination before 9 months, to a visual preference for face-to-face interactions at 10 months. These studies show that infants, like adults, discriminate and tend to prefer social interactions from visual information alone, using dynamic cues.

However, unlike research on adults, developmental research did not isolate the spatial cue of facingness. In above studies, infants' behavior could reflect the processing of dynamic features, or the recognition of meaningful social interactions, regardless of the spatial relation. Handl et al. (2013) started addressing this issue, presenting 9-month-old infants with still images of facing heads or non-facing heads without auditory cues. Similar to Augusti et al. (2010), infants shifted gaze more often between facing heads. These results thus support the hypothesis that facingness plays a role in infants' perception of social scenes already at 9 months of age. The current thesis (Chapters 2-3) aims at confirming this hypothesis and investigates whether, like in adults, the detection of facingness triggers a grouping mechanism and the inference of a social value.

## **4. Thesis overview**

### **4.1. Current research**

This thesis asks: how do we see relations between people? Specifically, it addresses how the visual system processes people in socially relevant spatial relations, focusing on the relation that here we call, *facingness*. I made the hypothesis that vision treats people facing each other as more than separate individuals: a structured whole that is one visual unit. Because such mechanism can facilitate the processing of a social scene, I made the hypothesis that it would be core to developing a representation of social relations, and thus, develop early in life.

This research unfolds in three chapters, addressing the following questions:

Chapter 1: Are face-to-face bodies integrated into a holistic perceptual unit?

Chapter 2: Do young infants process facing and non-facing bodies differently?

Chapter 3: When does the visual preference for face-to-face interactions develop, and what does it mean?

### **4.2. Overview of stimuli**

All experiments used identical stimuli. Stimuli were renderings of human bodies in different postures, randomly paired to form facing and non-facing dyads, with scant consideration of the meaningfulness of the interaction. The critical aspects of our stimuli is that facing and non-facing dyads always presented the exact same visual information, that is, the same bodies, except for the spatial relation between them. In sum, the spatial relation (facing /non-facing) was the only difference between the two main conditions of our studies and facingness was the only cue of social interaction, as stimuli were static and silent, providing no auditory or motion cues of interaction.

### 4.3. Overview of chapters

Chapter 1 asks whether facing bodies are represented differently than non-facing bodies in visual perception. In particular, building on previous studies (Adibpour et al., 2021; Papeo et al., 2017; 2019), we tested the hypothesis that facing bodies (but not non-facing bodies) are automatically grouped and, therefore, represented as a single perceptual unit. To test this, we used electroencephalography (EEG) in a dual frequency-tagging paradigm. Participants were shown two bodies, facing or non-facing, flickering at two different rates. To control for the specificity of effects to bodies, objects (chairs, machines) were presented as well, face-to-face or back-to-back, with the same flickering. Spectral analysis dissociated the responses to each individual body from the response to the dyad (intermodulation response). Results showed larger intermodulation response for facing bodies than for non-facing bodies or for facing and non-facing objects; concurrently, responses to individual bodies were weaker when the bodies were in a facing dyad compared to bodies in non-facing dyads. Thus, Chapter 1 suggests that facing people are grouped represented as a whole, and this mechanism affects the processing of individual bodies.

Chapter 2 tests the hypothesis that spatial relations between people are discriminated early on in infancy, anticipating the development of sophisticated social cognitive abilities. This was investigated measuring differential looking times in a preferential looking paradigm, on 6-month-olds. Experiment 1 shows that 6-month-olds discriminate facing bodies from non-facing bodies, looking longer at non-facing bodies. Comparing looking time at dyads (facing or non-facing) with looking time at single bodies, Experiment 2 shows that infants look as much at facing dyads and single bodies (and less than at non-facing bodies), as if they processed facing bodies as one single visual unit. Contrasting facing or non-facing dyads with non-reciprocal face-to-back dyads (suggesting a non-reciprocal interaction with one acting upon another), Experiment 3 shows that infants broadly discriminate spatial relations that cue interaction (whether reciprocal or non-reciprocal) from a spatial relation that implies unrelated bodies (back-to-back). Experiment 4 shows that the above effects are specific to human-human (social) dyads and do not generalize to human-object (non-social) pairs. Overall, Chapter 2 finds that infants as young as 6-months are sensitive to spatial relations between people, and, like adults, may process facing people more efficiently, as one visual unit.

Chapter 3 tests the hypothesis that a scene with facing people is spontaneously interpreted as a social interaction. This was investigated by measuring differential looking times between facing and non-facing bodies in human adults. In this paradigm, stimuli with a higher social value (facing dyads) are expected to be visually preferred (looked longer) over stimuli with a lower social value (non-facing dyads). This was tested in human adults (Experiment 1a) concurrently with social semantic evaluation (Experiment 1b), throughout human development (Experiment 2a: infancy at 7, 10, 15, 18 months; Experiment 2b: childhood at 3, 5 years) and in nonhuman social primates (Experiment 3: macaques). Experiment 1 finds that adults look longer at facing bodies (Experiment 1a), which they rate higher on social semantic dimensions (Experiment 1b) compared to non-facing dyads. Experiment

2 shows that this preference emerges at 5 years, following early stages in which infants discriminate between facing and non-facing dyads but do not show any preference for facing dyads, implying no appraisal of the social relevance of facingness. Experiment 3 finds that social primates such as macaques, like humans, prefer facing conspecifics. In sum, Chapter 3 indicates that by 5 years, children appreciate the social relevance of facingness between two social agents, and that this process is shared with nonhuman social primates.

This thesis argues that socially-relevant spatial relations alone, in static, speechless scenes, triggers special perceptual mechanisms, involving grouping. Face-to-face dyads, but not back-to-back dyads, are processed as a whole. This process, grouping facing bodies in units, facilitates visual processing – perhaps at the cost of processing individuals. Developing precociously, it helps preverbal infants to process people who look like interacting as visual units, which they discriminate from people who do not interact. During childhood, this precocious visual sensitivity develops into a visual preference, that may indicate a more complex process representing, in human and non-human social primates, relations between others. This process may be essential to understand social interactions and develop a social mind.

# Chapter 1 – Intermodulation responses show integration of interacting bodies in a new whole

*Are face-to-face bodies integrated into a holistic perceptual unit?*

This chapter reports and discusses the investigation of Goupil, Hochmann & Papeo (2023), published as a preprint on <https://osf.io/q9e8a/>.

## 1. Introduction

The human visual system is particularly attuned to social cues in the environment (Nakayama, 2010; New et al., 2007; Papeo, 2020). Going beyond the study of specialized processes for perception of social entities such as faces and bodies, recent studies have addressed the processing of social interactions (Abassi & Papeo, 2020; 2022; Bellot et al., 2021; Dima et al. 2022; Isik et al., 2017; Landsiedel et al., 2022; Quadflieg et al., 2015; Quadflieg & Koldewyn, 2017; Tarhan & Konkle, 2020; Walbrin et al., 2018; Wurm et al., 2017; Yang et al., 2015). Some of those studies have shown that, in visual areas specialized to person perception (i.e., face and body processing), perception of social interaction changes the visual representation of individual bodies and body movements (Abassi & Papeo, 2020; Bellot et al., 2021; Walbrin & Koldewyn, 2019). The representational change registered in visual areas may have a counterpart in behavioral phenomena showing that seemingly interacting bodies (e.g., face-to-face bodies) have faster access to visual attention and awareness, relative to the non-interacting (back-to-back) bodies (Papeo et al. 2017; Papeo et al., 2019; Vestner et al., 2019; Yin et al. 2018).

In the present study, we addressed a mechanistic explanation, whereby differences between face-to-face *versus* back-to-back bodies, at the level of neural representation and behavioral performance, would reflect the advantage of integrating two independent percepts (two bodies) into a unitary perceptual unit, or *new whole*, when represented as interacting.

We used a frequency-tagging electroencephalography (EEG) paradigm based on Steady-State Visually Evoked Potentials (SSVEP) to address this hypothesis. In female and male human adults, SSVEP were elicited by presenting two items on the right and left of central fixation, respectively, flickering at two different frequencies, during EEG recording. The periodic stimulation entrains the neuronal population that responds to that stimulation, to oscillate at a similar periodicity (Adrian & Matthews, 1934; Regan, 1966; Regan & Heron, 1969). This process is captured in the EEG signal, in the form of a periodic response at the stimulation frequency. Since the two bodies were presented at two different frequencies, we obtained two responses at the corresponding fundamental frequencies F1 and F2, and their harmonics. The neuronal populations that receive inputs from both stimulations, and in which those inputs interact non-linearly, may oscillate, can generate a response, at a frequency called intermodulation

frequency, corresponding to sum or difference of the individual terms ( $nF1 \pm mF2$ ). Because intermodulation frequencies do not necessarily overlap with fundamental and harmonic frequencies, they can allow distinguishing neural integration from the response to individual parts (here, single bodies; see Aissani et al., 2011; Alp et al., 2016; Appelbaum et al., 2008; Ratliff & Zemon, 1982; Victor & Conte, 2000; Zemon & Ratliff, 1984; Zhang et al., 2011).

Recent research has successfully adapted this method to study integration in visual processing of faces and biological motion (Adibpour et al., 2021; Alp et al., 2017; Boremanse et al. 2013; 2014). Neural integration was dissociated from part perception, showing different intermodulation response for parts that formed a unitary percept (e.g., a face) *versus* parts that did not form any meaningful configuration (e.g., misaligned face parts) (Boremanse et al. 2013, 2014). Using this approach, the current study sought to capture a similar dissociation between perception of multiple bodies and perception of multiple bodies in interaction. Dyads of bodies were presented in the visuo-spatial relation (proto)typical of a social interaction (face-to-face), or as unrelated (back-to-back). Individual bodies in a dyad flickered at two different frequencies (fundamental frequencies  $F1$  and  $F2$ ). We performed narrowband spectral analysis to distinguish the neural responses to individual bodies (at the fundamental and harmonic frequencies) from the neural signature of integration at the intermodulation frequency  $nF1 \pm mF2$ . The specificity of the effects to representation of social interaction was addressed by presenting objects with a clear anteroposterior morphology (i.e., chairs and machines), which, like bodies, could be presented face-to-face and back-to-back but, unlike bodies, yield no social interaction. In addressing the processing of body dyads with (face-to-face) or without (back-to-back) indication of relationship, the current study investigates how, from visual perception of individual bodies, a transformative visual process proceeds towards the representation of social interaction.

## 2. Materials & method

**Participants.** Thirty-two participants were recruited (20 females; age range: 18–34, *Mean* [ $M$ ] = 23, *Standard Deviation* [ $SD$ ] = 4) through social networks. Sample size was estimated from the size of the effect of positioning reported in a previous study using a similar experimental design (Adibpour et al., 2021), using the *ss.power.wa* function of the *BUCSS* package (Anderson & Kelley, 2020) in *R* ( $F = 5.5$ ,  $N = 20$ ,  $\alpha_{prior} = 1$ ,  $\alpha_{planned} = 0.05$ , power = 0.80, assurance = 50%). Participants reported no history of neurological or psychiatric disorders, taking no medication, being right-handed and possessing a normal (or corrected-to-normal) vision. The present experiment was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from participants before any assessment or data collection. All procedures were approved by the local ethics committee (CPP sud-est II), and conducted at the Institut des Sciences Cognitives Marc Jeannerod. Participants were paid 10€ for participating in a 30 minutes long experiment, for reimbursement of travel expenses.

**Stimuli.** Gray-scale images of profile views on a human body in four different postures (width:  $M = 1.7^\circ$  of visual angle,  $SD = 0.3^\circ$ ; height:  $M = 4.6^\circ$ ,  $SD = 0.2^\circ$ ) were created with *Daz3D* (Daz Productions, Salt Lake City, UT) and the *Image Processing Toolbox* of *MATLAB 2015b* (MathWorks Inc., Natick, MA). Postures were biomechanically possible, with comparable amount of visual information (pixels) in the anterior and posterior halves of the body. Six different dyads were formed by randomly pairing two different body postures (hereafter, bodies) face-to-face. For each face-to-face dyad, a version with the same bodies back-to-back was created. In each trial, the display showed a dyad with one individual on the left side and the other on the right side of a central fixation cross (Figure 1A). The fixation cross (length =  $0.3^\circ$  of visual angle; span =  $0.06^\circ$ ) was located horizontally at the center of the screen, vertically at the level of the upper part of bodies,  $1.3^\circ$  of visual angle from the screen center (visual angles were computed considering a distance of 100 cm from the screen). The cross was black, but turned red for short intervals during the experiment (see *Procedures*). The two bodies in a dyad were equally distant from the center of the screen ( $M = 0.5^\circ$ ,  $SD = 0.2$ ), and were equally distant from each other in face-to-face and back-to-back dyads ( $M = 1.1^\circ$ ,  $SD = 0.2$ ). Dyads had a width of  $M = 4.4^\circ$ ,  $SD = 0.3$  and height of  $M = 4.7^\circ$ ,  $SD = 0.1$ .

For the object conditions, twelve pairs of objects were created randomly coupling two of four different renderings of machines (pairs of machines:  $N = 6$ ; width:  $M = 4.2^\circ$ ,  $SD = 0.2$ ; height:  $M = 4.7^\circ$ ,  $SD = 0.02$ ; Figure 1B) or two of four different renderings of chairs (pairs of chairs:  $N = 6$ ; width:  $M = 4^\circ$ ,  $SD = 0.3$ ; height:  $M = 2.6^\circ$ ,  $SD = 0.02$ ; Figure 1C). The two items in all pairs were equally distant from the center of the screen ( $M = 0.5^\circ$ ,  $SD = 0.1$ ), and were equally distant from each other in face-to-face and back-to-back pairs ( $M = 0.9^\circ$ ,  $SD = 0.1$ ). Fixation cross was set at the high corresponding to the objects' most prominent anterior part ( $0.3^\circ$  from the screen center).

**Procedure.** The experiment consisted in 16 trials, 8 showing body dyads and 8 showing pairs of objects. Two different body dyads, one pair of chairs and one pair of machines were shown to each participant. Each body dyad was shown in 4 trials: 2 trials face-to-face and 2 trials back-to-back, counterbalancing the flickering rate of the two items (F1 and F2). Each object pair was shown in 2 trials, 1 face-to-face and 1 back-to-back, counterbalancing the flickering rate of the two items (F1 and F2). Body- and object-trials were presented in separate blocks, counterbalancing the order of body- and object-blocks across participants. Within object blocks, the order of chairs and machine trials was also counterbalanced across participant. Each block was composed of two face-to-face and two back-to-back trials, counterbalancing the side of F1 and F2. The order of presentation of the four first trials was randomized across participants. The next four trials were presented in reversed order. Each trial lasted 80 seconds (s), starting with a buffer period of 10 s, during which a central fixation cross was displayed. Individual stimuli flickered for the remaining 70 s of the trial. During this interval, the cross changed color from black to red for 200 ms, six to eight times. We opted for a relatively long stimulation duration as it enhances resolution in the frequency domain, concentrating the tagged responses in narrower bands, and reducing biological noise (Alonso-Prieto et al., 2013; Liu-Shuang et al., 2014; Regan & Regan, 1989).

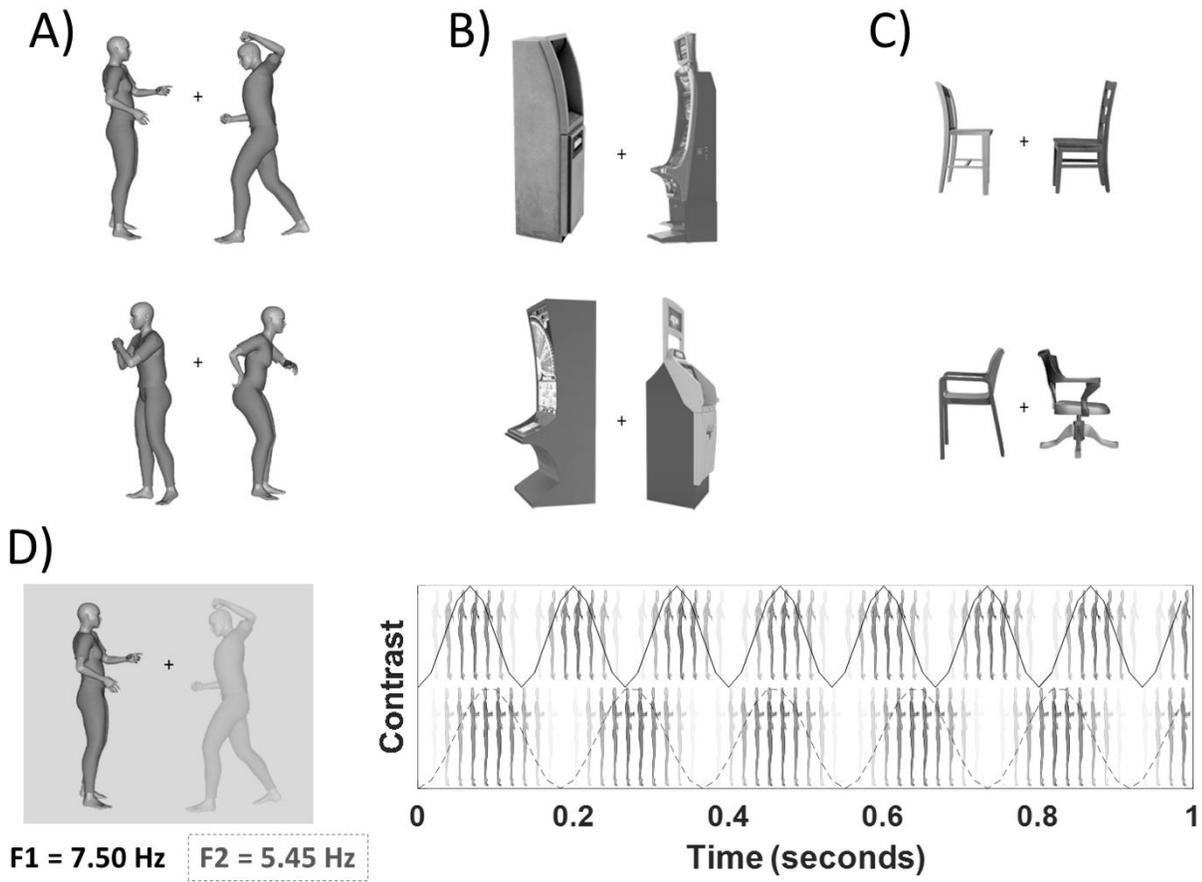
Participants sat on a chair, 100 cm away from the computer screen where stimuli were presented. The height and the position of the chair was adjusted to have the participant's eyes were aligned with the center of the screen. After the EEG preparation phase, participants were instructed to avoid movements during periods of recording, to fixate the cross and report the color change as soon as they saw it, by pressing with their right index finger, the left button of a mouse they held in their hands throughout the experiment. This task was included to maintain participants' attention constant throughout trials and conditions, while keeping the experimental manipulation implicit (Boremanse et al., 2013, 2014). Performance in this task did not differ ( $P_s > .05$ ) between conditions in terms of accuracy ( $M = 98\%$ ,  $SD = 0.03$ ) or reaction times ( $M = 397$  ms,  $SD = 43$ ).

Visual stimulation during EEG recordings was performed by *SinStim*, an application running on *MATLAB*, designed for frequency-tagging (Alonso-Prieto et al., 2013; Alp et al., 2016; Boremanse et al., 2013, 2014). Left and right items flickered at two different rates (F1 and F2), with a sinusoidal contrast modulation ranging from 0 (no contrast) to 1 (full contrast) (Figure 1D). Flickering rates were chosen after previous studies (Alp et al., 2017; Vergeer et al., 2018) and complied to several constraints. First, frequencies located in the *Theta* band avoid overlap with the higher noise level of the *Alpha* band, and maximize the response of the visual system to the complex stimuli presented (Alonso-Prieto et al., 2013; Norcia et al., 2015; Zemon & Ratliff, 1984). Second, the two frequencies were temporally close to obtain comparable SSVEP characteristics (Alonso-Prieto et al., 2013; Norcia et al., 2015), but separated by at least 44 frequency bins ( $\sim 0.6$  Hz) so as to avoid overlap in computing Signal-to-Noise Ratios (see below). Third, harmonics of stimulation rates did not overlap with one another, and with the intermodulation frequencies, as much as possible (see below). Finally, to ensure exact presentation rates, frequencies were chosen so that the periodic cycles corresponded to a fixed number of frames for a screen refreshing rate of 60 Hz: F1 = 7.50 Hz (cycle of 8 frames), F2 = 5.45 Hz (cycle of 11 frames). The actual refreshing rate, recorded during testing, was  $\sim 59.95$  Hz on average, slightly shifting F1 to 7.49 Hz. A photodiode, located on the right lower corner of the monitor, recorded trial onset/end, and monitored timing accuracy throughout the experimental session.

Brain activity was recorded using a 128-electrode EEG net (EGI, Eugene, USA), with reference on the vertex. Impedance was kept below 50 k $\Omega$ . Recordings were continuously digitized at a sampling rate of 1000 Hz (net amp 400 system EGI, Eugene, USA).

**Figure 1.**

*Stimuli and procedure.*



*Note.* Stimuli: face-to-face and back-to-back bodies (A), and face-to-face and back-to-back objects (50% machines (B) and 50% chairs(C)). All four individual exemplars for each category are shown. D) Illustration of periodic stimulation. The image on the left illustrates a stimulus-display. On a gray background, individual items flickered at two different rates ( $F1$  &  $F2$ ). The fixation cross was black most of the time, except for rapid changes of color from black to red. The image on the right illustrates 1 second of stimulation: Sinusoidal contrast modulation, from 0 (transparent) to 1 (full contrast) is plotted for  $F1$  on the upper row, and for  $F2$  on the lower row. The corresponding appearance of individual stimuli at each frame is shown in background. Note that the sinusoidal contrast modulation at 60 Hz was perceivable as continuous.

## Analysis

Signal processing was performed with *MNE python* (Gramfort et al., 2013) and handmade *MATLAB* scripts. Statistical analysis was performed in *R* 4.0.2 (R Core Team, 2020) using *ez* 4.4.0 (Lawrence, 2016) for computing ANOVAs. Scripts and data will be made available in a public repository.

**Preprocessing.** Trials were truncated over the stimulation period during which items flickered (10 to 80 s). Amplitude was normalized by subtracting from each electrode its mean amplitude. Trials were cropped to include a constant number of cycles of stimulation frequencies ( $F1 = 517$ ,  $F2 = 376$ ; duration = 68989 ms).

Trials of the same condition, as defined by stimulus category (bodies or objects) and configuration (face-to-face, back-to-back), were averaged in the time domain, to reduce EEG oscillatory activity that was not phase-locked to the stimulation (Boremanse et al., 2013, 2014). Trials were collapsed regardless of side (left item flickering at F1 or *vice versa*).

Signal was decomposed using the Fast Fourier Transform algorithm of Matlab, with a frequency resolution of 0.0145 Hz. Frequencies above the electrical noise (50 Hz) were excluded from further analysis. To take into account the 1/f noise, the Signal-to-Noise Ratio (SNR) was computed: the amplitude at each frequency was divided by the average amplitude of its 40 neighboring frequencies (20 on each side,  $\sim 0.3$ Hz). Immediately adjacent bins (2 on each side) were not included to reduce potential effects of spectral leakage (Yan et al., 2019); finally, bins with the highest amplitude (and lowest, for balance) were discarded to avoid including signal (e.g., at F1 and F2) in the estimation of noise (Beck et al., 2018).

**SNR for individual items (IND-SNR).** Flickering of individual items was expected to increase SNR at the corresponding fundamental frequencies ( $F1 = 7.49$  Hz and  $F2 = 5.45$  Hz) and their harmonics (up to  $6F1$  and  $9F2$  below 50 Hz). Each individual term ( $nF1$  or  $mF2$ ) was tested by comparing SNR (averaged over all electrodes and trials by participant) at the corresponding frequency with the chance level (1), using one-sided Wilcoxon signed-rank tests (Bonferroni corrected  $\alpha = 0.0033$  corresponding to 0.05/15) (Retter et al., 2021; Rossion et al. 2020). Next, for each electrode, the response to the flickering of individual bodies (IND-SNR) was measured averaging SNR across terms showing significant signal. The effect of Configuration was computed as the difference between the IND-SNR for face-to-face stimuli and the IND-SNR for back-to-back stimuli. The effect of Category was measured as the difference between the IND-SNR for Bodies and the IND-SNR for Objects. The Category by Configuration interaction was computed as the difference between the effect of Configuration for Objects and the effect of Configuration for Bodies. Significance was assessed with Wilcoxon signed-rank tests against chance (0). The spatial distribution of effects and interaction over the scalp was determined by testing each individual electrode. To correct for multiple comparisons, clusters of at least three adjacent electrodes showing significant effects ( $\alpha = .05$ ) were submitted to cluster-mass permutation analyses (Maris & Oostenveld, 2007). In significant clusters,

the IND-SNR was averaged across electrodes. In case of interaction, the effect of Configuration was computed and tested separately for Bodies and Objects.

***SNR for the intermodulation response (IM-SNR).*** Because the mutual terms ( $nF1 \pm mF2$ ) at which intermodulation will arise are not easy to predict (Appelbaum et al., 2008; Gordon et al., 2019; Regan & Regan, 1988, 1989), we considered any possible combination. We first considered second order mutual terms ( $F1 + F2$ ,  $F1 - F2$ ), since intermodulation has primarily been reported at these frequencies (Adibpour et al., 2021; Alp et al., 2017; Boremanse et al. 2013, 2014; Mersad & Caristan, 2021). Next, all mutual terms were considered up to 50 Hz, excluding redundant terms (e.g.,  $4f1 - 5f2 = 6f2 - 4f1 = 2.73$  Hz), yielding 56 different intermodulation frequencies (Table 1). Intermodulation (IM-SNR) was first localized over the scalp, by averaging SNR over all frequencies, by condition, and by electrode. The effect of Configuration was computed as the difference between the IM-SNR for face-to-face stimuli *vs.* the IM-SNR for back-to-back stimuli. The effect of Category was computed as the difference between the IM-SNR for Bodies *vs.* the IM-SNR for Objects. The interaction was computed as the difference between the Configuration effect with Bodies *vs.* Objects. Effects were localized computing Wilcoxon signed-rank tests against chance (0) at each electrode, using cluster-mass permutation testing to correct for multiple comparisons.

In order to determine the frequencies at which intermodulation was the strongest, the IM-SNR was averaged by conditions and frequency bands ( $0.3 - 3.5 =$  Delta,  $3.5 - 8 =$  Theta,  $8 - 13 =$  Alpha,  $13 - 30 =$  Beta,  $30 - 50 =$  Gamma), at the electrodes showing a significant effect. Repeated-measures ANOVA was computed with Category, Configuration and Frequency Band as factors. The third-order interaction was inspected by computing repeated-measures ANOVAs with Category and Configuration as factors, separately for each frequency band. In bands showing an interaction, Wilcoxon signed-rank tests evaluated the effect of Configuration for Bodies and Objects, separately.

**Table 1.***Intermodulation frequencies.*

Delta		Theta		Alpha		Beta		Gamma	
Term	Hz	Term	Hz	Term	Hz	Term	Hz	Term	Hz
3F1 - 4F2	0.68	2F1 - 2F2	4.09	4F1 - 4F2	8.18	4F1 - 3F2	13.63	7F2 - 1F1	30.66
3F2 - 2F1	1.36	5F2 - 3F1	4.77	3F2 - 1F1	8.86	4F2 - 1F1	14.31	2F1 + 3F2	31.34
1F1 - 1F2	2.04	3F1 - 3F2	6.13	2F1 - 1F2	9.54	5F1 - 4F2	15.67	5F1 - 1F2	32.02
4F1 - 5F2	2.73	4F2 - 2F1	6.81	6F2 - 3F1	10.22	3F1 - 1F2	17.03	3F1 + 2F2	33.38
2F2 - 1F1	3.41			3F1 - 2F2	11.58	6F2 - 2F1	17.71	6F1 - 2F2	34.06
				5F2 - 2F1	12.26	1F1 + 2F2	18.39	1F1 + 5F2	34.74
				1F1 + 1F2	12.94	4F1 - 2F2	19.08	4F1 + 1F2	35.43
						5F2 - 1F1	19.76	8F2 - 1F1	36.11
						2F1 + 1F2	20.44	2F1 + 4F2	36.79
						5F1 - 3F2	21.12	3F1 + 3F2	38.83
						7F2 - 2F1	23.16	6F1 - 1F2	39.51
						1F1 + 3F2	23.84	1F1 + 6F2	40.19
						4F1 - 1F2	24.53	4F1 + 2F2	40.88
						6F2 - 1F1	25.21	9F2 - 1F10	41.56
						2F1 + 2F2	25.89	2F1 + 5F2	42.24
						5F1 - 2F2	26.57	5F1 + 1F2	42.92
						3F1 + 1F2	27.93	3F1 + 4F2	44.28
						6F1 - 3F2	28.61	1F1 + 7F2	45.64
						1F1 + 4F2	29.29	4F1 + 3F2	46.33
								2F1 + 6F2	47.69
								5F1 + 2F2	48.37

*Note.* Term shows the relation of the frequency with stimulation rates. Hz is the frequency in Hertz.

### 3. Results

**Responses to individual bodies.** For both F1 and F2, significant responses were found at the fundamental frequency and up to the fourth harmonic (Table 2, Figure 2A). There was an effect of Category ( $p = .0004$ ) at posterior electrodes (Figure 2B, right), where IND-SNR was significantly higher for bodies ( $M = 1.74$ ,  $SD = 0.55$ ) than objects ( $M = 1.50$ ,  $SD = 0.46$ ). A significant interaction between Category and Configuration was found ( $p = .0068$ ) with overlapping distribution (73% overlap) (Figure 2B, left). In that cluster (Figure 2C), the IND-SNR was significantly higher for single bodies in back-to-back trials ( $M = 2.14$ ,  $SD = 0.83$ ) than for single bodies in face-to-face trials ( $M = 1.88$ ,  $SD = 0.69$ ;  $W = 54$ ,  $z = -3.93$ ,  $p < .001$ ), while there was no effect of Configuration for objects (face-to-face:  $M = 1.77$ ,  $SD = 0.70$ ; back-to-back:  $M = 1.72$ ,  $SD = 0.64$ ;  $W = 311$ ,  $z = 0.88$ ,  $p = .380$ ).

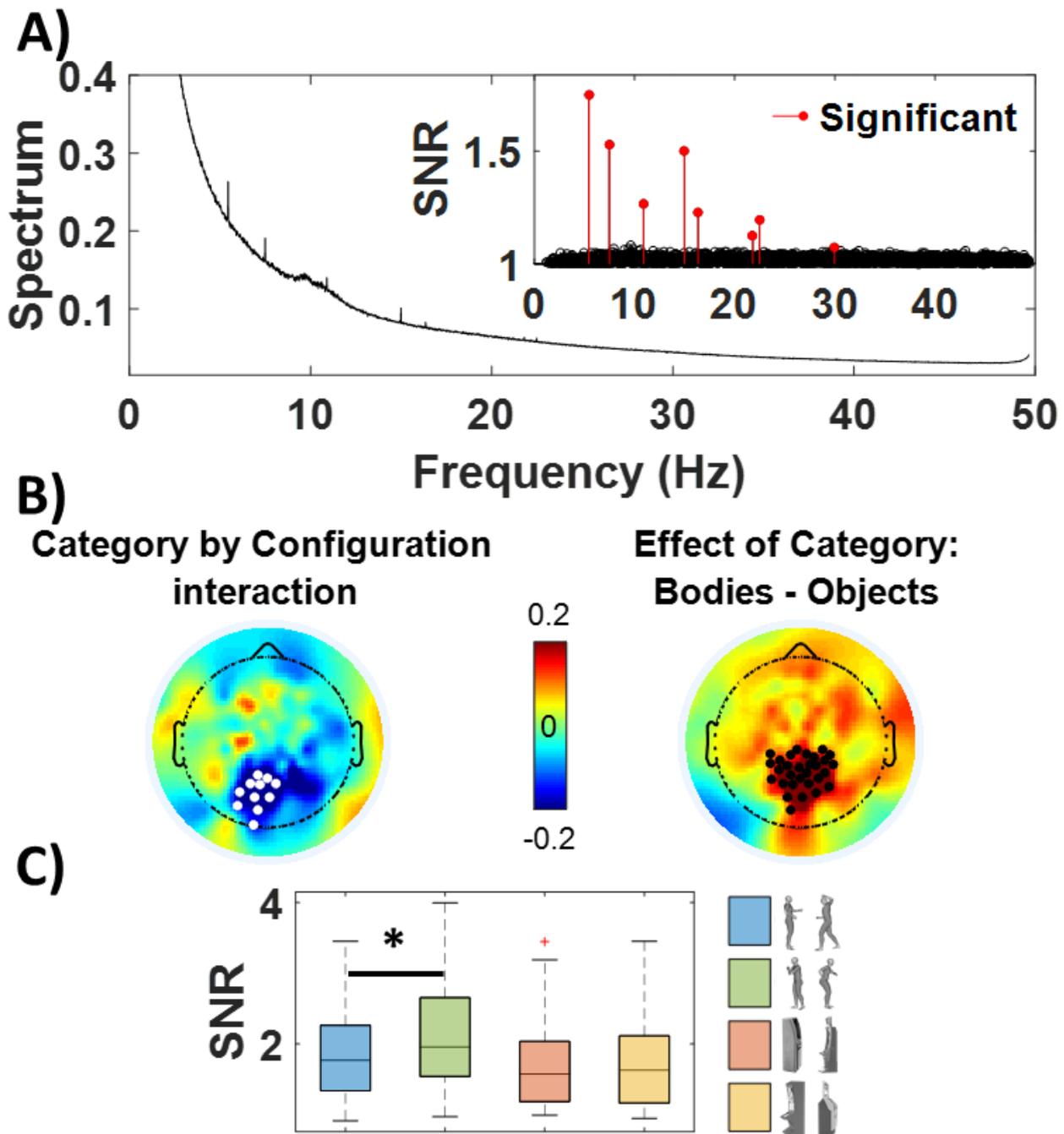
**Table 2.***Signal-to-Noise Ratio of the response to the individual stimuli.*

Term	Hz	SNR	<i>W</i>	<i>z</i> -value	<i>p</i> -value	Inclusion
1F1	7.494	1.530	510	4.591	< .00001	1
2F1	14.988	1.501	528	4.927	< .000001	1
3F1	22.482	1.194	500	4.404	< .00001	1
4F1	29.976	1.073	433	3.151	< .001	1
5F1	37.470	1.025	349	1.580	.057	0
6F1	44.964	1.042	381	2.178	.015	0
1F2	5.450	1.751	525	4.871	< .000001	1
2F2	10.900	1.266	503	4.460	< .00001	1
3F2	16.350	1.229	493	4.273	< .0001	1
4F2	21.800	1.124	475	3.936	< .0001	1
5F2	27.251	1.029	365	1.879	.030	0
6F2	32.701	1.028	367	1.917	.028	0
7F2	38.151	1.032	349	1.580	.057	0
8F2	43.601	1.020	373	2.029	.021	0
9F2	49.051	0.996	210	-1.019	.846	0

*Note.* Term shows the relation of the frequency with stimulation rates. Hz indicates the corresponding frequency in Hertz. SNR is the Signal-to-Noise Ratio averaged across participants. Next columns provide the statistics of the one-tailed Wilcoxon Signed-rank tests comparing the SNR averaged by participant against chance (1). Inclusion indicate whether the frequency was included in the IND-SNR ( $\alpha = .0033$ ).

Figure 2.

Effects of Category and Configuration on the response to individual bodies.



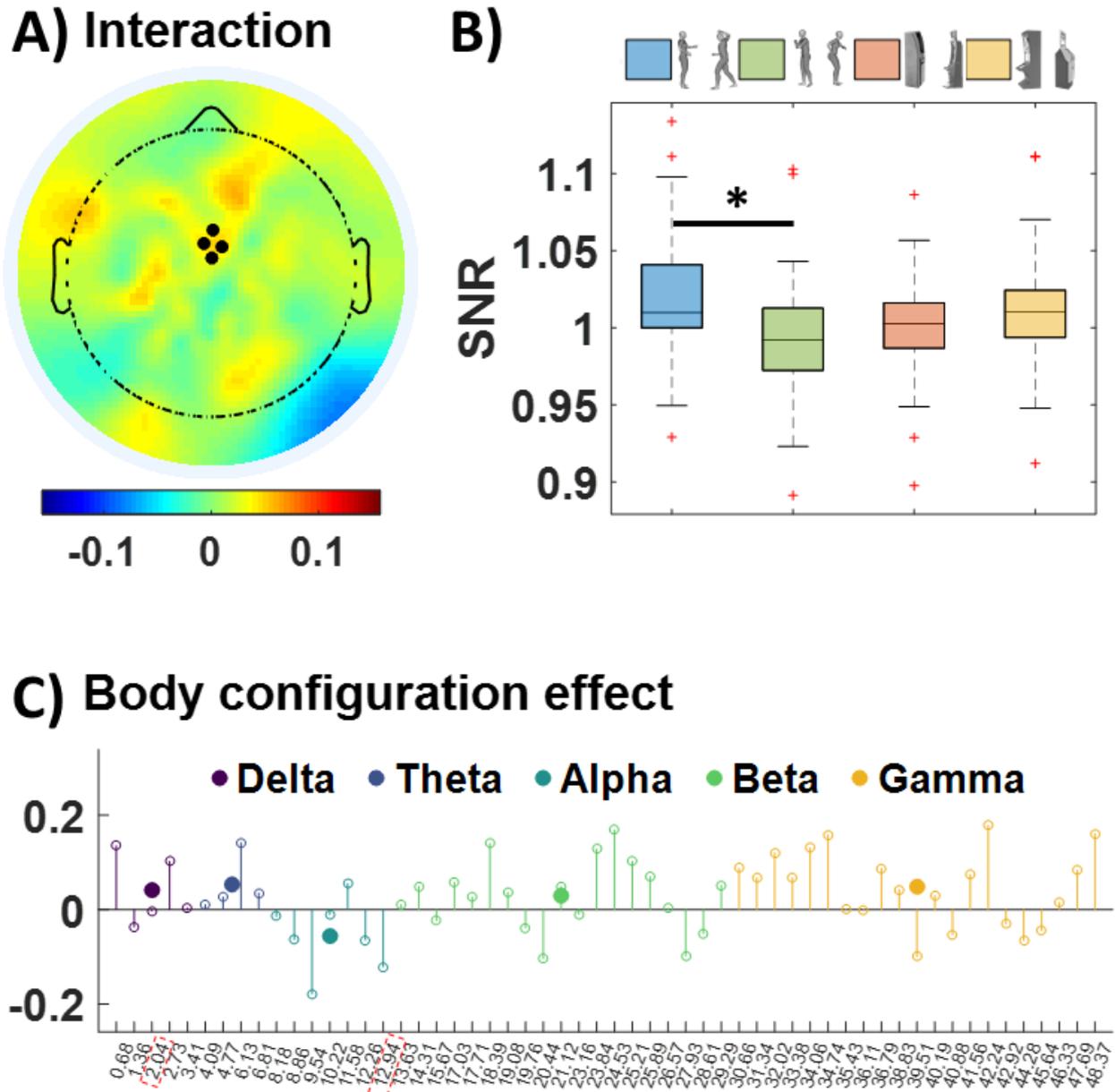
Note. A) Grand average of the frequency spectrum included in analysis (0–50 Hz, x axis) and the corresponding Signal-to-Noise Ratio (inset). Significant SNR is highlighted in red. B) Left: Scalp topography for the interaction between configuration and category: higher IND-SNR for single bodies in back-to-back trials than for single bodies in face-to-face trials but no difference for single object; Right: Scalp topography for the main effect of category: higher IND-SNR for bodies vs. objects. Positive differences are in warm colors, negative differences in cold colors (red – blue hues in the middle). Electrodes showing significant effects are shown with dots (in black or white). C) Boxplot of the IND-SNR for each condition, averaged over the electrodes in the cluster highlighted in B (left). Conditions are color-coded: blue = face-to-face bodies, green = back-to-back bodies, orange = face-to-face objects, yellow = back-to-back objects.

**Intermodulation response.** No cluster was found for the main effects of Category or Configuration or for the interaction of Category and Configuration, considering the second order mutual terms F1-F2 and F1+F2 (Figure 3C). Analyzing intermodulation responses averaging all 56 possible intermodulation frequencies, we found no main effect of Category or Configuration but a significant interaction ( $p = .0388$ ) in an anterior cluster (Figure 3A). There (Figure 3B), the IM-SNR was higher for bodies face-to-face ( $M = 1.02$ ,  $SD = 0.04$ ) than for bodies back-to-back ( $M = .99$ ,  $SD = 0.04$ ;  $W = 395$ ,  $z = 2.45$ ,  $p = .014$ ), while no difference was found between face-to-face ( $M = 1.00$ ,  $SD = 0.04$ ) and back-to-back objects ( $M = 1.01$ ,  $SD = 0.04$ ;  $W = 164$ ,  $z = -1.87$ ,  $p = .061$ ). We then tested the IM-SNR for each condition against the surrounding noise. Considering an  $\alpha$  level of .0125 to correct for multiple comparisons, the effect was only significant for bodies face-to-face ( $W = 414$ ,  $z = 2.80$ ,  $p = .005$ ), but not for bodies back-to-back ( $W = 186$ ,  $z = -1.46$ ,  $p = .145$ ), objects face-to-face ( $W = 283$ ,  $z = 0.355$ ,  $p = .722$ ) or objects back-to-back ( $W = 371$ ,  $z = 2.00$ ,  $p = .045$ ).

Next, we investigated whether the interaction between Category and Configuration on intermodulation responses was carried by specific frequency ranges (Figure 3C). IM-SNR in the identified cluster was averaged within each of five frequency bands (Delta, Theta, Alpha, Beta, Gamma). A repeated-measures ANOVA with Category, Configuration and Frequency Band as factors yielded a main effect of Frequency Band  $F(3.17,98.26) = 3.00$ ,  $p = .032$ , an interaction between Category and Configuration  $F(1,31) = 10.12$ ,  $p = .003$ , and an interaction between Category, Configuration and Frequency Band  $F(3.01,93.20) = 3.04$ ,  $p = .033$ . To follow up on the three-way interaction, a repeated-measures ANOVA with Category and Configuration as factors was run separately for each frequency band. Results showed a significant interaction between Category and Configuration ( $\alpha = .01$ , corrected for multiple comparisons) in the Gamma band,  $F(1,31) = 7.73$ ,  $p = .009$ , but not in other bands ( $Ps > .012$ ). In the Gamma band, the effect of Configuration (stronger response to face-to-face than back-to-back) was significant for Bodies  $W = 416$ ,  $z = 2.84$ ,  $p = .004$ , but not for Objects  $W = 277$ ,  $z = -0.24$ ,  $p = .808$ .

**Figure 3.**

*Effects of Category and Configuration on intermodulation frequencies.*



*Note.* A) Scalp topography of the Configuration by Category interaction on IM-SNR. Black dots represent electrodes at which the interaction was significant. B) Boxplot of the IM-SNR for each condition, averaged over the electrodes in the cluster identified in A. Conditions are color-coded: blue = face-to-face bodies, green = back-to-back bodies, orange = face-to-face objects, yellow = back-to-back objects. C) Effect of body configuration (y axis) by intermodulation frequency (x axis) across frequency bands. Colors range from the slowest (darkest) to the fastest (lighter) band. Second order intermodulation frequencies (F1+F2, F1-F2) are squared in red. Big solid dots represent band averages.

## 4. Discussion

We used frequency tagging EEG to measure the neural response to each of two bodies in a visual scene, and the corresponding intermodulation response, a neural correlate of the non-linear combination of the responses to individual bodies, taken as a marker of integration. Results showed that the intermodulation response was larger for face-to-face –seemingly interacting– bodies relative to the same bodies presented back-to-back, a difference that we did not observe for object pairs. Moreover, we found that the response to individual bodies was larger for bodies presented in back-to-back trials than for bodies in face-to-face trials, a difference that was not observed for non-body objects.

Both effects, the difference between single bodies in face-to-face *versus* back-to-back trials (at fundamental frequencies) and the difference between face-to-face and back-to-back dyads (at intermodulation frequencies), concur to demonstrate that spatial relations between bodies changed the neural response to those stimuli. In particular, we addressed the *a priori* hypothesis that perceiving two bodies tied in a relationship promotes integration of individual representations into a structured unitary percept. Supporting this hypothesis, we captured the transformative process towards an event representation with the intermodulation response in the EEG signal, which was larger for face-to-face than for back-to-back bodies.

### Response to face-to-face *versus* back-to-back bodies

Intermodulation is thought to reflect the activity of neurons that receive and integrate, in a non-linear fashion, signals from different inputs (Ratliff & Zemon, 1982; Regan & Regan, 1988; Victor & Conte, 2000). In the case of multipart stimuli, such as faces or point-light walkers, intermodulation has been proposed to reflect the formation of a new integrated percept, or *whole* (Boremanse et al. 2013; Alp et al., 2017). By extension, we interpret the larger intermodulation response for face-to-face than back-to-back bodies, as reflecting the rise of a new visual representation, integrating two seemingly interacting bodies. In forwarding this interpretation, we can exclude two alternative interpretations. First, the effect cannot be explained by the responses to individuals, as, if anything, those responses were larger for single bodies in back-to-back than in face-to-face dyads. Second, the effect cannot be interpreted as an unspecific correlate of face-to-face *versus* back-to-back stimulus' orientation, as we did not find a similar effect with non-body objects having a clear anteroposterior morphology. Thus, we propose that the larger intermodulation response for face-to-face (*vs.* back-to-back) bodies reported here reflects the effect of perceiving multiple bodies (a dyad, at least) in a relationship.

We note that the intermodulation signal reported here was smaller than in other studies (Alp et al., 2017; Boremanse et al., 2013, 2014; Mersad & Caristan, 2021). Differences at the level of stimuli and task could explain this circumstance. First, our greyscale static bodies flickering on a grey background with a sinusoidal contrast modulation, could be visually less salient than the colorful high-contrast or dynamic stimuli used in other studies (Alp et al., 2017; Boremanse et al., 2013, 2014). Second, it is possible that integration at the level of a scene (across objects) is fainter than integration at the level of an object (across parts): that is, the integration of parts in a face (Boremanse et al., 2013) or in a Gestalt percept (Alp et al., 2016) could be more compulsory than the

integration of objects in a scene, or in an event (see also Kaiser et al., 2019). The physical separation of bodies, as opposed to the continuation of parts (real in Boremanse et al., 2013; apparent in Alp et al., 2016) may have further contributed to reduce intermodulation in our study.

Two other studies have used a similar approach to ours, to tag integration of multiple objects, yielding intermodulation effects for human silhouettes (Mersad & Caristan, 2021) or bodies (Adibpour et al., 2021). The former study (Mersad & Caristan, 2021) found larger intermodulation response for two silhouettes next to each other than for two chairs next to each other; the latter (Adibpour et al., 2021), consistent with our study, reported larger intermodulation response for face-to-face (*vs.* back-to-back) bodies, primarily in an anterior cluster compatible with the current effect. Both studies supported the view that the representation of an interaction (arguably stronger in the case of silhouettes *vs.* chairs, and in the case of face-to-face *vs.* back-to-back bodies) elicits stronger integration of items in a scene.

At the same time, despite the remarkable similarity of stimuli and design, the intermodulation response in Adibpour et al. was larger than in our current study. The differences between Adibpour et al. (2021) and the current study deserve consideration as they can explain not only the difference in the intermodulation amplitude, but also how our study replicates and goes beyond the previous one. In Adibpour et al., the individual object frequencies (F1 and F2) were different from ours. This change might explain on its own the difference in the amplitude of both individual and intermodulation signals, as it is known that SSVEP amplitude changes across frequencies (Alonso-Prieto et al., 2013), and individual responses of different amplitudes alter the intermodulation signal (Regan & Regan, 1988). Additionally, and perhaps more importantly, in Adibpour et al., the intermodulation frequency F1 - F2 overlapped with the second harmonic of a third periodic visual stimulation, corresponding to the repeated synchronous presentation of the two bodies in full contrast. Thus, in Adibpour et al., it remained unknown how much the response to face-to-face bodies reflected the direct visual processing of two face-to-face bodies, and how much, the integration of the two individual bodies. The current study was designed to disentangle the contribution of the regular simultaneous full-contrast appearance of both bodies and intermodulation frequencies. Decreasing the contribution of simultaneous full-contrast appearance may have produced a weaker response in the current study, compared to Adibpour et al. (2021). In doing so, however, the current study overcomes the ambiguity in Adibpour et al., showing that face-to-face bodies give rise to larger integrative responses than back-to-back bodies.

Moreover, unlike Adibpour et al. (2021), here we included the test of object pairs, providing demonstration that the difference between face-to-face and back-to-back body dyads has to do with the representation of social relationship, rather than with a non-specific effect of spatial orientation. We do not exclude that other object pairs (e.g., a hammer towards a nail) that evoke interaction more strongly than two face-to-face chairs could yield as much integration as two face-to-face bodies (Kaiser et al., 2014; Green & Hummel, 2006; Roberts & Humphreys, 2010; Vestner et al., 2020). Future research will clarify whether the current effect of *facingness* is specific to the social-object domain or generalizes to other pairs of related items.

## Response to single bodies

We found a weaker response to individual bodies appearing in face-to-face *versus* back-to-back trials. This result was unpredicted, but not surprising. In effect, we have observed a similar phenomenon in a pilot study (unpublished data, available from the authors). Moreover, in the context of grouping, it has repeatedly been shown that the perception of a grouped representation modifies the responses to the constituent parts, making the parts less accessible, or accessible only after the global configuration (Poljac et al., 2012; Suzuki & Cavanagh, 1995). For example, Suzuki and Cavanagh (1995) found that, in visual search, the search for a whole face was more efficient (i.e., faster) than the search for a face part (e.g., the mouth), suggesting that vision gives priority to the highest level of representation available, to the detriment of local features. A similar effect has been reported for body dyads: Papeo et al. (2019; Experiment 5) showed that, in visual search through a crowded array, subjects were more likely to detect a single body (e.g., a body in a punching posture) when embedded in a non-interacting dyad than when presented in an interacting dyad. Thus, a poorer/delayed accessibility of the constituent parts, for the benefit of the global representation, appears to be a consistent feature of grouping, which we captured in the response to single bodies<sup>1</sup>.

## Conclusions

We found a larger intermodulation response to face-to-face (*vs.* back-to-back) bodies, suggesting that the visual system integrates seemingly interacting bodies in a non-linear fashion, giving rise to a representation that is different from the sum of the parts. Our results also showed a reduced response to single bodies in face-to-face trials, which we take to reflect reduced accessibility to parts of a grouped representation. In this sense, the two effects would reflect the two sides of the same coin, mirroring signatures of grouping reported in face perception (Purcell & Stewart, 1988; Suzuki & Cavanagh, 1995). The grouped representation of interacting bodies may account for the advantage of those stimuli, in terms of access to visual awareness (Papeo et al., 2017; Papeo & Abassi, 2019), recruitment of attention (Papeo et al., 2019; Vestner et al., 2019; 2020; 2021), representation in working memory (Ding et al., 2017; Paparella & Papeo, 2022; Stahl & Feigenson, 2014; Vestner et al., 2019), and neural representation (Abassi & Papeo, 2020; Bellot et al., 2021; Walbrin & Koldewyn, 2020). Mediated by grouping, the new representation that results from the perception of face-to-face bodies may constitute the earliest rudimentary representation of a social interaction.

---

<sup>1</sup> To prevent confusion, while there is evidence for reduced access to parts of a grouped representation, there is also evidence that, once accessed, parts of a meaningful configuration are represented and remembered better than parts in meaningless, scattered or scrambled configurations (Abassi & Papeo, 2020; Bellot et al., 2021; Ding et al., 2017; Heilbron et al., 2020; Kaiser et al., 2019; Purcell & Stewart, 1988). Rather than a conflict, the broader literature suggests a different effect of the visual context on access *versus* representation of parts: parts of grouped representations would be accessed later but encoded and retained better.

## Chapter 2 – Visual perception grounding of social cognition in preverbal infants

### *Do young infants process facing and non-facing bodies differently?*

This chapter reports and discusses the investigation of Goupil, Papeo & Hochmann (2022) published in *Infancy* (<https://doi.org/10.1111/infa.12453>). See Appendix D for the published version.

#### 1. Introduction

Human cognition has been massively shaped by the requirements of social life. Various perceptual adaptations have evolved to serve social life with remarkable efficiency. The social value of objects in the visual world is a guiding principle of human selective attention and perception. Socially relevant entities such as faces and bodies are attended to, detected and recognized with the highest priority, in very young infants (Farroni et al., 2005; Simion et al., 2008; Morton & Johnson, 1991) and throughout the life span (Gluckman & Johnson, 2013; New et al., 2007; Ro et al., 2007; Sun et al., 2017).

Early perceptual tuning for socially relevant entities is a pillar in the development of social cognition (Baron-Cohen et al., 2000), but understanding the social world entails processing social relationships beyond the social entities. How does the ability to understand social relationships come about? Approaching this question, we sought to identify visual perceptual mechanisms that may lay the foundations for detection and recognition of social interaction in the visual world.

Social relationships such as physical or communicative exchanges involve at least two entities, often close and perceptually accessible to one another, to allow fundamental social processes such as gaze following and shared attention (Baron-Cohen, 1997; Birmingham & Kingstone, 2009; Graziano & Kastner, 2011). While a reciprocal relationship illustrated by face-to-face social agents may constitute the prototypical representation of a social relationship, social relationships can also be non-reciprocal, as long as two social entities play a role in the same structure (e.g., in agent-patient relations, where one acts on another). Under the pressure of the social life to encode social relationships, human vision might have developed efficient coding of perceptual units larger and more complex than single social entities.

Encouraging this hypothesis, research on human adults has shown that two bodies facing toward each other, as if interacting, are detected and recognized more efficiently (i.e., faster and/or more accurately) than the same bodies facing away from each other, when stimuli are presented around perceptual threshold (Papeo et al., 2017), or in visual search through a crowd (Papeo et al., 2019; Vestner et al., 2019). Chapter 1, with previous research, has suggested that efficient perception of facing dyads is mediated by perceptual grouping, that is, the processing of two bodies as a unitary configuration (as opposed to two independent items), analogous to the processing of facial features in a face, or body parts in a body (Papeo et al., 2019; Vestner et al., 2019; Yin et al., 2018).

Visual sensitivity to facing –seemingly interacting– bodies in adults could be the product of decades of exposure to social events. Alternatively, it could be a signature of early developing perceptual mechanisms that contribute to the construction of cognitive representations toward a mature social cognitive system. The relationship between perceptual mechanisms that manifest early and the development of cognitive abilities is pervasive across domains of cognition (Carey, 2009). For example, perception of cohesion, boundedness and movement through a continuous path supports object representation (Spelke, 1990); sensitivity to spatiotemporal continuity in the motion of two objects (e.g., object A moves toward object B, which starts moving immediately after being contacted by A) triggers representation of cause-effect relationships (Kominsky et al., 2017; Michotte, 1963; Scholl & Tremoulet, 2000; Leslie, 1982; Leslie & Keeble, 1987; Muentener & Carey, 2010); eye-like stimuli (i.e., two dark spots on a brighter background) or self-propelled motion prompt attribution of agency and intentionality (Gergely & Csibra, 2003; Newman et al., 2010; Saxe et al., 2005).

Along these lines, early discrimination of visuo-spatial relations between social entities could mediate discrimination of basic social events (e.g., conflict or cooperation) and role attribution, promoting rapid enactment of contextually adaptive behaviors and learning about social norms and relationships. In this process, perceptual grouping triggered by certain visuo-spatial relations between human bodies may give an advantage to interacting, over non-interacting agents, which may streamline the processing across body perception and representation of social events.

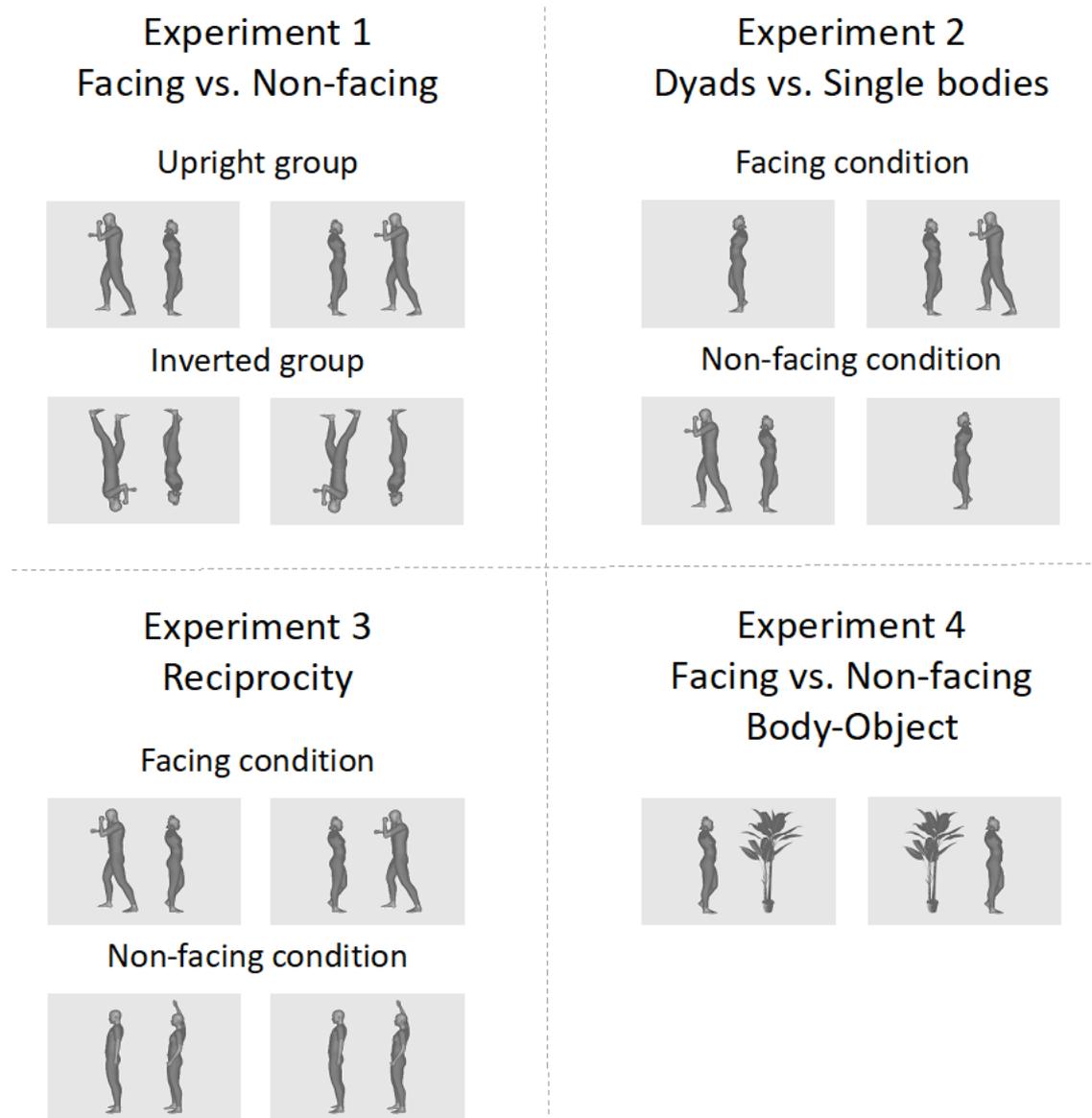
The ability to represent social events emerges early in life. Within the first year of life, infants discriminate between pro-social and anti-social actions (Hamlin et al., 2007; Sloane et al., 2012; Jin, & Baillargeon, 2017; Margoni et al., 2018), they infer dominance-relationship from physical interaction between two conflicting agents (Mascaro & Csibra, 2012; Thomsen et al., 2011), and an agent's affiliation to a group from her imitative motor behavior (Powell & Spelke, 2013; 2018a; 2018b). The relatively short ontological time course of social inference suggests early sensitivity to basic visual cues of interaction, upon which the representation of visual social events would be built.

In the current study, we first asked whether six-month-old infants could discriminate between multiple-person scenarios based on the mere relative positioning of bodies, which could cue interaction or not (Experiment 1). Discrimination could be revealed by longer looking times to facing dyads, or longer looking times to nonfacing dyads. The former effect would be compatible with the preference for socially relevant stimuli (e.g., faces), manifested early on in life (e.g., Farroni et al., 2005; Simion et al., 2008; Morton & Johnson, 1991). In the case of longer looking times toward non-facing dyads, an explanation based on preference would be less likely. The latter effect can rather be predicted based on the abovementioned research on adults, showing that two bodies, whose spatial positioning suggests relationship (e.g., face-to-face), are processed more efficiently (i.e., faster) than two unrelated bodies (Papeo et al., 2017; Papeo et al., 2019; Vestner et al., 2020; Yin et al., 2018). Since Experiment 1 showed evidence for the second effect, we carried out a follow-up study to highlight the processing advantage of facing over non-facing dyads. In Experiment 2, we asked whether facing dyads were processed analogously to single units (i.e., single bodies) in terms of looking times. Finally, we addressed whether

discrimination based on dyadic visuo-spatial relations extended to non-reciprocal relations involving two bodies (Experiment 3) or body-object pairs (Experiment 4). Across four experiments, eye-tracking was used to measure the relative time that infants looked at each of two images presented simultaneously on the two sides of a screen, and differing by one property only: spatial positioning of two bodies (Experiments 1, 3, 4) or number of bodies (one *versus* two; Experiment 2). We reasoned that, if the manipulated property is salient to infants, looking times to the two images should differ. With this logic, we examined the ability of preverbal infants to encode one of the most basic perceptual cues of social relationship, the relative positioning of spatially close bodies. As a result of our investigation, we describe a perceptual mechanism that can contribute to get the processing of visual social events off the ground.

**Figure 4.**

*Examples of the stimuli used in Experiments 1-4.*



## 2. Experiment 1 – Facing vs. Non-facing

Can infants distinguish two human groups involving identical people but in a different spatial relation? To address this question, 6-months-old infants were presented with two images at a time, showing the same two bodies facing toward and away from each other, respectively (Upright Group; see Figure 4). Facing and non-facing dyads differed only for the relative spatial positioning of the two bodies. A second group of infants was tested with the same stimuli, but inverted upside-down (Inverted Group; see Figure 4), to control for the effect of possible visual differences, other than body positioning. Inversion disrupts the body structure, while preserving all the low-level visual features of the images (e.g., shapes, luminance, contrast). By testing inverted body dyads, we could ascertain that any effect observed with upright bodies reflected the processing of bodies, rather than possible low-level visual differences between the two conditions.

### 2.1. Materials & Method

**Participants.** Two groups of 20 6-months-old infants were tested in Experiment 1 (Upright group: 9 females, 11 males, age range 5 months (m) 21 days (d) – 7 m 16 d; average 6 m 19 d; Inverted group: 6 females, 14 males, age range 5 m 14 d – 7 m 04 d; average 6 m 18 d). Data from one participant in the Inverted group were discarded after analysis, based on the exclusion criteria (see below). Experiment 1 was exploratory with respect to the sample size, although we note that a sample size of 20 corresponds to the average and median sample size of relevant infants' studies in the field (a meta-analysis of the 25 infants' studies cited in this article shows that sample sizes ranged from 4 to 42, with a mean value of 18.22,  $SD = 7.82$ , and median of 16; see Appendix B – Supplementary Table S1). The present and following experiments were conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the local ethics committee (CPP sud-est II), and conducted at the Institut des Sciences Cognitives Marc Jeannerod. Parents were given a 5€ for reimbursement of travel expenses.

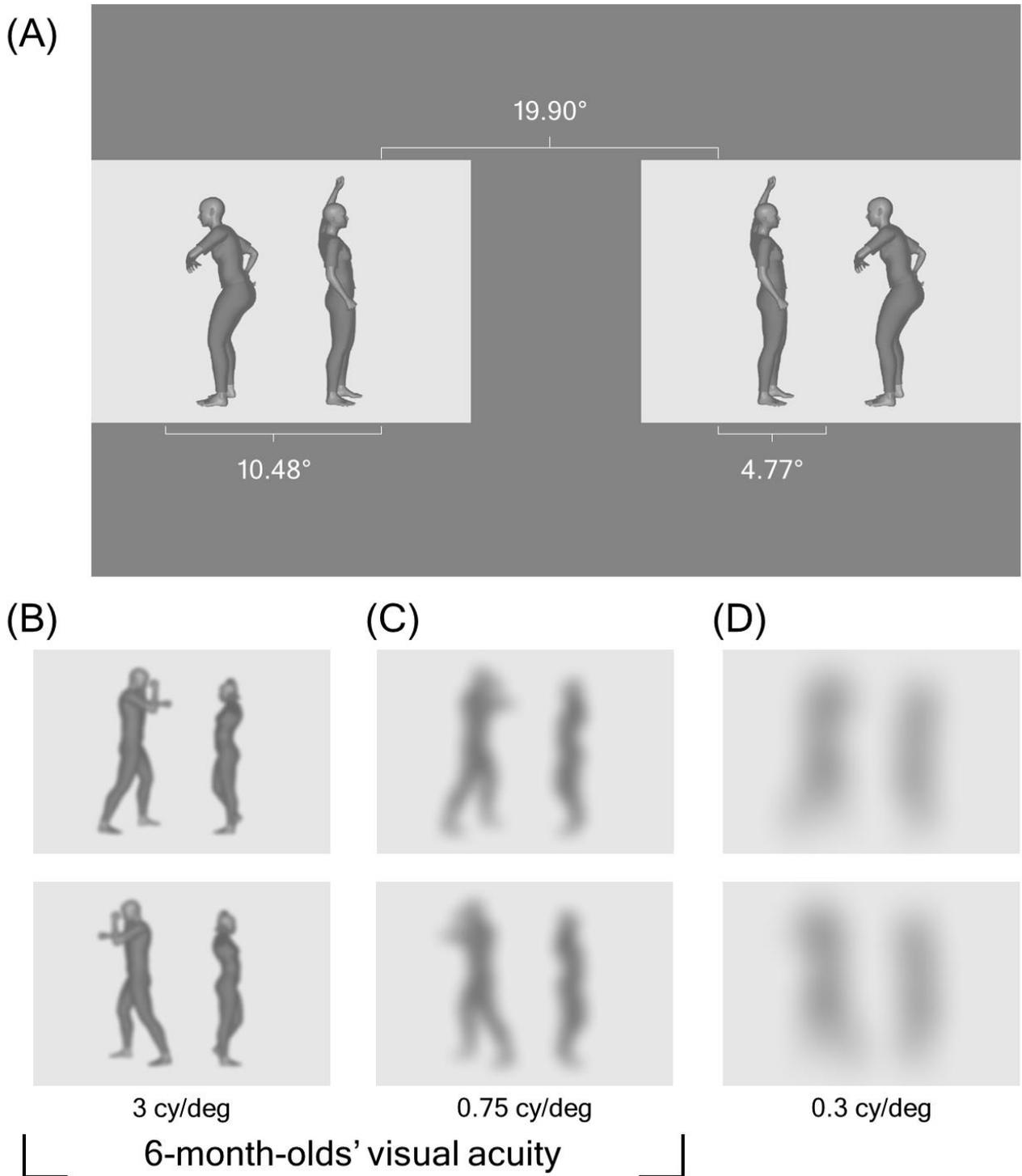
**Stimuli.** Stimuli were created using grey-scale renders of 16 human bodies (eight unique bodies in a lateral view and their mirrored images), edited with Daz3D (Daz Productions, Salt Lake City, UT) and the Image Processing Toolbox in MATLAB (The MathWorks, Natick, MA). Body poses were all biomechanically possible without fully outstretched limbs in order to have a comparable number of pixels in each half of the figure (the anterior and posterior ends), and to keep bodies at a comparable distance across facing and non-facing dyads. Sixteen unique facing body dyads were created combining the 16 bodies. By randomly combining poses in pairs, dyads did not give rise to any meaningful or familiar interaction. The two bodies in each facing dyad were swapped to create 16 non-facing dyads. Thus, facing and non-facing dyads only differed for the relative positioning of bodies. In Experiment 1, stimuli were 16 displays featuring one facing and the corresponding non-facing dyad, at either side of the screen (the facing dyad was on the left in 50% of displays). In all the displays, dyads were presented inside a rectangular box with a different background color (light grey) relative to the screen background (darker grey), so to create two clearly distinct areas for each dyad on the screen. The two areas were separated by  $\sim 19.9^\circ$ , so that the two bodies of a dyad were much closer to one another than to bodies of the other dyad (Figure 5A). A

dyad subtended  $\sim 10.48^\circ$  of visual angle ( $\sim 4.77^\circ$  for a single body). Stimuli of this size are clearly visible at 6 months, when contrast sensitivity peaks between .75 and 3 cy/deg (Gwiazda et al., 1997). Figure 5B-D shows low-pass filtering of an image within a range of .75-3 cy/deg, to illustrate of 6-months-old infants' perception of our stimuli. Dyads were presented in the upright canonical orientation to the Upright group, and rotated by  $180^\circ$ , to the Inverted group.

**Procedure.** Infants sat on their parent's lap at a distance of  $\sim 60$  cm from a Tobii T60XL eye-tracker screen, with a sampling rate of 60 Hz, yielding 300 data points per trial. All lights were switched off, except for those coming from the eye-tracker screen. Parents were instructed to close their eyes during the experiment to prevent biasing the child's behavior. Stimulus presentation and recording of eye-tracking data were controlled through PsyScope X (<http://psy.cns.sissa.it/>). In each experiment, infants were presented with 16 trials. Each trial began automatically, when the infant fixated a blinking cross at the center of the screen for at least 200 ms. The cross was then replaced by the stimulus, displayed for 5 s.

**Figure 5.**

*Illustration of the stimulation with (A) display layout and (B-D) stimulus visibility at 6 months.*



*Note.* (A) Image outlay and dimensions. (B-D) The same exemplar stimulus was filtered by 3 low-pass Gaussian filter. (B) Stimulus low-pass filtered at the highest spatial frequencies of the 6-months-old infants' contrast sensitivity peak range (3 cy/deg; sigma = 7). (C) Stimulus low-pass filtered at the lowest spatial frequencies of the infants' contrast sensitivity peak range (0.75 cy/deg; sigma = 25). (D) Stimulus low-pass filtered at spatial frequencies below the infants' contrast sensitivity peak range (0.3 cy/deg; sigma = 60).

## Analyses

**Exclusion Criteria.** For each participant, missing eye-tracking samples due to technical failure or to the infant looking away, were not interpolated. In Experiments 1-4, for each infant, we computed the cumulative looking time for the two images in each trial. Traditionally, in infant cognition research, individual participants judged as “fussy” are excluded from the final analysis. However, criteria to identify an infant as fussy vary across researchers and studies. Here, we used short looking times to identify trials where infants were inattentive, and low overall cumulative looking times to identify infants that were globally inattentive. In doing this, we set criteria to discard the data-points for which infants were inattentive, while providing data-driven, objective, and replicable cut-offs, which could keep as much information as possible, in the final dataset.

In particular, trials in which cumulative looking time towards the two images were less than 1 *SD* from the mean were excluded from further analyses. This resolved into discarding trials with mean duration of one second or less (Experiment 1:  $M = 962$ ,  $SD = 419$ ; Experiment 2:  $M = 1017$ ,  $SD = 476$ ; Experiment 3:  $M = 720$ ,  $SD = 441$ ; Experiment 4:  $M = 1158$  ms,  $SD = 458$ ). In addition, for each experimental group, participants whose cumulative looking time was below 2 *SD* from the group mean were excluded from further analyses (1 in Experiment 1 – inverted condition; 1 in Experiment 2 and 1 in Experiment 3). Table 3 reports the number of trials included for each experiment.

**Table 3.**

*Number of trials included and excluded in the final analyses.*

Experiment	Total #trial excluded	Total #included	Average # trial exclude per infant ( <i>SD</i> )
Experiment 1 Upright	66	254	3.30 (3.05)
Experiment 1 Inverted	58	262	2.90 (2.43)
Experiment 2	63	257	3.15 (3.80)
Experiment 3	56	264	2.80 (3.02)
Experiment 4	61	259	3.05 (2.54)

**Time course analysis and informative time window.** The time course analysis was carried out to detect transient, but statistically reliable effects that might be hidden in the average, when the entire, relatively long, trial duration is taken into account. Here, the trial duration was 5000 ms. However, infants are not expected to attend to the images for 5000 ms. In fact, every trial began with the infant fixating the center of the screen, meaning that in the very first part of the trial the infant did not fixate either stimulus. The first fixation on either stimulus occurred 500 to 1000 ms after the trial onset. Moreover, toward the end of the trial, infants tended to lose interest and look away from the screen. As a result, data at the beginning and at the end of the trial were noisier and less informative, than in the central part of the trial. To account for this, separately for each experiment, we implemented a data-driven approach to identify the *informative time window* (ITW), that is, the most informative interval, during

which the majority of participants looked at the screen in most of the trials. For every time point, we computed the proportion of trials, in which each infant looked at either image. We compared these proportions to chance (50%) with a cluster-mass permutation test (Hochmann & Papeo, 2014; Maris & Oostenveld, 2007) using Wilcoxon signed-rank test (one-tail), and defined the ITW as the largest cluster of adjacent time points where the p-value of the Wilcoxon test was below .01. While, as we verified for each experiment, there was no significant difference between conditions in the discarded time intervals, excluding those intervals significantly increased the power of our time course analysis.

Within the defined ITW, for each infant, at each time point, we computed the difference between the number of trials in which the infant looked at one image (e.g., the facing dyad in Experiment 1) and the number of trials in which the infant looked at the other image (e.g., the non-facing dyad in Experiment 1), divided by the total number of trials that the infant contributed. With this method, a positive value indicated higher number of looks toward the facing dyad; a negative difference indicated higher number of looks toward the non-facing dyad. The positive or negative value representing the difference between conditions was compared to chance (0) with a cluster-mass permutation test (relying on a Wilcoxon signed-rank test).

***Cumulative looking time analysis.*** With this analysis, we identified the relative time that infants spent on either type of image, presented in a given trial. In a first analysis, we considered the cumulative looking times for the two stimuli within the ITW, as defined for the above time course analysis. In a second analysis, we considered the cumulative looking times over the whole trial duration (5000 ms). For each analysis, for each trial, we quantified the infants' ability to discriminate between two body dyads as the difference between the cumulative looking time (LT) to one type of image and the LT to the other type of image divided by the sum of the two values:  $(LT_{\text{facing}} - LT_{\text{nonfacing}})/(LT_{\text{facing}} + LT_{\text{nonfacing}})$ . Here, a positive value indicated longer looking time toward the facing dyad; a negative difference indicated longer looking time toward the non-facing dyad. Difference scores were compared to 0. The mean time that infants spent looking at the stimuli and mean difference in looking times between the two stimuli are reported in Table 4.

**Table 4.**

Summary of the average total looking times.

Condition (ms)	ITW		5 seconds	
	LT <sub>stimuli</sub> (ms)	LT <sub>difference</sub> (ms)	LT <sub>stimuli</sub> (ms)	LT <sub>difference</sub>
<u>Experiment 1 (Facing vs. Nonfacing)</u>				
Upright	2102 (229)	-160 (223)	3320 (496)	-211 (404)
Inverted	1614 (324)	18 (248)	2522 (578)	-28 (386)
<u>Experiment 2 (Facing or Nonfacing vs. Single)</u>				
Facing	2615 (401)	-36 (461)	3322 (456)	-208 (470)
Nonfacing	2612 (323)	-264 (380)	3289 (505)	-429 (379)
<u>Experiment 3 (Facing or Nonfacing vs. Non-reciprocal)</u>				
Facing	1415 (165)	4 (312)	2979 (646)	55 (557)
Nonfacing	1400 (167)	-154 (249)	2932 (646)	-220 (423)
<u>Experiment 4 (Facing vs. Nonfacing)</u>				
body-plant dyads	2445 (205)	-5 (324)	3323 (384)	-39 (453)

Note. Average total looking times (and standard deviations) towards the stimuli during the Informative Time Window and during the whole trial (5 seconds), and the mean (and standard deviations) differences in cumulative looking times between conditions, for Experiments 1-4. Time is expressed in ms.

## 2.2. Results

**Time course analysis.** The ITW of Experiment 1 fell between 578 and 3094 ms (Appendix B – Supplementary Figure S2A). The cluster-mass permutation test revealed that infants in the Upright group looked longer at the non-facing dyads between 2040 and 2839 ms ( $p = .009$ ); whereas infants in the Inverted group showed no significant bias toward either stimuli, at any point in time (Figure 6A). Between-group comparison in the time course analysis showed a trend for a difference between 2176 and 2652 ms, which however did not reach the significance ( $p = .092$ ), possibly because of limited statistical power to detect small/moderate between-group effects. The between-group difference however was statistically reliable in the following cumulative looking time analysis.

**Cumulative looking time analysis.** We quantified the infants' ability to discriminate between two body dyads, in terms of normalized difference between the LT toward either type of image. This analysis within the ITW showed an effect of group (Upright vs. Inverted:  $U = 109$ ;  $p = .02$ ;  $r_{Rosenthal} = 0.36$ ; Wilcoxon rank-sum test; Figure 6B). In the Upright Group, 17 infants showed a negative difference (longer looking times toward non-facing dyads) and 3 infants showed a positive difference (longer looking times toward facing dyads); in the Inverted Group, 6 infants showed a negative difference and 13 showed a positive difference. The two distributions differed significantly ( $p = .001$ ; Fisher's exact test). Moreover, infants in the Upright Group looked longer at non-facing dyads than at facing dyads ( $M = -0.08$ ,  $SD = 0.11$  corresponding to a difference of  $160 \pm 223$  ms;  $W = 32$ ;  $p = .005$ ;

$r_{Rosenthal} = 0.61$ ; Wilcoxon signed-rank test); while no difference between the two types of dyads was found in the Inverted group ( $M = 0.01$ ,  $SD = 0.15$  corresponding to a difference of  $18 \pm 248$  ms;  $W = 112$ ;  $p = .52$ ;  $r_{Rosenthal} = 0.16$ ; Wilcoxon signed-rank test).

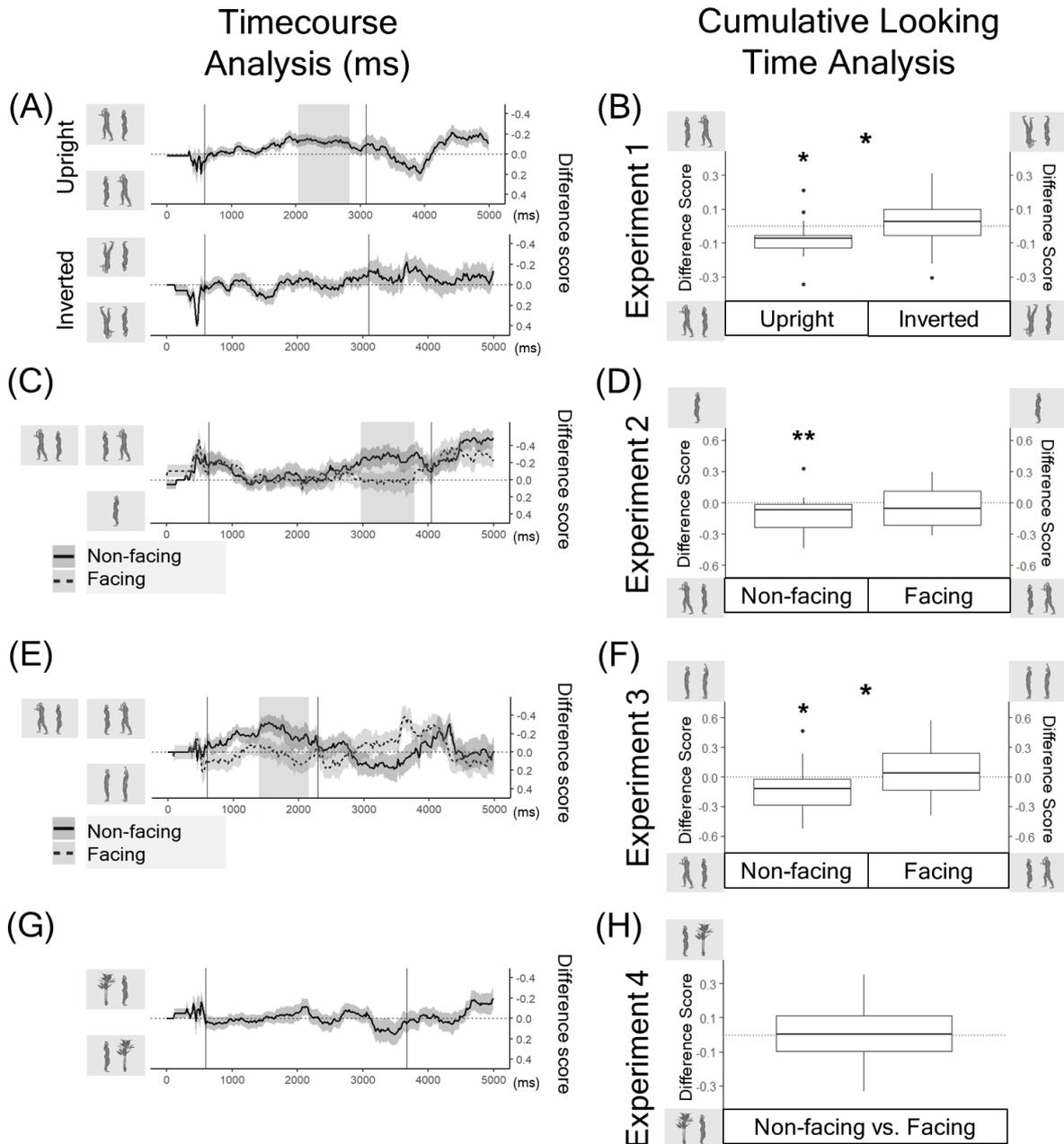
Considering the cumulative looking times over the whole trial duration, pairwise comparisons within each group confirmed the above results. In particular, infants looked longer at non-facing dyads than facing dyads in the upright group ( $M = -0.07$ ,  $SD = 0.12$  corresponding to a difference of  $211 \pm 404$  ms;  $W = 46$ ,  $p = .03$ ;  $r_{Rosenthal} = 0.49$ ), but not in the inverted group ( $M = -0.01$ ,  $SD = 0.17$  corresponding to a difference of  $28 \pm 386$  ms;  $W = 94$ ,  $p = .98$ ;  $r_{Rosenthal} = 0.01$ ). However, the difference between groups was not significant ( $U = 138$ ,  $p = .15$ ;  $r_{Rosenthal} = 0.23$ ), and neither was the difference between the two distributions (Upright Group: 15 infants looked longer toward non-facing dyads and 5 infants looked longer toward facing dyads; Inverted Group: 9 infants looked longer toward non-facing dyads and 10 looked longer toward facing dyads;  $p = .11$ ; Fisher's exact test).

In summary, 6-months-old infants could discriminate between two identical images that differed only for the relative spatial positioning of the two bodies in the scene: facing toward *versus* away from each other. Inversion cancelled the effect, suggesting that the difference between upright body dyads in the Upright Group resulted from processing of bodies and their spatial relations, rather than from other (unforeseen) visual differences between the images, which were all preserved in the inverted images.

These results show that infants encode the spatial relation between bodies in a visual scene. However, while differential looking times imply discrimination between two stimuli (Aslin, 2007), longer looking times to a given stimulus are ambiguous, in that they could index preference for that stimulus, greater complexity relative to the other stimulus (Kidd et al., 2012), or surprise (e.g., Baillargeon et al., 1985). Which one of these cognitive processes accounts for the differential looking times between facing and non-facing dyads? Recent studies on human adults, now supported by Chapter 1, have shown that face-to-face human bodies are processed more efficiently than independent bodies, possibly because they form a structured ensemble, or a group (Papeo, 2020; Papeo et al., 2017; Papeo et al., 2019; Vestner et al., 2019; Yin et al., 2018). In Experiment 2, we considered the possibility that results of Experiment 1 implied higher efficiency in processing facing *vs.* non-facing dyads.

**Figure 6.**

Results of the time course and cumulative looking time analyses for Experiments 1-4.



*Note.* Results of the time course analyses (A,C,E,G) and of the cumulative looking time analyses inside the Informative Time Window (B,D,F,H) for Experiments 1-4. (A,C,E,G) Each curve represents the time course of the difference scores, measuring the tendency to look towards one or the other type of stimuli, by condition. The Horizontal dotted line represents chance (0). Negative values on the Y-axis are plotted upward. Vertical lines mark the boundaries of the informative time window for each experiment. Grey areas highlight the time period of significant differences from chance (Experiments 1 and 4) or significant differences between two conditions (Experiments 2-3), as identified with a cluster mass permutation test. Light-grey areas indicate standard deviations from the mean. (B,D,F,H) Boxplot of the cumulative looking times analysis within the informative time window for each experiment. The thick horizontal bar represents the median. The lower and upper hinges correspond to the first and third quartiles. Whisker extends from the hinges to the largest/smallest values no further than 1.5 times the inter-quartile range. Single dots correspond to values beyond this range. The dotted horizontal line represents the chance level (0). \*  $p < .05$ ; \*\*  $p < .01$ .

### 3. Experiment 2 – 1 vs. 2

Longer looking times to non-facing dyads could index, against all odds, preference for non-interacting over seemingly interacting scenarios. Here, we addressed the alternative hypothesis that longer looking times to non-facing dyads reflected an effect of number, that is, the difference in the time required to process two independent items (two non-facing bodies) *versus* one structured unit (a facing dyad).

It has been shown that adults took longer to evaluate images with two bodies back-to-back, than with two bodies face-to-face; while the time to evaluate facing dyads was rather comparable to the time spent on single bodies (Papeo et al., 2017). In a similar vein, research on infants has shown that after witnessing a social interaction between two dolls, 16-months-old represent the two as one chunk in working memory (Stahl & Feigenson, 2014). Those results encourage the hypothesis that infants, just like adults, can process bodies that appear related to one another, more efficiently than independent bodies, and possibly analogously to one single unit.

To test this, in Experiment 2, we presented 6-months-old infants with displays featuring a dyad and a single body (Figure 4). On each display, the dyad could feature either facing bodies (Facing condition) or non-facing bodies (Non-facing condition). If in Experiment 1 longer looking times to non-facing (*vs.* facing) dyads reflected an effect of number, looking times should differ more between single bodies and dyads when dyads involved non-facing (*vs.* facing) bodies.

#### 3.1. Materials & Method

**Participants.** Experiment 2 involved a new group of 20 infants (14 females, 6 males, age range 5 m 14 d – 6 m 27 d; average 6 m 04 d). In Experiment 2 (and in the following experiments), we tested 20 infants like in each group of Experiment 1, which satisfied the minimal sample size ( $N = 18$ ) to obtain an effect of positioning comparable to that of Experiment 1 (i.e., longer looking times for non-facing *versus* facing dyads;  $t(19) = -3.18$ ;  $p = .005$ ;  $d_{zCohen} = -0.71$ ,  $\beta = 0.80$ ,  $\alpha = 0.05$ ; GPower 3.1 package). Data from one participant were discarded based on the exclusion criteria (see Analyses of Experiment 1), yielding a final sample size of 19.

**Stimuli and procedure.** The 16 displays used as stimuli in Experiment 1 were modified for Experiment 2, by replacing one of the two dyads (either the facing or the non-facing one) with a single body. Thus, stimuli of Experiment 2 were 16 displays, each featuring a facing or a nonfacing dyad on one side, and a single body on the other side (the dyad was on the left in 50% of displays). Stimulus presentation and every other aspect of the experimental procedure were identical to Experiment 1.

**Data analyses.** Data were analyzed as in Experiment 1. In the difference scores, computed separately for the two conditions  $(LT_{single} - LT_{dyad}) / (LT_{single} + LT_{dyad})$ , positive values indicated longer looking times towards the single body; negative values indicated longer looking times towards the dyad (facing or non-facing). The comparison between conditions was one-tailed, given that the direction of the effect (i.e., longer looking times to dyads in both conditions, and larger differences in the non-facing condition) was predicted based on the results of Experiment 1.

### 3.2. Results

**Time course analysis.** The ITW fell between 646 and 4046 ms (Appendix B – Supplementary Figure S2B). The time course analysis showed that the two conditions differed between 2907 and 3791 ms ( $p = .03$ ; one-tail, cluster mass permutation test). Within this interval, infants looked toward the dyad more often in the non-facing condition than in the facing condition (Figure 6C). In the non-facing condition, the difference in looking time toward single bodies vs. dyads was significant between 2958 and 3621 ms ( $p = .03$ ; one-tail, cluster mass permutation test). No significant difference was found at any point in time between single bodies and dyads, in the facing condition.

**Cumulative looking time analysis.** Consistent with the above findings, within the ITW, infants spent significantly longer in fixating the dyad than the single body, in the non-facing condition ( $M = -0.12$ ,  $SD = 0.19$  corresponding to a difference of  $264 \pm 380$  ms;  $W = 25$ ;  $p = .003$ ;  $r_{Rosenthal} = 0.65$ ), but not in the facing condition ( $M = -0.05$ ,  $SD = 0.19$  corresponding to a difference of  $36 \pm 461$  ms;  $W = 64$ ;  $p = .23$ ;  $r_{Rosenthal} = 0.29$ ; Figure 6D), although the difference between the two conditions (facing vs. non-facing) did not reach significance ( $W = 129$ ;  $p = .09$ ;  $r_{Rosenthal} = 0.31$ ; one-tail Wilcoxon signed-rank test). The analysis over the whole trial duration showed that infants looked longer at dyads than single bodies, in both the non-facing condition ( $M = -0.16$ ,  $SD = 0.16$  corresponding to a difference of  $429 \pm 379$  ms;  $W = 16$ ,  $p = .001$ ;  $r_{Rosenthal} = 0.73$ ) and the facing condition ( $M = -0.10$ ,  $SD = 0.20$  corresponding to a difference of  $208 \pm 470$  ms;  $W = 45$ ,  $p = .04$ ;  $r_{Rosenthal} = 0.46$ ). Again, the two conditions did not differ significantly ( $W = 129$ ,  $p = .09$ ;  $r_{Rosenthal} = 0.31$ ; one-tailed).

Experiment 2 confirmed and extended the results of Experiment 1. Infants processed differently two images that were identical except for the relative positioning of bodies. The difference between facing and non-facing dyads here was highlighted in relation to a third stimulus, the single body: although dyads were overall more attractive than single bodies (especially at the beginning of the trial), this effect was consistently found across the three analyses, only when dyads depicted non-facing bodies.

As for the difference between the two conditions, results of the cumulative looking time analysis are not conclusive, though the stronger effect is observed when single bodies were compared with non-facing dyads. More conclusive are the results of the time course analysis, which captured a significant difference between the facing and the non-facing conditions between 2907 and 3791 ms. Interestingly, despite the fact that facing dyads contain twice more visual information than single bodies (twice more bodies, twice more pixels), no significant difference between facing dyads and single bodies was found in any time interval.

In summary, Experiment 2 revealed a consistent effect of number (longer looking times to dyads than single bodies) when single bodies were compared with non-facing dyads, and only an inconsistent effect when single bodies were compared with facing dyads. These results lend credit to the hypothesis that, like adults, infants as young as 6 months may process facing –seemingly interacting– bodies more similarly to a single percept than to two independent items (i.e., two unrelated bodies).

Results of Experiments 1-2 might capture the infants' discrimination between a dyadic relation with two mutually accessible (face-to-face) bodies *versus* the absence of such relation. Alternatively, they might imply a more general discrimination between any relation *versus* no relation. Experiment 3 addressed the categorical distinction underlying the above discrimination between facing and non-facing dyads in infants.

#### 4. Experiment 3 – Reciprocity

In Experiment 3, we introduced a new set of dyads, in which one body faced another, who faced away, in a standing pose (Figure 4). We labelled this type of dyad “non-reciprocal”, as the relative positioning and the passive pose of the facing-away body imply a unidirectional relation with one body addressing the other and no sign of reciprocation. A group of 6-months-old infants saw two types of displays: displays with a facing dyad and a non-reciprocal dyad (Facing condition) and displays with a non-facing (back-to-back) dyad and a non-reciprocal dyad (Non-facing condition).

We reasoned that, if infants represent the broad distinction between presence of relation (reciprocal or not) *versus* no relation, they should discriminate between non-facing and non-reciprocal dyads, but not between facing dyads and non-reciprocal dyads, as the last two would both implement a relation. In contrast, if infants can also represent the narrower distinction between reciprocal and non-reciprocal relations, they should discriminate between facing and non-reciprocal dyads.

##### 4.1. Materials & Method

**Participants.** Experiment 3 involved a new group of 20 infants (9 females, 11 males, age range 6 m 0 d – 7 m 08 d; average 6 m 26 d). One additional infant was tested, but data were not analyzed because of an abrupt interruption (by the infant's sister) of the experiment. Data from one participant were discarded based on the exclusion criteria (see Analyses of Experiment 1), yielding a final sample size of 19.

**Stimuli, procedure and data analyses.** The same displays of Experiment 2 were used in Experiment 3, except that the single body was replaced by non-reciprocal dyads, in which one body faced the other who faced away (the non-reciprocal dyad was on the left in 50% of displays). The 16 non-reciprocal dyads were created by pairing each of the 16 bodies with the same body facing away, in a standing pose. The choice of a passive pose was made to emphasize the representation of one body acting on another, with no reciprocation. Procedures and data analyses were identical to Experiment 1. Difference scores were computed as  $(LT_{\text{non-reciprocal}} - LT_{\text{facing/nonfacing}})/(LT_{\text{non-reciprocal}} + LT_{\text{facing/nonfacing}})$ . Thus, positive difference values indicated longer looking time to non-reciprocal dyads; negative values indicated longer looking time to facing or non-facing dyads.

##### 4.2. Results

**Time course analysis.** Within the ITW, identified for this experiment between 595 and 2295 ms (Appendix B – Supplementary Figure S2C), a difference between the two conditions (facing *vs.* non-reciprocal and non-facing

vs. non-reciprocal) emerged between 1394 and 2159 ms ( $p = .01$ , cluster mass permutation test)<sup>2</sup>. Comparisons within each condition showed that infants looked longer at non-facing dyads relative to non-reciprocal dyads between 952 and 1887 ms ( $p = .01$ , cluster mass permutation tests); while at no point in time they showed a bias to either facing or non-reciprocal dyads (Figure 6E).

**Cumulative looking time analysis.** The analysis of cumulative looking times over the whole trial duration showed no significant difference between the two conditions ( $W = 130$ ,  $p = .17$ ;  $r_{Rosenthal} = 0.32$ ). Within each condition, there was no difference between non-reciprocal and facing dyads ( $M = -0.006$ ,  $SD = 0.27$  corresponding to a difference of  $55 \pm 557$  ms;  $W = 95$ ,  $p = 1$ ;  $r_{Rosenthal} = 0$ ), and only a trend for the difference between non-reciprocal and non-facing dyads ( $M = -0.10$ ,  $SD = 0.22$  corresponding to a difference of  $220 \pm 423$  ms;  $W = 49$ ,  $p = .07$ ;  $r_{Rosenthal} = 0.43$ ).

The analysis of cumulative looking times constrained by the ITW proved more sensitive to the effects of conditions. Here we found a significant difference between the two conditions ( $W = 162$ ;  $p = .005$ ;  $r_{Rosenthal} = 0.62$ ; Figure 6F), whereby infants looked longer at non-facing than non-reciprocal dyads ( $M = -0.18$ ,  $SD = 0.29$ ; corresponding to a difference of  $154 \pm 249$  ms;  $W = 32$ ,  $p = .009$ ;  $r_{Rosenthal} = 0.58$ ), but they spent a comparable amount of time on facing and non-reciprocal dyads ( $M = 0.05$ ,  $SD = 0.27$  corresponding to a difference of  $4 \pm 312$  ms;  $W = 109$ ;  $p = .59$ ;  $r_{Rosenthal} = 0.13$ ).

Overall, the results of Experiment 3 highlighted stronger discrimination between non-reciprocal and non-facing dyads than between non-reciprocal and facing dyads. We found no evidence that infants could discriminate facing and non-reciprocal dyads. Our results show that infants could distinguish non-facing (back-to-back) dyads possibly cueing absence of relationship, from facing and non-reciprocal dyads, possibly cueing presence of relationship, but they could not distinguish between two different types of relationships: facing/reciprocal and non-reciprocal. Thus, reciprocity, or the mutual perceptual accessibility of the two bodies held in facing dyads, is not the key feature for infants to discriminate between different spatial relations. Results of Experiment 3 rather suggest that infants may hold a general representation with a body facing toward another, whether the other faces back or not. Such general representation could virtually apply to social scenes (i.e., scenes involving multiple people such as the current dyads), as well as to non-social contexts, for instance, involving human-object relationships. Experiment 4 addressed this possibility.

## 5. Experiment 4 – Body-object dyads

Experiment 3 suggests that infants can discriminate between scenes in which a body faces toward another (who does or does not reciprocate) vs. scenes in which two bodies face away. Can infants perform such discrimination

---

<sup>2</sup> Visual inspection of the time course shows an apparent difference between conditions, in a late time interval outside the ITW. To test the significance of this effect, we carried out a cluster mass permutation test over the whole trial duration, which confirmed the effect found within the ITW ( $P = .035$ ), and showed no significant difference at the later time interval ( $P_s > .77$ ).

only in the context of social scenes? Experiment 4 addressed whether infants can discriminate between scenes in which a body faces toward another *vs.* scenes in which a body faces away from another, when the other is a non-body object. A new group of 6-months-old infants saw displays featuring two images with a body facing toward and away from a non-body object (i.e., a plant), respectively. On each display, the two images differed only for the relative positioning of the two items.

### 5.1. Materials & Method

**Participants.** Experiment 4 involved a new group of 20 infants (6 females, 14 males, age range 5 m 23 d – 7 m 00 d; average 6 m 20 d). One additional infant was tested in Experiment 4, but data were discarded because a technical failure stopped the experiment. All the remaining participants in Experiment 4 met the inclusion criterion, yielding a final sample size of 20.

**Stimuli, procedure and analyses.** In Experiment 4, 16 new facing pairs were created by coupling each of the 16 single bodies with one of four plants created with the same procedures and software packages as the bodies. By swapping the body and the plant in each facing pair, 16 non-facing pairs were created. Sixteen displays were created with one facing and one non-facing pair showing the same body and plant (the facing pair was on the left in 50% of displays). Procedures and analyses were identical to Experiment 1.

### 5.2. Results

**Time course analysis.** The ITW covered an interval between 595 and 3672 ms (Appendix B – Supplementary Figure S2D). The time course analysis did not reveal any bias towards facing or non-facing body-plant pairs, at any point in time (Figure 6G).

**Cumulative looking time analysis.** Difference scores did not differ from chance, when considering the cumulative looking times within the ITW ( $M = 0.004$ ,  $SD = 0.15$  corresponding to a difference of  $5 \pm 324$  ms;  $W = 111$ ;  $p = .84$ ;  $r_{Rosenthal} = 0.05$ ; Figure 6H), or over the whole trial duration ( $M = -0.01$ ,  $SD = 0.16$  corresponding to a difference of  $39 \pm 453$  ms;  $W = 100$ ,  $p = .87$ ,  $r_{Rosenthal} = 0.04$ ).

The cumulative looking time analyses and the time course analysis failed to identify any significant difference between the infants' looking behavior towards facing and non-facing body-plant dyads. Experiments 1-3 had shown that infants discriminate between scenes in which a body faces toward a second body from scenes in which two bodies face away from each other. Since the directionality of the second body did not seem to be important for the above discrimination (i.e., the second body could reciprocate, facing back to the first, or not), in Experiment 4, we asked whether the identity of the second object (i.e., the fact that it was a body) mattered, or whether we could obtain a similar effect (discrimination between facing toward *vs.* facing away) with non-social scenes, where a body faces toward or away from a non-body object. The null effect of Experiment 4 suggests that infants are especially sensitive to visuo-spatial relations between multiple items, in the context of social (multiple-body) scenarios. Thus, rather than a general action scheme (*someone acting over anything*), the effects reported here imply representation of a social scheme with *someone acting with and/or over someone else*.

**Table 5.***Summary of results.*

Condition	ITW	Time course	Cumulative LTs in ITW
Experiment 1 ( <i>Facing vs. Non-facing</i> )			
Upright	578-3094	Non-facing > Facing (2040-2839, $p = .009$ )	Non-facing > Facing ( $p = .005$ ) <sup>a</sup>
Inverted		n.s.	n.s.
Upright vs. Inverted		Upright > Inverted (2176-2652, $p = .09$ )	Upright > Inverted ( $p = .02$ ) <sup>b</sup>
Experiment 2 ( <i>Facing or Non-facing vs. Single</i> )			
Facing	646-4046	n.s.	n.s.
Non-facing		Non-facing > Single (2958-3621, $p = .03$ )	Non-facing > Single ( $p = .003$ ) <sup>a</sup>
Facing vs. Non-facing		Non-facing > Facing (2907-3791, $p = .03$ )	Non-facing > Facing ( $p = .09$ ) <sup>a</sup>
Experiment 3 ( <i>Facing or Non-facing vs. Non-reciprocal</i> )			
Facing	595-2295	n.s.	n.s.
Non-facing		Non-facing > Non-reciprocal (952-1887, $p = .01$ )	Non-facing > Non-reciprocal ( $p = .009$ ) <sup>a</sup>
Facing vs. Non-facing		Non-facing > Facing (1394-2159, $p = .01$ )	Non-facing > Facing ( $p = .005$ ) <sup>a</sup>
Experiment 4 ( <i>Facing vs. Non-facing</i> )			
body-plant dyads	595-3672	n.s.	n.s.

*Note.* Difference significance and time intervals are reported for Experiments 1-4. <sup>a</sup> Within group comparisons were assessed through Wilcoxon signed-rank tests. <sup>b</sup> Between group comparison was assessed through a Wilcoxon rank-sum test. n.s. =  $p > .10$ .

## 6. Discussion

Across four experiments, we have shown that the visuo-spatial relation between two bodies is a distinctive feature of a scene, to which humans are sensitive early on in life. The relative positioning of two bodies affected the time that 6-months-old spent looking at an image (Table 5). If the looking time is a measure of processing time, then infants processed face-to-face bodies faster than the same two bodies appearing as unrelated (i.e., back-to-back; Experiment 1). Moreover, the difference in looking times was more pronounced between non-facing dyads and single bodies than between facing dyads and single bodies, with the longest looking times toward non-facing dyads (Experiment 2). Such behavior encourages the thinking that, as for adults, also for infants, spatial relations that cue relationship favor the parsing of complex (multi-body) scenarios, producing a perceptual advantage reflected in more efficient (e.g., faster) processing of the stimuli. Infants also discriminated between two bodies back-to-back and a dyad with a body facing toward another, who faced away (non-reciprocal dyads; Experiment 3). Discrimination between non-facing and non-reciprocal dyads was stronger than discrimination between facing and non-reciprocal dyads. Finally, the effect of positioning (facing toward *vs.* away from) was found for visual scenes that involved dyads of bodies, but not human-object pairs (Experiment 4). While acknowledging the relatively small sample size of each individual experiment ( $N = 20$ ) as a limitation of the study, we point out that Experiments 1-3 provide three conceptual replications of the same effect, which emerged most clearly from the time course analysis (see Table 5). In particular, across Experiments 1-3, the time-course analysis consistently showed longer looking times towards non-facing (back-to-back) dyads: infants looked longer at non-facing than

at facing (upright) dyads in Experiment 1; they looked longer at non-facing dyads than at single bodies in Experiment 2 (where the difference was significantly smaller between facing dyads and single bodies); and they looked longer at non-facing dyads than at non-reciprocal dyads in Experiment 3 (where the difference was smaller between facing dyads and non-reciprocal dyads). While most effects were corroborated by the analysis of difference scores based on the cumulative looking times, the time course analysis proved particularly useful to characterize the infants' behavior over time, detecting transient effects that might go lost when considering the average behavior over a quite long time period (i.e., the whole trial duration).

Individual body postures involved in the dyads (as opposed to the whole dyad) cannot explain the current effects of body positioning. First, the very same body postures were used in facing and non-facing dyads. Thus, the specific body postures cannot account for the difference between facing and non-facing dyads, when contrasted against each other directly (Experiments 1 and 4) or against another (common) stimulus type (single bodies in Experiment 2 and non-reciprocal dyads in Experiment 3). Moreover, in Experiment 1, the inversion of bodies abolished the difference between facing and non-facing dyads, thus ruling out the role of low-level features that could differ between conditions. Another feature of our stimuli, which could in principle play a role in the current effects of body positioning, is the face/head/body/gaze direction. Gaze/head/body direction is known to be a powerful cue to orient attention (Driver et al., 1999; Friesen & Kingstone, 1998; Frischen et al., 2007; Langton et al., 2000; Vestner et al., 2020), to which six-months-old may already be sensitive (Reid et al., 2004; Senju et al., 2008). However, the effects that could be predicted based on body direction are opposite to what we observed. The two bodies in a facing dyad would direct the attention toward each other, thus keeping attention on the site (i.e., in between them). Such mechanism predicts that infants would look longer at the facing than at the non-facing dyad; but this is not what we observed. The direction of bodies in non-facing dyads would also bias attention toward the facing dyad, since the two bodies looked away from one another and one of them (the one closer to the central fixation) looked toward the facing dyads. But, again, an attentional bias toward facing dyads is not what we observed. Related to body direction, recent research suggests that, in a dyad, two heads/bodies oriented toward each other can create an attentional *hot spot* that captures the observer's attention more rapidly than dyads in other (non-facing) spatial relations (Vestner et al., 2020). As a result, facing dyads would be looked at, and processed faster, than non-facing dyads. To assess whether this mechanism could explain the current pattern of results, we considered infants' first fixation in each trial, which reflects attentional capture more reliably than looking times. This analysis, reported as supplementary material (Appendix B – Supplementary Table S2), showed no reliable effect, ruling out the role of an unbalanced capture of attention in the infants' looking behavior reported here.

In the current study, back-to-back body dyads elicited the longest visual inspection, relative to all other stimuli considered here. One common interpretation of looking times is preference: between two stimuli, infants would look longer toward the preferred one. On this rationale, a preference for socially relevant stimuli (e.g., faces) has been repeatedly shown in infants (Farroni et al., 2005; Simion et al., 2008). Such preference, however, was found when a social stimulus was compared with a non-social stimulus (e.g., a face *vs.* a scrambled face). In our study,

both conditions (facing and non-facing dyads) had social value (they both involved human bodies) and, as such, they were generally attractive for infants. In this context, attention is likely to be attracted by the weird –i.e., less familiar or less expected– social stimulus, more than by the normal –i.e., familiar or expected– social stimulus (Lewkowicz & Ghazanfar, 2012). Thus, one possible explanation is that longer looking times toward non-facing dyads indicate that those stimuli violated infants’ expectations about the positioning of two nearby bodies. Infants may have an expectation for sociality such that, when two bodies come spatially close, they would engage in some sort of (reciprocal or non-reciprocal) exchange or relationship. Thus, while the current results directly demonstrate the infants’ sensitivity to different spatial configurations of multiple bodies, they open to the possibility that six-months-olds are indeed sensitive to the different representational contents implied by facing and non-facing bodies, showing surprise when two spatially close social agents are non-facing as if ignoring each other, rather than interacting.

On another explanation, if looking times index processing times (see Aslin, 2007; Kidd et al., 2012), non-facing dyads might be more complex than the other stimuli because they consist of two perceptual units that are processed independently from one another. This explanation takes into account a growing corpus of studies on adults, showing that multiple bodies, whose spatial positioning suggests relationship, are processed more efficiently (i.e., faster and/or more accurately) than unrelated bodies (Papeo et al., 2017; Papeo et al., 2019; Vestner et al., 2019; see also Kaiser et al., 2019). Behavioral effects (and related neural effects; see Chapter 1; Abassi & Papeo, 2020; Bellot et al., 2021; Walbrin & Koldewyn, 2019; Walbrin et al., 2020; Wurm & Caramazza, 2019) concur to demonstrate efficient coding of human groups in visual perception, which might reflect grouping, the representation of objects as a unitary whole, as opposed to multiple individual items. Grouping would increase efficiency by reducing the representational complexity of the stimuli and the competition for neural representation and processing resources (Kaiser et al., 2019; see also Desimone & Duncan, 1995).

We shall note that the two explanations –the one based on expectation of *facingness* and/or interaction for two nearby bodies and the one based on grouping of seemingly related bodies– are not mutually exclusive and may actually be two sides of the same coin: a tradition of research on perceptual efficiency has indeed shown that expected stimuli are often the easiest to process (Bukach et al., 2006).

The current results on infants add to the extant literature on adults, suggesting that the visual discrimination of multiple-body configurations does not require extensive experience in the social world; it rather appears early on in life, implying a perceptual adaptation that precedes the expression of social inference in infancy.

Early perceptual adaptations act as powerful learning mechanisms, biasing attention and expectations about the environment (Carey, 2009). In the social domain, high visual sensitivity to certain spatial configurations accounts for the infants’ attentional bias toward faces, a mechanism that mediates learning about people in the infants’ environment and discrimination across human groups (e.g., other-species and other-race effects; Kelly et al., 2007; Pascalis et al., 2005). Early acoustic sensitivity to speech sounds participates in infants’ discovery of the rhythm (Nazzi et al., 1998) and phonology (Werker & Tees, 1984) of their native language. In this spirit, early discrimination of spatial relations between bodies might contribute to encoding of social events, and learning

about aspects of social interaction such as social norms and roles (e.g., who is the dominant party, when and how to show subordination/dominance, to reciprocate and so on), which partly but reliably correlate with visuo-spatial relations.

Facing dyads embody the prototypical social exchange, perhaps because *facingness* favors a number of basic social processes such as shared attention and gaze following. While adults' vision seems to be especially attuned to face-to-face bodies (Papeo & Abassi, 2019), infants do not seem to distinguish between face-to-face (e.g., reciprocal) and non-reciprocal dyads. Other research has shown that non-reciprocal events such as *giving*, with an agent and a recipient, are interpreted as inherently social by 12-months-old infants (Tatone et al., 2015). The representation of social (i.e., human-human) relationship in infancy can start off with a single broad category that embeds both reciprocal and non-reciprocal schemas. Further research shall address the possibility of a perceptual development yielding the visual specialization for the relation *facingness* seen in adults. Moreover, Experiment 4 with body-plant pairs addressed whether having a body facing anything (another body or an object) was sufficient to observe the effects reported in Experiments 1-3. While results showed that infants are especially sensitive to visuo-spatial relations in social (two-body) dyads, and to a lesser extent, to relations in non-social body-object pairs, further research shall address infants' sensitivity to non-social (body-object or even object-object) pairs, when non-body objects have a more unambiguous antero-posterior organization (i.e., a front and a back), and/or evoke stronger representations of interaction, than plants. Building on current findings, it shall also be studied what infants represent when they see a facing dyad, and through which transformations a dyadic schema with mutually accessible body-shapes becomes the representation of social interaction. Here, we manipulated spatial relations under the hypothesis that they can cue social relationship. The specific representational content evoked by two facing or two non-facing bodies, in infants, but also in adults, remains a fascinating question for further research.

Past research has exposed perceptual adaptations in young infants, which define expectations about the world, guide perception and shape cognition. Here, we report evidence for early developing discrimination of perceptual proxies of dyadic social interactions, based on internal spatial relations between participants in a visual scene (i.e., individual bodies). Infants' discrimination resolves in the broad distinction between implied presence and absence of relationship between two people. The mechanism revealed here may contribute to explain how infants' perception prepares the discovery of the keel and backbone of social cognition: relations.

## Chapter 3 – A visual preference for face-to-face social relations in humans and macaque monkeys

*When does the visual preference for face-to-face interactions develop, and what does it mean?*

This chapter reports and discusses an investigation by Nicolas Goupil, Holly Rayson, Émilie Serraille, Alice Massera, Pier Francesco Ferrari, Jean-Rémy Hochmann & Liuba Papeo.

### 1. Introduction

Impressions about others' interactions rely on visual cues such as the spatiotemporal contingencies of movements, postures, direction, orientation and so on (Bassili et al., 1976; Gao & Scholl, 2011; Heider & Simmel, 1944; Scholl & Tremoulet, 2000). Spatiotemporal patterns also change how a scene is visually represented. For instance, people are discriminated and represented better when moving toward one another (Bellot et al., 2021; Neri et al., 2006), and actors moving in synchrony are perceived as a whole (Alp et al., 2017; Cracco et al., 2021). Researchers have proposed that the social brain network, highly specialized in recognizing socially relevant information in the visual world, is especially tuned to dynamic stimuli (Landsiedel et al., 2022; Pitcher & Ungerleider, 2021; Weiner & Gomez, 2021). This visual sensitivity may support the development of core socio-cognitive skills in human and non-human primates, for learning about, or from conspecifics.

Humans exhibit a precocious ability for representing social interaction based on spatiotemporal (dynamic) information (Grossmann, 2021). Young infants discriminate relations between agents like chasing, hitting or caressing, using the same visual cues that adults use to form impressions about such interactions (Premack & Premack, 1997; Rochat et al., 1997). Infants look longer at animations showing agents moving toward one another or in synchrony, as if interacting, than animations depicting non-interacting agents (i.e., agents moving away from each other or desynchronized; Galazka et al., 2014; Rochat et al., 2004; Thiele, Hepach, Michel, & Haun, 2021a). By analogy to the longer looking times toward faces *versus* inverted faces (Farroni et al., 2005), or toward point-light walkers *versus* inverted/scrambled walkers (Bardi et al., 2011; Simion et al., 2008), longer looking times toward animations of social interactions suggest that infants can recognize the higher social value of those stimuli.

Notwithstanding the importance of dynamic information, static visuo-spatial information alone can convey rich information about social interaction. Human adults can rapidly decipher ongoing interactions, thematic roles of the agents and their intentions, based on spatial cues such as spatial proximity, body positioning, and/or body postures, as well as semantic and contextual information (Glanemann et al., 2016; Hafri et al., 2013; 2018; Zhou et al., 2019). Particularly key information is the spatial relation between people (Papeo, 2020). Two people face-to-face are more likely to be represented as interacting, as *facingness* allows for the mutual perceptual accessibility of social agents, facilitating access to facial expressions, lips, gaze cues, or body gestures, and optimizing non-verbal communication (Bavelas & Chovil, 2000) as well as the joining of attention (Baron-Cohen,

1997) and action (Sebanz et al., 2006). This spatial relation scaffolds social interaction from the earliest stages of development, not only in humans but also nonhuman primates like macaques (Ferrari, et al., 2009; Scaife & Bruner, 1975; Striano & Stahl, 2005; Tollefsen, 2005; Trevarthen, 1974). Thus, face-to-face interaction is ubiquitous in social species (Dunbar et al., 1995; James, 1951; Mazur, 1985).

Previous work has shown that *facingness* strongly influences the visual processing of body dyads and that young infants show sensitivity to the relative positioning of human bodies. In particular, as shown in Chapter 2 (see also Goupil et al., 2022), 6-month-old infants look longer at non-facing dyads than facing dyads. This effect would result from the fact that infants process body dyads more efficiently (i.e., as efficiently as single bodies) when they are presented face-to-face, than back-to-back. This interpretation is compatible with the results from studies on adults showing that, relative to non-facing bodies, facing dyads are processed more efficiently (Papeo et al., 2017; 2019) and evoke stronger integrative neural response (intermodulation in Chapter 1), suggesting that they are represented as unitary wholes with consequent processing advantages. Shorter looking time toward facing (vs. non-facing) dyads in infants, however, is opposite to the effect that would be predicted by visual preference: longer looking times toward facing than non-facing dyads. This suggests that the effect found in 6-month-olds is purely perceptual, reflecting the different visual complexity of facing and non-facing dyads, rather than the social evaluation of those stimuli.

Here we asked: *does facingness (i.e., facing dyads) evoke a visual preference effect analogous to effects reported for other socially relevant stimuli?* We first addressed this question in human adults (Experiment 1). Using the same stimuli and preferential looking paradigm as in Chapter 2, we presented images of facing and non-facing dyads simultaneously on the screen. If, by implying interaction, *facingness* adds social value to body stimuli, facing dyads should be looked longer than non-facing dyads, in the same way that socially relevant stimuli (e.g., a face or animations depicting biological motion) are looked longer than visually-matched stimuli (e.g., inverted face or scrambled animations of biological motion; Bardi et al., 2011; Farroni et al., 2005; Galazka et al., 2014; Simion et al., 2008). In addition, we asked a new sample of subjects to rate facing and non-facing dyads with respect to social semantic dimensions (emotional content, intentionality and meaningfulness) as well as a perceptual dimension (implied motion), to test whether positioning (facing/non-facing) actually changed the attribution of socially-relevant properties to identical bodies. To anticipate our results, we found a visual preference for facing dyads (longer looking times toward facing vs. non-facing dyads) in adults, accompanied by higher ratings in all social semantic dimensions. Since, as mentioned above, we had previously shown that young infants (6-month-olds) exhibited the opposite pattern (longer looking times toward non-facing dyads), in Experiment 2, we asked when an adult-like preference for facing dyads emerges during development. Using the same stimuli and procedures as in adults, we tested infants in the first year of life (7 and 10 months) or in the second year of life (15 and 18 months) (Experiment 2a) and young (3 years) and older (5 years) children (Experiment 2b). Finally, since visual preferences for social stimuli in humans (e.g., for faces, biological motion or directed gaze) reveal ancient biologically-constrained mechanisms for social living, and as such are typically shared with other group-living species, in Experiment 3, we asked whether the visual sensitivity to facingness

observed in humans, could also be found in macaques (*macaca mulatta*). These animals live a rich social life (Cheney et al., 1986; de Waal & Luttrell, 1988) and, like humans, develop early visual preferences for socially relevant stimuli such as faces (Kuwahata et al., 2004) or directed gaze (Muschinski et al., 2016). Macaques were tested with stimuli and procedures similar to those used in human infants, children and adults.

## 2. Experiment 1 – Human adults

Human adults were presented with two dyads at a time, featuring two bodies facing each other and the very same bodies facing away from each other. We measured the preferential looking times with an eye-tracker. To test whether a possible difference in looking times could be explained by differences in low-level visual features between facing and non-facing dyads, participants were also tested on inverted images.

### 2.1. Materials & Method

**Participants.** Twenty-four participants were tested (13 females; age range 19 years – 34, age  $M = 23.45$ ,  $SD = 4.37$ ). This sample size was defined *a priori* with a power analysis estimating the minimal sample for detecting a large effect in a two-tailed one-sample *t*-test ( $d = .60$ ,  $\alpha = .05$ ,  $\beta = .80$ ) using *GPower* 3.1 (Faul et al., 2007). Participants reported no history of neurological or psychiatric disorders, declared to take no medication, were right-handed and had normal or corrected-to-normal vision. The present and following experiments on human participants were conducted according to guidelines of the Declaration of Helsinki, with written informed consent obtained from the participant before any assessment or data collection. All procedures were approved by the local ethics committee (CPP sud-est II), and conducted at the *Institut des Sciences Cognitives Marc Jeannerod*. Participants were paid 5€.

**Stimuli.** Stimulus displays were identical to the stimuli presented in Chapter 2 (Figure 7A). Stimuli were created from grey-scale renderings of 16 human bodies (eight unique bodies in a lateral view and their mirrored images), edited with Daz3D (Daz Productions, Salt Lake City, UT) and the Image Processing Toolbox in MATLAB (The MathWorks, Natick, MA). Bodies had 16 different poses, all biomechanically possible. Sixteen unique facing body dyads were created combining the 16 bodies, with each body appearing twice across all dyads, once in profile looking left and the other time in profile looking right, each time paired with a different body. The positions of the two bodies in each facing dyad were swapped to create 16 non-facing (i.e., back-to-back) dyads. Across all dyads, the centers ( $5.3^\circ$ ) and the closest points ( $M = 3.12^\circ$ ,  $SD = 0.69$ ) of the two bodies were equally distant from each other,  $t(30) = 0$ ,  $p = 1$ . Thus, facing and non-facing dyads only differed for the relative positioning of bodies. Dyads were used to create 16 different displays, used as stimuli during eye-tracking. Each display featured one facing dyad, and the corresponding non-facing dyad. The facing dyad was on the left in 50% of displays. Dyads appeared in a rectangular box highlighted with a different background color (lighter gray) relative to the screen background (darker gray), delineating two clearly distinct areas for each dyad on the screen. The two areas were separated by  $19.9^\circ$ , so that the two bodies of a dyad were much closer to one another than the two bodies of the other dyad. A dyad subtended on average  $10.05^\circ$  ( $SD = 0.71$ ) of visual angle (for a single body  $M = 3.52^\circ$ ,  $SD = 1.11$ ). Inverted displays were generated by rotating images by  $180^\circ$ .

**Procedure.** Participants sat ~60 cm away from a Tobii T60XL eye-tracker screen in a dark, soundproof cabin. The experiment began after eye-tracker calibration. Participants were instructed to pay attention to each display, without further instruction. They saw 16 trials showing upright displays and 16 trials with inverted displays. In every trial, displays showed two face-to-face bodies on one side of the screen, and the two same bodies back-to-back, on the other side. Trials began automatically after the participant fixated a cross blinking in the center of the screen for at least 200 ms. The cross disappeared and the test display appeared for 2.5 s. Throughout the experiment, stimulus presentation, recording of eye-tracking data and behavioral responses were controlled through *PsyScope X* (<http://psy.ck.sissa.it>). At the end of the experiment, participants completed a recognition test, in which they saw 8 body dyads presented during the experiment, randomly interleaved with 8 new body dyads. For each dyad, participants had to report whether it was shown during the eye-tracking experiment by pressing “o”, or not by pressing “w” (this mapping was counterbalanced for half participants) (Appendix C – Supplementary material 3.2). This memory task was included to ascertain that participants paid attention to the stimuli during the experiments.

**Analysis.** Analyses were computed in *R* 4.0.2 (R Core Team, 2020), using *eyetrackingR* 0.1.8 (Dink & Ferguson, 2018) for processing eye-tracking data, and *ggplot2* 3.3.6 (Wickham, 2016) for data visualization.

**Preprocessing.** Each trial lasted 2500 ms. Series of up to 5 missing samples (below the threshold of 100ms defining the minimal duration of a fixation; van Renswoude et al., 2018; Wass et al., 2014) were linearly interpolated. Samples were coded depending on whether they were located on-image or off-image (e.g., missing samples, blinks, gaze in the center, on the background, or off-screen). On-image samples included fixations on either dyad. Off-image samples included fixations on any other location on the screen.

**Informative time window.** We used the same data-driven approach described in Chapter 2, to determine the *informative time window* (ITW), corresponding to the points in time in which the majority of the participants looked at images in most trials (see Chapter 2; Goupil et al., 2022). For each participant, at every time point, the proportion of off-image eye-tracking samples was subtracted from the proportion of on-image eye-tracking samples. A positive score indicated a higher number of looks on-image, a negative score a higher number of looks off-image. For each point in time, the distribution of this score across participants was compared against chance (difference = 0) with a one-sample *t*-test. A cluster-mass permutation test was performed to identify significant clusters (Hochmann & Papeo, 2014; Maris & Oostenveld, 2007). The ITW was defined as the largest cluster of adjacent time points, with significant *t* values ( $p < 0.01$ , one-tail). The subsequent analyses were run within this ITW.

**Differential looking time.** Differences in looking times were tested at each point in time within the ITW. For each participant, at every time point, differential looking times between facing and non-facing dyads were computed, separately for upright and inverted displays, as the difference between the proportion of eye-tracking samples on facing dyads minus the proportion of samples on non-facing dyads, divided by the proportion of samples on-image (sum of the two). Positive differences indexed a higher number of looks on facing dyads, negative differences indicated a higher number of looks on non-facing dyads. For time samples at which infants

did not attend to images (the proportion of samples on-image was equal to 0), the difference was set to 0. For each point in time, differential looking times were tested against chance (difference = 0), with a one-sample *t*-test; significant clusters (consecutive significant time points) were identified with a cluster-mass permutation test. Differential looking times for upright displays were then compared with differential looking times for inverted displays (*t*-test, two-tailed), and tested with a cluster-mass permutation test. In significant clusters, differential looking times were averaged across participants, for upright and inverted displays separately, and then tested against chance (difference = 0) with a one-sample *t*-test.

## 2.2. Results

We identified an ITW starting 350 ms after the trial onset and lasting until the end of the trial ( $p < .001$ ). The analysis within the ITW revealed a cluster between 717 and 1167 ms, in which participants looked significantly longer at facing, than at non-facing dyads when upright ( $p = .039$ ; Figure 7B). No significant cluster was found for inverted displays ( $ps > .365$ ). Furthermore, differential looking times for upright images (facing minus non-facing) were significantly different from differential looking times for inverted images (467-950 ms;  $p = .026$ ). In the significant cluster between 717 and 1167 ms (Figure 7C), average differential looking times were positive for upright displays ( $M = 0.10$ ,  $SD = 0.18$ ,  $t(23) = 2.60$ ,  $p = .016$ ,  $d = 0.53$ ) showing that participants looked more at facing dyads than non-facing dyads. There was no significant effect when displays were inverted ( $M = -0.02$ ,  $SD = 0.18$ ,  $t(23) = -0.62$ ,  $p = .543$ ,  $d = -0.13$ ).

In sum, human adults looked at facing bodies longer than at non-facing bodies. This behavior was associated with the processing of bodies and their relations, as it disappeared when this information was disrupted by inversion. By analogy with interpretations of looking time differences between social and non-social stimuli (Bardi et al., 2011; Farroni et al., 2005; Galazka et al., 2014; Simion et al., 2008), this behavior may reflect a preference for facing dyads as the stimuli with the highest social value.

## 3. Experiment 1b – Rating study on human adults

We propose that two identical dyads of bodies differing only with respect to their spatial positioning are represented differently because one (the facing dyad) is assigned a higher social value than the other (the non-facing one). Here, we verify the hypothesis that *facingness* adds social value to the visual scene, by asking human adults to rate facing and non-facing dyads, as well as individual bodies, with respect to perceptual dimensions (implied motion) and social semantic dimensions (meaningfulness of the scene, emotional content, and intentionality).

### 3.1. Materials & Method

**Participants.** A total of 138 English-speaking participants (male and female human adults) were recruited and tested on Amazon Mechanical Turk®; data from one subject were discarded due to a technical failure. We considered this sample size sufficient to measure differences in ratings across conditions. We confirmed this, with a sensitivity analysis reported in the Results section.

**Stimuli.** Body dyads, created as explained above (Method of Experiment 1a), consisted of 30 unique bodies in as many unique body postures, which were randomly paired to create fifteen unique facing dyads. Fifteen non-facing dyads were created by swapping the two bodies in each facing dyad. In each dyad, the centers of the two bodies were at the same distance from the center of the image ( $1.8^\circ$ ), which corresponded to the center of the screen. Moreover, the distance between the closest points of two bodies in a dyad was matched across facing and non-facing dyads (facing:  $1.22^\circ$ ; non-facing:  $1.24^\circ$ ;  $t(29) = 0.292, p = .250$ ).

**Procedure.** From Amazon Mechanical Turk®, participants were redirected to the online platform Testable® (<https://www.testable.org/>) where the experiment was implemented. During the experiment, participants saw facing dyads, non-facing dyads and each individual body that composed the dyads. They had to rate, on a 10-points Likert scale, each image with respect to each of three social semantic dimensions (meaningfulness, emotional content, and intentionality) and a perceptual dimension (implied motion). Each dimension was rated in four separate blocks of 20 stimuli (5 facing dyads and 5 non-facing dyads and 10 individual bodies). In each block, in each trial, a stimulus was presented for 1.5 s. A 10-points Likert scale was shown on the bottom of the screen and remained until the response. Participants had unlimited time to respond. Each participant saw one of three different lists of stimuli, in which each body appeared only once (i.e., in one of the three conditions). The order of blocks and the order of stimuli within a block were randomized. The rating itself was preceded by the calibration of the screen (automated by Testable®), the informed consent and the task instructions (in English). At the beginning of each block, participants were shown a brief reminder of the instructions for that block (i.e., the dimension to evaluate).

### 3.2. Results

As the main purpose of this study was to test differences in the participants' judgment of facing and non-facing dyads, we first focused on pairwise comparisons (*t*-test) between the two types of dyads, separately for each dimension. Results (Appendix C – Supplementary figure S3) showed that facing dyads were rated significantly higher than non-facing dyads for all social semantic dimensions (Meaningfulness:  $t(137) = 6.72, p < .001, d = 0.57$ ; Emotional content:  $t(137) = 5.13, p < .001, d = 0.44$ ; Intentionality:  $t(137) = 5.55, p < .001, d = 0.47$ ) but not for implied motion ( $t(137) = 1.10, p = .272, d = 0.09$ ). These results demonstrated that, although facing and non-facing dyads were matched for perceptual dimensions (here captured by ratings of implied motion), their representation substantially differed with respect to dimensions that are important in encoding social interactions. Put in another way, the spatial relation between bodies in visually-matched images, changed the representation of conceptual dimensions such as meaningfulness, emotional content and intentionality.

*Extended results:* We considered the ratings for individual bodies and tested how they related to rating of facing and non-facing dyads. Results showed that across all the social semantic dimensions, facing dyads were rated higher than individual bodies (Meaningfulness:  $t(137) = 7.05, p < .001, d = 0.60$ ; Emotional content:  $t(137) = 7.07, p < .001, d = 0.60$ ; Intentionality:  $t(137) = 7.55, p < .001, d = 0.64$ ), while ratings did not differ between individual bodies and non-facing dyads (Meaningfulness:  $t(137) = -1.95, p = .053, d = -0.17$ ; Emotional content:  $t(137) = 1.44, p = .153, d = 0.12$ ; Intentionality:  $t(137) = 0.36, p = .719, d = 0.03$ ). In contrast, individual

bodies were rated lower in implied motion, relative to both facing,  $t(137) = 7.06, p < .001, d = 0.60$ , and non-facing dyads,  $t(137) = 5.79, p < .001, d = 0.49$ .

*Sensitivity analysis:* Sensitivity analysis (*GPower* 3.1; Faul et al., 2007) showed that current sample size was sufficient to detect small differences in two-tailed paired *t*-tests ( $t = 1.98, df = 137, d = .21, \alpha = .05, \beta = .80$ ).

## 4. Experiment 2 – Human infants and children

Our previous research (see Chapter 2) had shown that 6-month-olds looked longer at non-facing than at facing dyads. Given that human adults here exhibited the opposite behavior, we reasoned that a shift must take place during the development, from longer looking times to non-facing dyads to a preference for facing dyads. Here we sought to establish when such shift takes place on an ontogenetic timescale, by testing infants at different stages of development. We considered infants (Experiment 2a) distinguishing between young infants (< 1 year: 7 and 10 months) and older infants (>1 year: 15 and 18 months), and children (Experiment 2b), distinguishing between young children (3 years) and older children (5 years).

## 5. Experiment 2a – Infants

### 5.1. Materials & Method

*Participants.* Infants in the first year of life were 7-month-old ( $N = 20$ ; 7 females; age range 6 months (m) 15 days (d) – 7 m 21 d,  $M = 7$  m 03 d,  $SD = 11$  d), and 10-month-old ( $N = 20$ ; 9 females; age range 10 m 06 d – 11 m 17 d,  $M = 10$  m 22 d,  $SD = 13$  d). Infants in the second year of life were 15-month-old ( $N = 20$ , 11 females; age range 15 m 05 – 15 m 27 d,  $M = 15$  m 16 d,  $SD = 8$  d) and 18-month-old ( $N = 20$ , 11 females; age range 18 m 02 d – 19 m 04 d,  $M = 18$  m 20 d,  $SD = 9$  d). The sample size of 20 was chosen following a power analysis based on results in Goupil et al. (2022; Chapter 2; Experiment 1:  $d = -.71, \beta = .80, \alpha = .05$ ; minimal sample size  $N = 18$ ; *GPower* 3.1). Six more infants were tested but rejected because of fussiness (see below). Written informed consent was obtained from the infant's parents before any assessment or data collection. Parents were given 5€ for reimbursement of travel expenses.

*Stimuli and procedure.* The same stimuli, paradigm and procedure as in the adults' experiment were used, except for the following changes. First, only upright displays were shown; second, stimuli stayed on the screen for 5 s, rather than 2.5 s, to take into account infants' slower processing of visual information (Hochmann & Kouider, 2022); third, participants received no explicit instruction. Throughout the experiment, infants sat on their parent's lap, at a distance of ~60 cm from the eye-tracker screen. The size of body dyads was sufficient to be clearly visible at 7 months (Goupil et al. 2022; Gwiazda et al., 1997). Parents were instructed to close their eyes during the experiment, to prevent biasing infants' responses to the stimuli and interference with the eye-tracking. The experiment included 16 trials.

*Analysis.* Fussiness was evaluated using a data-driven approach, as described in Goupil et al. (2022; see Chapter 2). This approach aims at introducing objective criteria to define fussiness across experimenters and studies. In this approach, short looking times are used to identify trials in which infants are inattentive (trials with cumulative

looking times on dyads more than 1 *SD* below the mean), and low cumulative looking times averaged across all trials are used to identify infants who are globally inattentive (infants with looking time cumulated over all trials below 2 *SD* from the group mean). With these criteria, trials with a mean cumulative looking times shorter than two seconds were excluded (7 months:  $M = 1495$ ,  $SD = 572$ ; 10 months:  $M = 1238$ ,  $SD = 486$ ; 15 months:  $M = 1275$ ,  $SD = 751$ ; 18 months:  $M = 1932$  ms,  $SD = 666$ ), and six infants ( $N = 1$  at 10 & 15 months;  $N = 2$  at 7 & 18 months). Excluded infants were replaced to achieve the desired sample size. In the final sample, an average of 17.70 % ( $SD = 15.92$ ) of trials was discarded (7 months:  $M = 19.94$ ,  $SD = 21.99$ ; 10 months:  $M = 19.30$ ,  $SD = 13.98$ ; 15 months:  $M = 15.19$ ,  $SD = 11.27$ ; 18 months:  $M = 16.35$ ,  $SD = 15.21$ ). In order to test groups (and differences between groups) over the same time interval, looking times were analyzed within a common ITW. This was defined by computing the ITW of each group, and selecting the time period that was common to the ITWs of all groups. Differential looking times were computed inside this common ITW. Age differences were tested at each time point, by regressing the effect of age (7, 10, 15, 18 months) on the differential looking times. A cluster-mass permutation test (permuting difference score signs) was used for multiple comparison correction.

## 5.2. Results

Each ITW started with a similar delay across age groups (7 months: 550 ms, 10 months: 533 ms, 15 months: 483 ms, 18 months: 500 ms; all  $ps < .001$ ), and lasted until 5000 ms. Thus, the common ITW was between 550 and 5000 ms. Over this period, using age as regressor, we found three significant clusters (Figure 7B), at three consecutive intervals (2617–2833 ms,  $p = .046$ ; 2983–3267 ms,  $p = .023$ ; 3450–3733,  $p = .020$ ). To explain this effect, differential looking times were averaged by infant across clusters. Under the *a priori* hypothesis that there would be a shift from longer looking times on non-facing dyads in younger infants (see Chapter 2), to longer looking times on facing dyads, as found in adults (see Experiment 1a), each age group was compared to the older age groups with one-tail *t*-tests. We found no difference between 7 and 10 months,  $t(38) = -0.48$ ,  $p = .318$ ,  $d = -0.15$ , and no difference between 15 and 18 months  $t(38) = -0.07$ ,  $p = .471$ ,  $d = -0.02$ . In contrast, we found significant differences between 7 and 15 months  $t(38) = -1.74$ ,  $p = .045$ ,  $d = -0.55$ , between 7 and 18 months  $t(38) = -1.86$ ,  $p = .035$ ,  $d = -0.59$ , and between 10 and 18 months  $t(38) = -1.86$ ,  $p = .036$ ,  $d = -0.59$ , and a trend for a difference between 10 and 15 months  $t(38) = -1.66$ ,  $p = .052$ ,  $d = -0.53$ . These results indicate a discontinuity in the infants' behavior between the first and second year of life: younger infants (< 1 year) looked longer at non-facing dyads; this difference disappeared in older infants (> 1 year) (Figure 7C). This developmental change was confirmed, by a *t*-test comparing infants <1 year (7- and 10-month-olds) against infants >1 year (15- and 18-month-olds),  $t(78) = -2.52$ ,  $p = .007$ ,  $d = -0.56$ . A one-sample *t*-test against chance (difference = 0) found differential looking times significantly lower than 0 in the younger group, meaning that they looked longer at non-facing dyads ( $M = -0.10$ ,  $SD = 0.31$ ;  $t(39) = -2.11$ ,  $p = .042$ ,  $d = -0.33$ ). In contrast, differential looking times did not differ from 0, in the older group ( $M = 0.05$ ,  $SD = 0.24$ ;  $t(39) = 1.39$ ,  $p = .171$ ,  $d = 0.22$ ). In sum, differential looking times showed that infants within the first year of life looked longer at non-facing dyads. These findings replicated previous findings in 6-months-old infants, which were interpreted as reflecting a greater visual

complexity of non-facing dyads, relative to facing dyads (Goupil et al., 2022; see Chapter 2). The difference between the two conditions disappeared in the second year of life.

## 6. Experiment 2b – Children

### 6.1. Materials & Method

**Participants.** We tested preschoolers at 3 years ( $N = 20$ , 8 females; age range 37 m 04 d – 47 m 25 d,  $M = 42$  m 09 d,  $SD = 154$  d) and 5 years ( $N = 20$ , 10 females; age range 60 m 17 d – 71 m 26 d,  $M = 65$  m 06 d,  $SD = 119$  d). All children were attentive and were included in the final analyses. Written informed consent was obtained from a parent before any assessment or data collection. Parents were given 5€ for reimbursement of travel expenses.

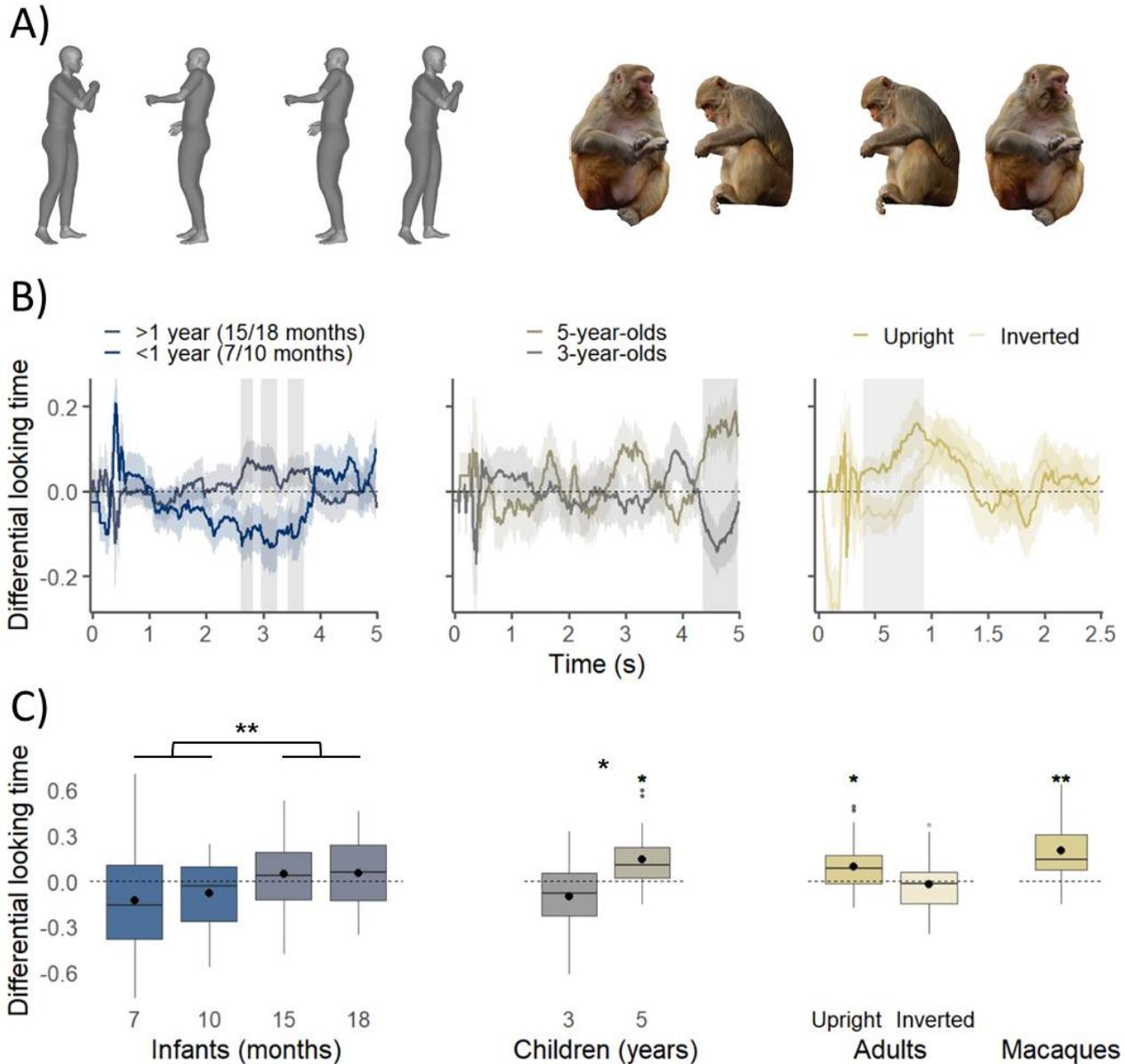
**Stimuli, procedures and analysis.** Stimuli and procedures were identical to Experiment 2a, and so were the analyses. Trials with overall cumulative looking time below 1  $SD$  from the group mean were discarded, which resulted in excluding 16.41 % ( $SD = 13.25$ ) trials on average (3 years:  $M = 16.88$ ,  $SD = 13.16$ ; 5 years:  $M = 15.94$ ,  $SD = 13.67$ ). Excluded trials had a duration shorter than two seconds on average (3 years:  $M = 1413$ ,  $SD = 671$ ; 5 years:  $M = 1904$ ,  $SD = 697$ ). The ITW for each age group was computed on the remaining trials. Then, differential looking times were computed on the common ITW comprising the time points common to the ITWs of the two groups. Age differences were tested at each time point, by comparing the differential looking times of the two groups (two-tailed  $t$ -tests). A cluster-mass permutation test (permuting difference score signs) was used to correct for multiple comparisons (permuting difference score signs).

### 6.2. Results

Each ITW started with a similar delay in both groups (3 years: 500 ms, 5 years: 467 ms; all  $ps < .001$ ) and lasted until 5000 ms. In the common ITW, from 500 to 5000 ms, we found a significant difference between groups in a cluster between 4367 and 4983 ms ( $p = .026$ ; Figure 7B). Differential looking times inside this cluster were averaged for each subject, and tested against chance with one-sample  $t$ -tests (two-tailed). Showing continuity with infants in the second year of life, differential looking times did not differ from chance in 3-year-olds,  $M = -0.10$ ,  $SD = 0.26$ ,  $t(19) = -1.71$ ,  $p = .103$ ,  $d = -0.38$ . However, they were significantly above chance in 5 year-olds, revealing an adult-like propensity to look longer at facing dyads,  $M = 0.14$ ,  $SD = 0.20$ ,  $t(19) = 3.13$ ,  $p = .005$ ,  $d = 0.70$  (Figure 7C). A cluster-mass permutation test comparing the time course of differential looking times of 5-years-olds against chance showed a significant cluster between 4400 and 4983 ms ( $p = .043$ ). There was no significant effect for the 3-year-olds.

**Figure 7.**

*Stimuli and results of Chapter 3.*



*Note:* **A) Stimuli.** Examples of facing and non-facing dyads presented to humans (left) and macaques (right). In each trial, a facing dyad and a non-facing dyad featuring the very same bodies were presented simultaneously on the screen. **B) Results of the time course analysis** testing whether and when in a trial, subjects looked more at facing or non-facing dyads. For each time point, the curves show the proportion of looks to the facing dyads minus the proportion of looks to non-facing dyads, divided by the sum of the two. Horizontal dotted lines denote the chance level (0); positive values mean that subjects looked more to facing dyads; shaded areas around the curves denote standard errors from the mean; intervals highlighted by gray areas are those where significant differences between groups (left and central plots) or conditions (right plot) were found with cluster-mass permutation tests. From left to right: results of infants (below *versus* above than 1 year), children (3 *versus* 5 years), and adults (upright *versus* inverted displays). Note that infants and children only saw upright displays). **C) Results of the analyses on the differential looking times** averaged across all the time points in the intervals where significant differences were found (gray areas in B). In the boxplots, dots indicate means, thick horizontal bars medians, lower and upper hinges 1<sup>st</sup> and 3<sup>rd</sup> quartiles respectively, whiskers the span encompassing values

largest/smallest than 1.5 times the inter-quartile range, small dots values beyond this range, and horizontal dotted lines the chance level (0). From left to right: results of infants, children, adults (upright and inverted stimuli) and macaques. Stars above boxes denote effects different from chance (0) (one-sample *t*-tests), stars between boxes denote significant differences between conditions (pairwise *t*-tests) \*  $p < .05$ ; \*\*  $p < .01$ .

## 7. Experiment 3 – Macaques

We tested the visual preference for facing dyads in a group of macaques, using video recording to measure looking times toward static images of facing vs. non-facing macaques.

### 7.1. Materials & Method

**Participants.** Twenty-one juvenile rhesus macaques (*Macaca mulatta*; 11 females; approximate age: 2.5 years) were tested in an indoor environment, in which they were free to move around, without any restraining device. All subjects had previously taken part to experiments with visual stimuli presented on the computer monitor and, therefore, were familiar with the current setting. All housing and procedures conformed to current guidelines regarding the care and use of laboratory animals (European Community Council Directive No. 86–609), and were approved by our local ethics board (03.10.18) and the French Ministry of Research (10.10.18) (Appendix C – Supplementary material 3.2).

**Stimuli.** A set of ten colored images was created including five unique photographs of macaques (open licensed pictures available on Google Image) and their mirrored images (Figure 7A). In each image, the monkey appeared on a white background in lateral view, sitting in a natural posture with neutral facial expression, and gaze, head and body oriented in the same direction (leftward or rightward). Twenty unique facing dyads were created combining the ten photographs. Each body was presented once in each view (i.e., oriented leftward or rightward) paired with another individual. Individual bodies were placed at the same distance from the center of images (visual angle:  $10^\circ$ ). To create 20 non-facing dyads, the positions of the two bodies in each facing dyad were swapped. Thus, facing and non-facing dyads only differed for the relative positioning of bodies. Stimuli for the experiment consisted of displays featuring a facing dyad and the corresponding non-facing dyad image (visual angle:  $20 \times 15.33^\circ$ ), next to each other (the facing dyad was on the left in 50% of displays), separated by  $2^\circ$  of visual angle. Both dyads showed the same monkeys, only differing for the positioning of bodies. Each dyad appeared once on the left side of the screen, once on the right side. Dyads were equally distant from the center of the screen.

**Procedures.** Each subject was temporarily separated from the social group and placed into the testing area, a large cage (87 x 100 x 120 cm) in which the animal was free to move, with a front, delimited by a large-mesh metallic grid. A computer monitor (35 x 61cm; 2560 x 1440 resolution) was placed 60 cm from the grid. Subjects were given five minutes to habituate to the testing area, before the experiment began. A moving geometric pattern accompanied by a non-social sound appeared in the center of the screen to attract the subject's attention; when the animal looked toward the screen, a stimulus was shown for 5 s. Stimulus presentation was triggered by an experimenter, who monitored the animal's behavior on a separate monitor connected to a webcam (not visible to

the subject). During the experiment, subjects were recorded with a webcam (30 fps) placed on the top-center of the screen. Video recording onset/offset was automatically triggered at the start/end of each stimulus presentation, controlled through Psychopy v1.90.2 (Peirce et al. 2019). In total, each subject was presented with a maximum of 10 trials. The spatial arrangement of dyads on the first trial (e.g., facing dyad left, non-facing dyad right) was counterbalanced across subjects. The positioning of facing and non-facing dyads on the display (left or right) alternated across trials.

**Analysis.** Subjects' gaze position (left/right/other/off-screen) was manually coded offline, frame-by-frame, by a researcher (HR) blind to the position of the two dyads on the screen. This researcher had been established as reliable using this coding scheme, with very good reliability scores ( $\kappa = 0.84$ ) obtained in a previous study with the same paired-stimuli presentation set-up and the same coding scheme (Rayson et al., 2021). On each video-frame, the coder decided whether the monkey looked at the right image, the left image, in an ambiguous location or in a task irrelevant location (off the display). Each entry of the coding file indicated the number of consecutive frames during which the monkey looked in either direction. Next, this number was multiplied by the frame duration (in seconds), to obtain a looking time. Trials in which the monkey looked at the two dyads for less than 500 ms in total, were discarded. Subjects with less than two trials were discarded. For the remaining monkeys, for each trial, differential looking time was computed as the difference between looking time to the facing dyad minus looking time to the non-facing dyad, divided by the total looking time (sum of the two). For each subject, differential looking times were averaged across trials and tested against chance (difference = 0) with a one-sample *t*-test, where positive values denoted longer looking times toward facing dyad and negative values denoted longer looking times toward non-facing dyad.

## 7.2. Results

Five monkeys were excluded as they never attended to the displays. One more subject provided only one trial above the inclusion criterion (looking time >500 ms) and was excluded from subsequent analyses. For the remaining subjects ( $N = 15$ ; 7 females), on average the 35% ( $SD = 29$ ) of trials were discarded due to looking times <500 ms. The analysis of the remaining trials of these 15 subjects revealed significantly positive difference scores ( $M = 0.20$ ,  $SD = 0.22$ ,  $t(14) = 3.56$ ,  $p = .003$ ,  $d = 0.92$ ) (Figure 7C), indicating longer looking times for facing than for non-facing dyads. Thirteen out of 15 macaques exhibited longer looking times for facing than non-facing dyads (Exact binomial test  $p = .007$ ).

## 8. Discussion

Human adults, children and infants as well as macaques were tested in a preferential looking paradigm, to measure a possible visual bias for facing *versus* non-facing body dyads. On our hypothesis, while both facing and non-facing bodies are highly salient social stimuli, if facingness adds critical social value to the stimuli, participants should spontaneously look longer toward facing (*vs.* non-facing) dyads, similarly to what is observed when other socially relevant stimuli (faces or biological motion) are compared with visually-matched stimuli (Bardi et al., 2011; Farroni et al., 2005; Galazka et al., 2014; Simion et al., 2008). Differential looking times

showed a bias for facing dyads in human adults. Consistent with the interpretation of this effect in terms of preference, human adults rated facing dyads higher than non-facing dyads in social semantic dimensions such as intentionality, meaningfulness, or emotional content. A similar preference was also found in children by 5 years, but not in younger infants/children, suggesting that the preference for facing dyads is a relatively late achievement, which might signal the development of sophisticated social cognitive abilities. Finally, macaques, nonhuman primates living in complex social systems and developing visual processes close to humans, also looked longer at facing (*vs.* non-facing) conspecifics. These results and their implications are discussed in the following.

### **The visual preference for facing dyads**

We interpret the preference for facing bodies in human adults and children as analogous to preferences observed for other socially relevant stimuli such as faces (Farroni et al., 2005) or biological motion (Bardi et al., 2011; Simion et al., 2008) when compared with visually-matched stimuli. Consistent with this interpretation, human adults evaluated that facing bodies looked more intentional, carried richer emotional content and gave rise to a more meaningful scene than the very same bodies presented back-to-back. These results are in line with previous research showing that orienting individuals toward one another increased observers' ratings of the scene as depicting social interaction (Isik et al., 2020; Zhou et al., 2019). Moreover, they add to behavioral studies showing that humans readily extract information about social interaction from static images (Glanemann et al., 2016; Hafri et al., 2013; 2018; Isik et al., 2020; Zhou et al., 2019). These researches indicate that static visual features, like dynamic features, contribute to representing social events. Facingness, illustrating the need to access each other while interacting, may be one such feature.

### **The emergence of the visual preference for facing dyads through development**

The visual preference for facing dyads was not found in the first years of life: In contrast to adults and 5-year-olds, infants did not prefer people facing over people non-facing. Seven- and 10-months old infants looked longer at non-facing bodies than facing bodies, replicating results of Chapter 2 (Goupil et al., 2022). As discussed in Chapter 2, this behavior can have two interpretations, not mutually exclusive. First, infants' longer looking time to non-facing bodies might reflect surprise, or violation of the expectation that two nearby bodies are oriented toward one other, and even interact. In this perspective, longer looking times to non-facing bodies *versus* facing bodies would be similar to previously reported effects of longer looking times toward uncanny faces *versus* natural faces (Lewkowicz & Ghazanfar, 2012), or disarticulated bodies *versus* regularly structured bodies (Slaughter et al., 2002; Zieber et al., 2015), and reflect an expectation regarding how people should be positioned when they are spatially close.

A second possible explanation is that young infants look longer at non-facing bodies because they process them more slowly than facing dyads. As shown in Chapter 2 (Experiment 2), infants looked at facing dyads for as long as they looked at single bodies; and the looking time toward facing dyads and single bodies was shorter than the looking time toward non-facing dyads. This interpretation is encouraged by a wealth of results suggesting that human adults process facing dyads as a unitary structure, with an increase of the computational efficiency, also

captured by a decrease in processing times, while they process non-facing bodies as separate perceptual units (see Adibpour et al., 2021; Kaiser et al., 2019; Papeo, 2020). Note that a similar interpretation is also encouraged by the results reported in Chapter 1. There, we targeted a neural signature of integration, previously associated with perceptual grouping, and showed that this response was larger for facing than non-facing bodies.

It should be noted that those effects in human adults were found in the neural responses, or in behavioral measures (i.e., reaction times or accuracy rates) during difficult visual tasks (with fast, near-threshold visual stimulation, with visual masking; see Papeo, 2020). Relatively small differences in visual perception require specific paradigms to be highlighted in adults, and can hardly be captured during passive observation, as used in the current study, where we found no evidence for less efficient processing of non-facing bodies in adults. The current paradigm rather emphasized the explicit evaluation of the stimuli, which would reflect the social values of the stimuli, under the hypothesis that individuals have a spontaneous tendency to attend to more (*vs.* less) socially relevant stimuli (see Di Giorgio et al., 2021; New et al., 2007). As a result, individuals looked longer at facing (*vs.* non-facing) dyads. This effect was first found in 5-year-olds suggesting that at this age: *a*) the perceptual system is mature enough to process with efficiency both facing and non-facing dyads (unless the task is designed to highlight subtle differences); and *b*) children have developed the abilities for appreciating differences in the social value of facing and non-facing dyads.

The relatively late emergence of the preference for facingness suggests the computational difficulty of the process. Before 5 years, infants show a preference for directed gaze/movements (Farroni et al., 2004), and can understand mutual gaze between others as a social act, but only when accompanied with other powerful cues of social interaction such as conversation or motion (Beier & Spelke, 2012; Galazka et al., 2014; Thiele, Hepach, Michel, & Haun, 2021a). What our results suggest is that the social value of facingness is much harder to infer from static body postures of a third-party event. This would require the observer to represent each individual's orientation of attention and put the two in relation, as well as understanding that facingness is as relevant to others as it is to oneself (i.e., when directed to oneself), and attributing to others the ability to interpret gaze/body orientation, a core component of theory of mind.

Finally, we note that it is possible that the abilities for the social evaluation of facing/non-facing dyads, which would resolve into a preference for facing dyads, emerged before 5 years, but were masked by the effect of visual complexity producing the opposite effect (longer looking times for non-facing dyads). In this case, the two effects would cancel each other out resolving into a null effect in the second and third year of life. It also remains possible that the effect of visual complexity vanishes in these age groups, but the abilities for social evaluation are not yet developed. Future research, with paradigms designed to untangle the effects of visual complexity and social preference, will clarify the developmental change that leads from an effect (longer looking times toward non-facing dyads) to the other (longer looking times toward facing dyads).

## **Evolution of a social mind**

Like humans, macaques exhibit a preference for socially relevant stimuli. For instance, infant macaques look longer at schematic faces relative to visually matched schemata (Kuwahata et al., 2004), and at directed gaze relative to averted gaze (Muschinski et al., 2016). Like humans, macaques here looked at facing conspecifics longer than at non-facing conspecifics, suggesting a preference for the former stimuli. Like in humans, in the macaques' group-living, facingness is core to social interaction since infancy (Ferrari, et al., 2009), supporting the development of socio-cognitive skills like joint attention (Simpson et al., 2016), and aspects of socialization such as hierarchical ranking (Mazur, 1985). The preference for facing dyads reported here indicates that macaques understand the declarative referential content of facingness, that is, they are capable of generalizing the social meaning of facingness, which they appreciated when self-directed, to third-party interactions.

## **Conclusions**

Present works suggest that the spatial relation *facingness* between two conspecifics elicits longer looking times in human adults, and children, as well as in monkeys, which we interpret as preference. The social value of spatial relations between individuals may have pressured the primate visual system to tune to this information precociously. In fact, at six months, infants can already use the relative spatial positioning of bodies to discriminate between otherwise identical visual scenes (Chapter 2-3). However, understanding the social significance of people facing may require the development of complex cognitive abilities that are achieved later on. A spontaneous visual preference for facing dyads in humans as well as monkeys can be crucial for selecting and processing aspects of the visual environment from which individuals can learn the most, namely, social interaction and relational structures, setting individuals on the path of social learning.

## General discussion

Beyond individual behaviors, people's interactions structure the social world into networks of relations. Building on previous work (see Papeo, 2020 for an overview), we hypothesized that, in order to detect and recognize interactions, human vision may have developed specialized mechanisms, sensitive to the characteristic visual features of social interaction. This thesis focused on *facingness*, a spatial positioning common to many social interactions, for it supports mutual perceptual access. We addressed the specialization for processing face-to-face people, asking the following questions: Are facing people processed differently from non-facing people in visual perception (Chapter 1)? How early in life infants manifest a sensitivity to socially-relevant spatial relations between people (Chapter 2)? What is the relationship between representing *facingness* (i.e., a spatial relation) and representing social interaction (Chapter 3)?

### 1. Grouping bodies

Building on previous work, showing that facing bodies are perceptually, automatically grouped, in Chapter 1 we developed a paradigm based on frequency-tagged EEG to measure a physiological correlate of grouping. We found that facing (vs. non-facing) dyads elicit a stronger intermodulation response in the EEG signal. This response could only emerge if neural populations had received and integrated non-linearly visual inputs from both bodies (Ratliff & Zemon, 1982; Regan & Regan, 1988; Victor & Conte, 2000). Hence, it indicates that the human adult brain automatically integrates the individual representation of face-to-face bodies in a holistic percept. By analogy with studies evidencing signals of intermodulation with face features forming a face (Boremanse et al. 2013; 2014), agents moving synchronously (Alp et al., 2017), and social groups (Mersad & Caristan, 2021), this result suggests that perceptual mechanisms analogous to those that underlie integration of low visual information in a gestalt (Alp et al., 2016), or figure ground segregation (Appelbaum et al., 2008), can also operate to integrate higher visual information signaling meaningful relations between these entities (here, bodies).

Moreover, this result converges with previous work in showing that human vision may have developed perceptual mechanisms analogous to those involved in individual face and body perception, for processing dyads of face-to-face (seemingly interacting) people as single visual units (Papeo, 2020). Using fMRI, Abassi & Papeo (2020) reported evidence of integration of facing dyads in the form of neural activation predicted by a non-linear combination of the neural response to the individual bodies. Furthermore, the processing of facing dyads has been associated with an increased inversion effect, a behavioral index of the integrated processing of parts into whole (Papeo & Abassi, 2019; Papeo et al., 2017), which also predicts the response to body dyads in areas of the visual cortex (Abassi & Papeo, 2022).

Concurrently with the stronger intermodulation signal for facing versus non-facing dyads, in Chapter 1 we also reported that the neural response at fundamental and harmonic frequencies was reduced for single bodies presented in facing (vs. non-facing) dyads. This effect is in line with evidence showing that grouping reduced access to the individual parts of the grouped representation (here, single bodies). In particular, in a visual search

experiment (Papeo et al., 2019), it was shown that while subjects were more efficient at finding a facing dyad in a crowd, relative to a non-facing dyad, they were less efficient at finding an individual body-target when it was embedded in a facing dyad, than in a non-facing dyad. This pattern was reminiscent of the so-called “object inferiority effect”, that is, the less efficient detection of face features such as an eye or a mouth, presented in a regular face schema, relative to a scrambled face (Suzuki & Cavanagh, 1995). By showing that the grouped representation of facing dyads interferes with access to individual bodies, these findings support the hypothesis that bodies in that configuration are processed at a global level, a grouped representation.

Another line of visual search experiments supports the automatic grouping of facing bodies. Earlier work showed that the serial search for a target-object through distractor-objects is more efficient when distractors are arranged in a meaningful way, forming functionally related sets (mirror above sink, monitor above keyboard, etc.) (Kaiser et al., 2014). Kaiser et al. (2014; Kaiser et al., 2019) proposed that objects in meaningful spatial relations are automatically grouped, effectively reducing the number of distractors to process in the search array. We reported a similar effect with facing (vs. non-facing) bodies (Papeo, Goupil & Soto-Faraco, 2019). In Experiment 3, we showed that participants search more efficiently for a non-facing dyad (target) among facing dyads (distractors), than *vice versa*. Because this difference increased with the number of distractors, the advantage depended on the processing of distractors (i.e., how efficiently distractors could be checked and rejected in the search). Appendix E of this thesis reports a replication and an extension of the visual search work, defining the task- and stimulus-level conditions that determine how meaningful spatial relations between people and objects affect visual search (see also Goupil, Kaiser & Papeo, 2023). In the above studies grouping was inferred from behavioral patterns of performance (cost of inversion effect on body recognition, or visual search performance in different visual search task conditions) or from *the lack of a linearity* in the neural response to facing bodies. In Chapter 1, by measuring intermodulation response, we provide the possibly most direct evidence that two facing bodies, more strongly than two non-facing bodies, evoke automatic integration; in sum, they are grouped in a single perceptual unit.

Chapter 2 further investigated –and demonstrated– that an analogous grouping mechanism can apply to the infants’ processing of multiple bodies. This was shown in a series of experiments measuring the differential looking times of 6-months-old infants presented with two scenes, each consisting in one or two bodies. First, we showed that infants are sensitive to spatial relations between people, particularly distinguishing between bodies in a positioning that cue interaction/relation (face-to-face or face-to-back) *versus* unrelated (back-to-back) bodies. Second, we showed that related (face-to-face) bodies were looked at for as long as single bodies, but for a shorter time than unrelated (back-to-back) bodies. We proposed that this effect reflected more efficient processing of facing dyads, as efficient as the processing a single unit (e.g., a single body). On this interpretation, 6-months-old infants, like adults, may group facing bodies in one visual unit, which reduces processing time.

## **2. Social evaluation**

Our experimental work has shown that facing people are represented as a single group or visual unit, that is more than the sum of the individual bodies. But what is “more”? What additional value or information results from the representation of grouped bodies?

The facing dyads used in our (and previous) experiments did not necessarily depict familiar interactions, as bodies were paired without regards to the meaning of individual postures (for instance, a guy dancing could be paired with someone punching him). Yet, in Chapter 3, we showed that adults rate facing dyads higher than non-facing dyads on social semantic dimensions. Moreover, adults, 5-years-old children and macaques showed spontaneous visual preference for facing dyads *vs.* non-facing dyads, similar to the visual preference extensively documented for socially relevant stimuli such as faces (*vs.* scrambled faces), or biological motion (*vs.* random motion) (Bardi et al., 2011; Farroni et al., 2005; Simion et al., 2008). Hence, even though dyadic scenes were not systematically interpretable, the mere facingness enhanced the social value assigned to the images and evoked a spontaneous visual preference, suggesting that facingness, just like faces and biological motion, works as a powerful social signal, holding an individual attention.

The idea that social information yields spontaneous visual preference is common in the literature (End & Gamer, 2017; New et al., 2007). Walbrin & Koldewyn (2019) have proposed that interacting people are processed more extensively than non-interacting people for they convey “something more”, some socially relevant information that may spontaneously attract adults’ attention, relative to less relevant information (see also Skripkauskaitė et al., 2022). According to Thiele, Hepach, Michel, Gredebäck, et al. (2021), third-party interactions spontaneously attract attention for they are intrinsically rewarding. With this hypothesis, Thiele, Hepach, Michel, & Haun (2021a) observed that infants at the end of their first year of life prefer social scenes presenting powerful cues of interactions, such as conversation or motion, over non-interactive scenes. In our studies, visual preference for facing dyads emerged much later, at 5 years, suggesting that evaluating the social value of a scene from static body postures of individuals, without any further interaction cue, is more difficult.

Altogether, the findings reported in Chapter 2-3 suggest a two-step process of social interactions from static depiction of human bodies. In the first step (perceptual analysis), bodies are grouped as a structured unit, on the basis of visuo-spatial cues such as spatial proximity and facingness. The first step is already social in a certain sense, as it only applies to dyads of social agents (bodies). Replacing one body with a plant (Chapter 2), or both bodies with objects (Chapter 1), halted grouping. In the second step (social evaluation), this unit is further analyzed and evaluated as a social interaction. Developmental data shown in Chapters 2-3 suggest these hypothesized two steps may develop sequentially. Six- to 10-months-old infants supposedly processed a facing dyad as a group, but there was no evidence that they saw them as more social compared to non-facing dyads, at least considering the visual preference measure. In contrast, 5-years-old children and adults’ visual preference for facing dyads (Chapter 3) may indicate that they recognize facing dyads as socially relevant signal, possibly, as social interactions.

### **3. Outstanding questions and limitations**

The studies reported in this thesis suggest that facing dyads are automatically grouped into social wholes. What are the limits of our studies? What alternative interpretation may be proposed? What happens next?

### 3.1. The “attentional hotspot” hypothesis

In several occasions, the phenomena we observed did not generalize to other classes of objects, indicating that they are specific to social objects. In Chapter 1, intermodulation was found with facing bodies, but not facing objects. In Chapter 2, infants did not discriminate relations between bodies and plants.

In contrast, other lines of research have argued that relations between non-social objects can affect how they are perceived. Vestner et al. (2020; 2021; 2022) argued that facingness can facilitate the visual processing of any object with a clear antero-posterior organization (e.g. cars, fans, arrows, etc.). They proposed that, when organized in facing pairs, those objects direct attention in-between them, generating an “attentional hotspot” that captures attention. However, attentional capture in visual search experiments was found with some categories of facing objects (e.g. cars, fans, bicycles), it failed with other categories with similar antero-posterior organizations and pointing directions (e.g. shoes, guns, chairs) (Vestner et al., 2020; 2021; 2022). Since it is unclear how this effect generalizes, it is unclear what produces it; having an anterior-posterior structure does not seem to be sufficient. Most critically, as detailed below, the attentional hotspot hypothesis does not fit the data in the present thesis.

In Chapter 1, different patterns were observed for bodies on the one hand, and chairs and machines on the other hand, despite that all stimuli had an antero-posterior structure. There was no hint of intermodulation oscillations for chairs or machines. An exploratory analysis (Appendix A – Supplementary material 1.1) more directly addressed the attentional hypothesis, by measuring alpha band suppression, a common marker of attentional enhancement (Klimesch, 1999; Klimesch et al., 1998; Sauseng et al., 2005). This analysis found lower alpha levels with facing bodies and objects compared to non-facing bodies and objects, indicating that facingness could enhance attention for both types of stimuli, consistent with the attentional hotspot hypothesis. In sum, facing objects may indeed create an attentional hotspot, but this effect does not necessarily pattern with intermodulation oscillations, and thus, cannot account for the specific integrated representation of facing bodies.

In Chapter 2, the attentional hotspot hypothesis critically failed to predict infants’ looking times. Indeed, despite being already sensitive to directional cues provided by heads, gaze or hands (Butterworth & Itakura, 2000; Farroni et al., 2004; Reid et al., 2004; Scaife & Bruner, 1975), 6-months-old infants looked longer at *non-facing dyads* than at facing dyads. Moreover, facing dyads failed to capture attention, as analyzing first fixations found that infants did not orient their first look on facing dyads but equally between both types of stimuli (Appendix B – Supplementary Table S2).

In Chapter 3, if the attentional hotspot enhancement could successfully account for longer looking times at facing dyads in adults and 5-year-olds, it fails to account for the behavior of younger populations similarly to how it fails in Chapter 2. Again, it is unclear why the structure of bodies, to which infants are sensitive since 3.5 months (Zieber et al., 2015), would fail to direct attention before 5 years of development when directional cues provided by heads, gaze or hands already direct the attention of infants and even neonates (Butterworth & Itakura, 2000; Farroni et al., 2004; Reid et al., 2004; Scaife & Bruner, 1975).

### 3.2. Integrating visual and social information

How does a visuo-spatial relation like facingness become a social information? What mechanism may attach a social value to the grouped percept of facing bodies?

Chapter 3 suggests that facingness adds value and meaning to body dyads, consistent with research showing how static visual features such as spatial proximity, relative positioning, or body postures speak to human observers, who readily interpret social relations from them, assign roles and intentions to the interacting people, and identify the type of interaction going on (Glanemann *et al.*, 2016; Hafri *et al.*, 2013; 2018; Isik *et al.*, 2020; Zhou *et al.*, 2019). The proficiency of humans in doing so contrasts with the difficulty of up-to-date artificial models of vision, like feedforward neural networks expert at object recognition, in recognizing the structure of social scenes such as relations between people (Ben-Yosef *et al.*, 2017; Lake *et al.*, 2017; Summerfield *et al.*, 2020). This highlights the lack of a computational account on how visual features cueing relations, such as facingness, allow a system to classify such relations. Recent advances in this direction have integrated a stage for the explicit processing of visuo-spatial relations between social agents (e.g., facingness) in a visual neural network (i.e., graph neural network model), and shown a significant improvement towards human-like interaction recognition in the network's performance (Malik & Isik, 2022). The two-step process proposed in the discussion of this thesis may guide future investigations on this issue. In particular, the second step suggests that understanding social relations needs processes beyond integrating visual information about the bodies and spatial relations of people. In agreement with this, Isik *et al.* (2020) found that meaningful social interactions are represented neurally at a slower speed than predicted by feedforward models of object recognition, suggesting additional iterative processes.

## 4. Conclusion

This thesis addressed how humans represent socially-relevant relations between social agents (people). We focused on facingness, a visuo-spatial relation which supports mutual perceptual accessibility between two agents. We showed that the individuals represent facing people as wholes, but not non-facing people (Chapter 1). Infants as young as 6 months show sensitivity to spatial relations between people and seem to process facing dyads like single visual units as if they grouped them like adults do (Chapter 2). This visual sensitivity slowly evolves into a visual preference by 5 years of life, a preference that gregarious non-human primates show as well, and may signal the recognition of social interaction from the mere face-to-face body positioning of social agents (Chapter 3). To conclude, experimental data were discussed in a model proposing that visual perception supports the representation of social interaction, by grouping individual persons into a holistic percept, available for later socio-cognitive processes encoding its social meaning.

## Bibliography

- Abassi, E., & Papeo, L. (2020). The Representation of Two-Body Shapes in the Human Visual Cortex. *Journal of Neuroscience*, 40(4), 852-863. <https://doi.org/10.1523/JNEUROSCI.1378-19.2019>
- Abassi, E., & Papeo, L. (2022). Behavioral and neural markers of visual configural processing in social scene perception. *NeuroImage*, 260, 119506. <https://doi.org/10.1016/j.neuroimage.2022.119506>
- Adibpour, P., Hochmann, J.-R., & Papeo, L. (2021). Spatial Relations Trigger Visual Binding of People. *Journal of Cognitive Neuroscience*, 33(7), 1343-1353. [https://doi.org/10.1162/jocn\\_a\\_01724](https://doi.org/10.1162/jocn_a_01724)
- Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, 11(2), 231-239. [https://doi.org/10.1016/S0959-4388\(00\)00202-6](https://doi.org/10.1016/S0959-4388(00)00202-6)
- Adrian, E. D., & Matthews, B. H. C. (1934). The Berger rhythm: Potential changes from the occipital lobes in man. *Brain*, 57(4), 355-385. <https://doi.org/10.1093/brain/57.4.355>
- Aissani, C., Cottureau, B., Dumas, G., Paradis, A.-L., & Lorenceau, J. (2011). Magnetoencephalographic signatures of visual form and motion binding. *Brain Research*, 1408, 27-40. <https://doi.org/10.1016/j.brainres.2011.05.051>
- Alonso-Prieto, E., Belle, G. V., Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2013). The 6Hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex. *Neuropsychologia*, 51(13), 2863-2875. <https://doi.org/10.1016/j.neuropsychologia.2013.08.018>
- Alp, N., Kogo, N., Van Belle, G., Wagemans, J., & Rossion, B. (2016). Frequency tagging yields an objective neural signature of Gestalt formation. *Brain and Cognition*, 104, 15-24. <https://doi.org/10.1016/j.bandc.2016.01.008>
- Alp, N., Nikolaev, A. R., Wagemans, J., & Kogo, N. (2017). EEG frequency tagging dissociates between neural processing of motion synchrony and human quality of multiple point-light dancers. *Scientific Reports*, 7(1), 44012. <https://doi.org/10.1038/srep44012>
- Anderson, S. F., & Kelley, K. (2020). Bias-and Uncertainty-Corrected Sample Size. R package version 1.2. 1. <https://CRAN.R-project.org/package=BUCSS>
- Appelbaum, L. G., Wade, A. R., Pettet, M. W., Vildavski, V. Y., & Norcia, A. M. (2008). Figure-ground interaction in the human visual cortex. *Journal of Vision*, 8(9), 8. <https://doi.org/10.1167/8.9.8>
- Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48-53. <https://doi.org/10.1111/j.1467-7687.2007.00563.x>
- Augusti, E.-M., Melinder, A., & Gredebäck, G. (2010). Look Who's Talking: Pre-Verbal Infants' Perception of Face-to-Face and Back-to-Back Social Interactions. *Frontiers in Psychology*, 1. <https://doi.org/10.3389/fpsyg.2010.00161>
- Aviezer, H., Trope, Y., & Todorov, A. (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, 338(6111), 1225-1229. <https://doi.org/10.1126/science.1224313>
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, 20(3), 191-208. [https://doi.org/10.1016/0010-0277\(85\)90008-3](https://doi.org/10.1016/0010-0277(85)90008-3)

- Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: Role of dynamic and configural properties: Biological motion preference. *Developmental Science*, *14*(2), 353-359. <https://doi.org/10.1111/j.1467-7687.2010.00985.x>
- Baron-Cohen, S. (1997). How to build a baby that can read minds: Cognitive mechanisms in mindreading. In *The Maladapted Mind: Classic Readings in Evolutionary Psychopathology* (p. 207-239). Psychology Press.
- Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., & Williams, S. C. R. (2000). The amygdala theory of autism. *Neuroscience & Biobehavioral Reviews*, *24*(3), 355-364. [https://doi.org/10.1016/S0149-7634\(00\)00011-7](https://doi.org/10.1016/S0149-7634(00)00011-7)
- Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology*, *33*(6), 680-685. <https://doi.org/10.1037/0022-3514.33.6.680>
- Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59-78. <https://doi.org/10.1017/S0140525X11002202>
- Bavelas, J. B., & Chovil, N. (2000). Visible Acts of Meaning: An Integrated Message Model of Language in Face-to-Face Dialogue. *Journal of Language and Social Psychology*, *19*(2), 163-194. <https://doi.org/10.1177/0261927X00019002001>
- Beck, A. A., Rossion, B., & Samson, D. (2018). An objective neural signature of rapid perspective taking. *Social Cognitive and Affective Neuroscience*, *13*(1), 72-79. <https://doi.org/10.1093/scan/nsx135>
- Beier, J. S., & Spelke, E. S. (2012). Infants' Developing Understanding of Social Gaze. *Child Development*, *83*(2), 486-496. <https://doi.org/10.1111/j.1467-8624.2011.01702.x>
- Bellot, E., Abassi, E., & Papeo, L. (2021). Moving Toward versus Away from Another: How Body Motion Direction Changes the Representation of Bodies and Actions in the Visual Cortex. *Cerebral Cortex*, *31*(5), 2670-2685. <https://doi.org/10.1093/cercor/bhaa382>
- Ben-Yosef, G., Yachin, A., & Ullman, S. (2017, mars 20). A Model for Interpreting Social Interactions in Local Image Regions. *2017 AAAI Spring Symposium Series*. 2017 AAAI Spring Symposium Series. <https://www.aaai.org/ocs/index.php/SSS/SSS17/paper/view/15354>
- Bhatt, R. S., Hock, A., White, H., Jubran, R., & Galati, A. (2016). The Development of Body Structure Knowledge in Infancy. *Child Development Perspectives*, *10*(1), 45-52. <https://doi.org/10.1111/cdep.12162>
- Birmingham, E., & Kingstone, A. (2009). Human social attention. In *Progress in Brain Research* (Vol. 176, p. 309-320). Elsevier. [https://doi.org/10.1016/S0079-6123\(09\)17618-5](https://doi.org/10.1016/S0079-6123(09)17618-5)
- Boremanse, A., Norcia, A. M., & Rossion, B. (2013). An objective signature for visual binding of face parts in the human brain. *Journal of Vision*, *13*(11), 6-6. <https://doi.org/10.1167/13.11.6>
- Boremanse, A., Norcia, A. M., & Rossion, B. (2014). Dissociation of part-based and integrated neural responses to faces by means of electroencephalographic frequency tagging. *European Journal of Neuroscience*, *40*(6), 2987-2997. <https://doi.org/10.1111/ejn.12663>
- Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159-166. <https://doi.org/10.1016/j.tics.2006.02.004>

- Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, 18(1), 25-50. <https://doi.org/10.1348/026151000165553>
- Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology: Issues, News, and Reviews*, 5(5), 172-180. [https://doi.org/10.1002/\(SICI\)1520-6505\(1996\)5:5<172::AID-EVAN6>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1520-6505(1996)5:5<172::AID-EVAN6>3.0.CO;2-H)
- Carey, S. (2009). *The origin of concepts*. Oxford university press.
- Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development*, 21(2), 315-330.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893-910. <https://doi.org/10.1037/0022-3514.76.6.893>
- Cashon, C. H., & Cohen, L. B. (2004). Beyond U-Shaped Development in Infants' Processing of Faces: An Information-Processing Account. *Journal of Cognition and Development*, 5(1), 59-80. [https://doi.org/10.1207/s15327647jcd0501\\_4](https://doi.org/10.1207/s15327647jcd0501_4)
- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage*, 12(3), 314-325. <https://doi.org/10.1006/nimg.2000.0612>
- Centelles, L., Assaiante, C., Nazarian, B., Anton, J.-L., & Schmitz, C. (2011). Recruitment of Both the Mirror and the Mentalizing Networks When Observing Social Interactions Depicted by Point-Lights: A Neuroimaging Study. *PLOS ONE*, 6(1), e15749. <https://doi.org/10.1371/journal.pone.0015749>
- Cheney, D. L., Seyfarth, R. M., & Smuts, B. (1986). Social relationships and social cognition in nonhuman primates. *Science (New York, N.Y.)*, 234(4782), 1361-1366.
- Cloutier, J., Cardenas-Iniguez, C., Gyurovski, I., Barakzai, A., & Li, T. (2016). Neuroimaging Investigations of Social Status and Social Hierarchies. In *Neuroimaging Personality, Social Cognition, and Character* (p. 187-203). Elsevier. <https://doi.org/10.1016/B978-0-12-800935-2.00009-9>
- Cracco, E., Lee, H., van Belle, G., Quenon, L., Haggard, P., Rossion, B., & Orgs, G. (2021). EEG Frequency Tagging Reveals the Integration of Form and Motion Cues into the Perception of Group Movement. *Cerebral Cortex*, bhab385. <https://doi.org/10.1093/cercor/bhab385>
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 358(1431), 447-458. <https://doi.org/10.1098/rstb.2002.1235>
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 1149-1157. <https://doi.org/10.1098/rstb.2010.0319>
- Decety, J., & Cacioppo, S. (2012). The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology*, 108(11), 3068-3072. <https://doi.org/10.1152/jn.00473.2012>
- Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, 18(1), 193-222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>

- de Waal, F. B. M., & Luttrell, L. M. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology*, *9*(2), 101-118.  
[https://doi.org/10.1016/0162-3095\(88\)90016-7](https://doi.org/10.1016/0162-3095(88)90016-7)
- Di Giorgio, E., Frasnelli, E., Salva, O. R., Scattoni, M. L., Puopolo, M., Tosoni, D., Simion, F., & Vallortigara, G. (2016). Difference in Visual Social Predispositions Between Newborns at Low- and High-risk for Autism. *Scientific Reports*, *6*(1), 1-9. <https://doi.org/10.1038/srep26395>
- Dima, D. C., Tomita, T. M., Honey, C. J., & Isik, L. (2022). Social-affective features drive human representations of observed actions. *Elife*, *11*, e75027. <https://doi.org/10.7554/eLife.75027>
- Ding, X., Gao, Z., & Shen, M. (2017). Two Equals One: Two Human Actions During Social Interaction Are Grouped as One Unit in Working Memory. *Psychological Science*, *28*(9), 621 1311–1320.  
<https://doi.org/10.1177/0956797617707318>
- Dink, J., & Ferguson, B. (2018). eyetrackingR. *R package version 0.1*, 7.
- Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93*(1), B27-B38. <https://doi.org/10.1016/j.cognition.2003.10.010>
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science, New Series*, *293*(5539), 2470-2473.
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509-540. <https://doi.org/10.1080/135062899394920>
- Dunbar, R. I. M., Duncan, N. D. C., & Nettle, D. (1995). Size and structure of freely forming conversational groups. *Human Nature*, *6*(1), 67-78. <https://doi.org/10.1007/BF02734136>
- Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science (New York, N.Y.)*, *317*(5843), 1344-1347. <https://doi.org/10.1126/science.1145463>
- End, A., & Gamer, M. (2017). Preferential Processing of Social Features and Their Interplay with Physical Saliency in Complex Naturalistic Scenes. *Frontiers in Psychology*, *8*.  
<https://www.frontiersin.org/article/10.3389/fpsyg.2017.00418>
- Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, *102*(47), 17245-17250. <https://doi.org/10.1073/pnas.0502205102>
- Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze Following in Newborns. *Infancy*, *5*(1), 39-60.  
[https://doi.org/10.1207/s15327078in0501\\_2](https://doi.org/10.1207/s15327078in0501_2)
- Farrow, T. F. D., Jones, S. C., Kaylor-Hughes, C. J., Wilkinson, I. D., Woodruff, P. W. R., Hunter, M. D., & Spence, S. A. (2011). Higher or lower? The functional anatomy of perceived allocentric social hierarchies. *NeuroImage*, *57*(4), 1552-1560. <https://doi.org/10.1016/j.neuroimage.2011.05.069>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.  
<https://doi.org/10.3758/BF03193146>

- Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101-107. <https://doi.org/10.1016/j.cognition.2016.03.005>
- Fawcett, C., & Tunçgenç, B. (2017). Infants' use of movement synchrony to infer social affiliation in others. *Journal of Experimental Child Psychology*, *160*, 127-136. <https://doi.org/10.1016/j.jecp.2017.03.014>
- Ferrari, P. F., & Gallese, V. (2007). Mirror neurons and intersubjectivity. *On Being Moved: From Mirror Neurons to Empathy*, 68. <https://doi.org/10.1075/aicr.68.08fer>
- Ferrari, P. F., Paukner, A., Ionica, C., & Suomi, S. J. (2009). Reciprocal Face-to-Face Communication between Rhesus Macaque Mothers and Their Newborn Infants. *Current Biology*, *19*(20), 1768-1772. <https://doi.org/10.1016/j.cub.2009.08.055>
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490-495. <https://doi.org/10.3758/BF03208827>
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*, 694-724. <https://doi.org/10.1037/0033-2909.133.4.694>
- Frith, C. d., Wolpert, D. m., Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 459-473. <https://doi.org/10.1098/rstb.2002.1218>
- Galazka, M. A., Roché, L., Nyström, P., & Falck-Ytter, T. (2014). Human Infants Detect Other People's Interactions Based on Complex Patterns of Kinematic Information. *PLOS ONE*, *9*(11), e112432. <https://doi.org/10.1371/journal.pone.0112432>
- Gao, T., & Scholl, B. J. (2011). Chasing vs. stalking: Interrupting the perception of animacy. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 669-684. <https://doi.org/10.1037/a0020735>
- Gazes, R. P., Hampton, R. R., & Lourenco, S. F. (2017). Transitive inference of social dominance by human infants. *Developmental Science*, *20*(2). <https://doi.org/10.1111/desc.12367>
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287-292. [https://doi.org/10.1016/S1364-6613\(03\)00128-1](https://doi.org/10.1016/S1364-6613(03)00128-1)
- Gergely, G., & Csibra, G. (2013). Natural pedagogy. *Navigating the social world: What infants, children, and other species can teach us*, 127-132.
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165-193. [https://doi.org/10.1016/0010-0277\(95\)00661-H](https://doi.org/10.1016/0010-0277(95)00661-H)
- Gibson, J. J. (1961). Ecological optics. *Vision Research*, *1*(3), 253-262. [https://doi.org/10.1016/0042-6989\(61\)90005-0](https://doi.org/10.1016/0042-6989(61)90005-0)
- Glanemann, R., Zwitserlood, P., Bölte, J., & Dobel, C. (2016). Rapid apprehension of the coherence of action scenes. *Psychonomic Bulletin & Review*, *23*(5), 1566-1575. <https://doi.org/10.3758/s13423-016-1004-y>
- Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00527>

- Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two Takes on the Social Brain: A Comparison of Theory of Mind Tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803-1814. <https://doi.org/10.1162/jocn.2007.19.11.1803>
- Gordon, N., Hohwy, J., Davidson, M. J., van Boxtel, J. J. A., & Tsuchiya, N. (2019). From intermodulation components to visual perception and cognition-a review. *NeuroImage*, *199*, 480-494. <https://doi.org/10.1016/j.neuroimage.2019.06.008>
- Goupil, N., Hochmann, J.-R., & Papeo, L. (2023). *Intermodulation responses show integration of interacting bodies in a new whole*. OSF Preprints. <https://doi.org/10.31219/osf.io/q9e8a>
- Goupil, N., Kaiser, D., & Papeo, L. (2023). *Category-specific effects of high-level relations in visual search*. OSF Preprints. <https://doi.org/10.31219/osf.io/n7uvk>
- Goupil, N., Papeo, L., & Hochmann, J.-R. (2022). Visual perception grounding of social cognition in preverbal infants. *Infancy*, *27*(2), 210-231. <https://doi.org/10.1111/infa.12453>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. <https://doi.org/10.3389/fnins.2013.00267>
- Graziano, M. S. A., & Kastner, S. (2011). Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cognitive neuroscience*, *2*(2), 98-113. <https://doi.org/10.1080/17588928.2011.565121>
- Green, C., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1107–1119. <https://doi.org/10.1037/0096-1523.32.5.1107>
- Grossmann, T. (2021). Developmental Origins of the Pathway for Social Perception. *Trends in Cognitive Sciences*, *25*(7), 546-547. <https://doi.org/10.1016/j.tics.2021.03.003>
- Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain Areas Involved in Perception of Biological Motion. *Journal of Cognitive Neuroscience*, *12*(5), 711-720. <https://doi.org/10.1162/089892900562417>
- Gwiazda, J., Bauer, J., Thorn, F., & Held, R. (1997). Development of Spatial Contrast Sensitivity from Infancy to Adulthood: Psychophysical Data. *Optometry and Vision Science*, *74*(10), 785-789. <https://doi.org/10.1097/00006324-199710000-00017>
- Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880-905. <https://doi.org/10.1037/a0030045>
- Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*, 36-52. <https://doi.org/10.1016/j.cognition.2018.02.011>
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557-559. <https://doi.org/10.1038/nature06288>

- Handl, A., Mahlberg, T., Norling, S., & Gredebäck, G. (2013). Facing still faces: What visual cues affect infants' observations of others? *Infant Behavior and Development*, 36(4), 583-586. <https://doi.org/10.1016/j.infbeh.2013.06.001>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, 57(2), 243. <https://doi.org/10.2307/1416950>
- Heilbron, M., Richter, D., Ekman, M., Hagoort, P., & de Lange, F. P. (2020). Word contexts enhance the neural representation of individual letters in early visual cortex. *Nature Communications*, 11(1), 321. <https://doi.org/10.1038/s41467-019-13996-4>
- Hesse, E., Mikulan, E., Decety, J., Sigman, M., Garcia, M. del C., Silva, W., Ciraolo, C., Vaucheret, E., Baglivo, F., Huepe, D., Lopez, V., Manes, F., Bekinschtein, T. A., & Ibanez, A. (2016). Early detection of intentional harm in the human amygdala. *Brain*, 139(1), 54-61. <https://doi.org/10.1093/brain/awv336>
- Hirai, M., & Kakigi, R. (2009). Differential orientation effect in the neural response to interacting biological motion of two agents. *BMC Neuroscience*, 10(1), 39. <https://doi.org/10.1186/1471-2202-10-39>
- Hochmann, J.-R., & Kouider, S. (2022). Acceleration of information processing en route to perceptual awareness in infancy. *Current Biology*. <https://doi.org/10.1016/j.cub.2022.01.029>
- Hochmann, J.-R., & Papeo, L. (2014). The Invariance Problem in Infancy: A Pupillometry Study. *Psychological Science*, 25(11), 2038-2046. <https://doi.org/10.1177/0956797614547918>
- Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3(1), 80-84. <https://doi.org/10.1038/71152>
- Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*, 21(3), 1167-1173. <https://doi.org/10.1016/j.neuroimage.2003.11.013>
- Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, 114(43), E9145-E9152. <https://doi.org/10.1073/pnas.1714471114>
- Isik, L., Mynick, A., Pantazis, D., & Kanwisher, N. (2020). The speed of human social interaction perception. *NeuroImage*, 215, 116844. <https://doi.org/10.1016/j.neuroimage.2020.116844>
- James, J. (1951). A Preliminary Study of the Size Determinant in Small Group Interaction. *American Sociological Review*, 16(4), 474-477. <https://doi.org/10.2307/2088278>
- Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The Faces in Infant-Perspective Scenes Change over the First Year of Life. *PLOS ONE*, 10(5), e0123780. <https://doi.org/10.1371/journal.pone.0123780>
- Jeannerod, M. (2001). Neural Simulation of Action: A Unifying Mechanism for Motor Cognition. *NeuroImage*, 14(1), S103-S109. <https://doi.org/10.1006/nimg.2001.0832>

- Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural Synchronization during Face-to-Face Communication. *Journal of Neuroscience*, 32(45), 16064-16069. <https://doi.org/10.1523/JNEUROSCI.2926-12.2012>
- Jin, K.-S., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences of the United States of America*, 114(31), 8199-8204. <https://doi.org/10.1073/pnas.1706286114>
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2), 201-211. <https://doi.org/10.3758/BF03212378>
- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object Vision in a Structured World. *Trends in Cognitive Sciences*, 23(8), 672-685. <https://doi.org/10.1016/j.tics.2019.04.013>
- Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences*, 111(30), 11217-11222. <https://doi.org/10.1073/pnas.1400559111>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, 17(11), 4302-4311. <https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. *Psychological Science*, 18(12), 1084-1089. <https://doi.org/10.1111/j.1467-9280.2007.02029.x>
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLOS ONE*, 7(5), e36399. <https://doi.org/10.1371/journal.pone.0036399>
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2), 169-195. [https://doi.org/10.1016/S0165-0173\(98\)00056-3](https://doi.org/10.1016/S0165-0173(98)00056-3)
- Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience Letters*, 244(2), 73-76. [https://doi.org/10.1016/S0304-3940\(98\)00122-0](https://doi.org/10.1016/S0304-3940(98)00122-0)
- Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and Constraints in Causal Perception. *Psychological Science*, 28(11), 1649-1662. <https://doi.org/10.1177/0956797617719930>
- Kujala, M. V., Carlson, S., & Hari, R. (2012). Engagement of amygdala in third-person view of face-to-face interaction. *Human Brain Mapping*, 33(8), 1753-1762. <https://doi.org/10.1002/hbm.21317>
- Kuwahata, H., Adachi, I., Fujita, K., Tomonaga, M., & Matsuzawa, T. (2004). Development of schematic face preference in macaque monkeys. *Behavioural Processes*, 66(1), 17-21. <https://doi.org/10.1016/j.beproc.2003.11.002>
- Lahnakoski, J. M., Forbes, P. A. G., McCall, C., & Schilbach, L. (2020). Unobtrusive tracking of interpersonal orienting and distance predicts the subjective quality of social interactions. *Royal Society Open Science*, 7(8), 191815. <https://doi.org/10.1098/rsos.191815>

- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, e253. <https://doi.org/10.1017/S0140525X16001837>
- Landsiedel, J., Daughters, K., Downing, P. E., & Koldewyn, K. (2022). The role of motion in the neural representation of social interactions in the posterior temporal cortex. *NeuroImage*, 262, 119533. <https://doi.org/10.1101/2022.08.01.502134>
- Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, 107(1), 330-342. <https://doi.org/10.1016/j.cognition.2007.07.012>
- Lau, T., Gershman, S. J., & Cikara, M. (2020). Social structure learning in human anterior insula. *eLife*, 9, e53162. <https://doi.org/10.7554/eLife.53162>
- Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. <https://CRAN.R-project.org/package=ez>
- Leslie, A. M. (1982). The Perception of Causality in Infants. *Perception*, 11(2), 173-186. <https://doi.org/10.1068/p110173>
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3), 265-288. [https://doi.org/10.1016/S0010-0277\(87\)80006-9](https://doi.org/10.1016/S0010-0277(87)80006-9)
- Lewkowicz, D. J., & Ghazanfar, A. A. (2012). The Development of the Uncanny Valley in Infants. *Developmental Psychobiology*, 54(2), 124-132. <https://doi.org/10.1002/dev.20583>
- Liberman, Z., Kinzler, K. D., & Woodward, A. L. (2014). Friends or foes: Infants use shared evaluations to infer others' social relationships. *Journal of experimental psychology. General*, 143(3), 966-971. <https://doi.org/10.1037/a0034481>
- Liberman, Z., Kinzler, K. D., & Woodward, A. L. (2021). Origins of homophily: Infants expect people with shared preferences to affiliate. *Cognition*, 212, 104695. <https://doi.org/10.1016/j.cognition.2021.104695>
- Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia*, 52, 57-72. <https://doi.org/10.1016/j.neuropsychologia.2013.10.022>
- Lloyd, D. M., & Morrison, C. I. (2008). 'Eavesdropping' on social interactions biases threat perception in visuospatial pathways. *Neuropsychologia*, 46(1), 95-101. <https://doi.org/10.1016/j.neuropsychologia.2007.08.007>
- Macrae, C. N., & Quadflieg, S. (2010). Perceiving people. In *Handbook of social psychology* (p. 428-463). John Wiley & Sons Inc.
- Malik, M., & Isik, L. (2022, November 3). Relational Visual Information explains Human Social Inference: A Graph Neural Network model for Social Interaction Recognition. <https://doi.org/10.31234/osf.io/5cuyr>
- Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences*, 115(38), E8835-E8843. <https://doi.org/10.1073/pnas.1801677115>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177-190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>

- Marr, D. (1982). The Philosophy and the Approach. In *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Martin, A., & Weisberg, J. (2003). Neural Foundations for Understanding Social and Mechanical Concepts. *Cognitive Neuropsychology*, 20(3-6), 575-587. <https://doi.org/10.1080/02643290342000005>
- Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, 109(18), 6862-6867. <https://doi.org/10.1073/pnas.1113194109>
- Mascaro, O., & Csibra, G. (2014). Human Infants' Learning of Social Structures: The Case of Dominance Hierarchy. *Psychological science*, 25(1), 250-255. <https://doi.org/10.1177/0956797613500509>
- Matsuno, T. (2021). Visual Body Perception in Primates: From Individual to Social Dyad. In J. R. Anderson & H. Kuroshima (Éds.), *Comparative Cognition: Commonalities and Diversity* (p. 83-103). Springer. [https://doi.org/10.1007/978-981-16-2028-7\\_6](https://doi.org/10.1007/978-981-16-2028-7_6)
- Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255-260. [https://doi.org/10.1016/S1364-6613\(02\)01903-4](https://doi.org/10.1016/S1364-6613(02)01903-4)
- Mazur, A. (1985). A Biosocial Model of Status in Face-to-Face Primate Groups\*. *Social Forces*, 64(2), 377-402. <https://doi.org/10.1093/sf/64.2.377>
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90(3), 215-238. <https://doi.org/10.1037/0033-295X.90.3.215>
- McGregor, P. K., Butlin, R. K., Guilford, T., & Krebs, J. R. (1993). Signalling in territorial systems: A context for individual identification, ranging and eavesdropping. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 340(1292), 237-244. <https://doi.org/10.1098/rstb.1993.0063>
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science, New Series*, 198(4312), 75-78.
- Mersad, K., & Caristan, C. (2021). Blending into the Crowd: Electrophysiological Evidence of Gestalt Perception of a Human Dyad. *Neuropsychologia*, 160, 107967. <https://doi.org/10.1016/j.neuropsychologia.2021.107967>
- Michotte, A. (1946/1963). *The perception of causality*, Methuen, London.
- Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, 98, 164-181. <https://doi.org/10.1037/0033-295X.98.2.164>
- Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, 61(2), 63-86. <https://doi.org/10.1016/j.cogpsych.2010.02.001>
- Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current directions in psychological science*, 16(5), 269-274. <https://doi.org/10.1111/j.1467-8721.2007.00518.x>
- Muschinski, J., Feczko, E., Brooks, J. M., Collantes, M., Heitz, T. R., & Parr, L. A. (2016). The development of visual preferences for direct versus averted gaze faces in infant macaques (*Macaca mulatta*). *Developmental Psychobiology*, 58(8), 926-936. <https://doi.org/10.1002/dev.21421>

- Nakayama, K. (2010). Introduction: Vision Going Social. *The Science of Social Vision*, xv-xxix.
- Nazzi, T., Bertoni, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756-766. <https://doi.org/10.1037//0096-1523.24.3.756>
- Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, 9(9), 1186-1192. <https://doi.org/10.1038/nn1759>
- Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, 395(6705), 894-896. <https://doi.org/10.1038/27661>
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, 104(42), 16598-16603. <https://doi.org/10.1073/pnas.0703913104>
- Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 107(40), 17140-17145. <https://doi.org/10.1073/pnas.0914056107>
- Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottureau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, 15(6). <https://doi.org/10.1167/15.6.4>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087-11092. <https://doi.org/10.1073/pnas.0805664105>
- Paparella, I., & Papeo, L. (2022). Chunking by social relationship in working memory. *Visual Cognition*, 30(5), 354-370. <https://doi.org/10.1080/13506285.2022.2064950>
- Papeo, L. (2020). Twos in human visual perception. *Cortex*, 132, 473-478. <https://doi.org/10.1016/j.cortex.2020.06.005>
- Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human-human interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 45(7), 877-888. <https://doi.org/10.1037/xhp0000646>
- Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual Search for People Among People. *Psychological Science*, 0956797619867295. <https://doi.org/10.1177/0956797619867295>
- Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychological Science*, 28(3), 369-379. <https://doi.org/10.1177/0956797616685769>
- Paradise, R., & Rogoff, B. (2009). Side by Side: Learning by Observing and Pitching In. *Ethos*, 37(1), 102-138. <https://doi.org/10.1111/j.1548-1352.2009.01033.x>
- Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, 8(8), 636-648. <https://doi.org/10.1038/nrn2195>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195-203. <https://doi.org/10.3758/s13428-018-01193-y>

- Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, 25(2), 100-110. <https://doi.org/10.1016/j.tics.2020.11.006>
- Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14), 5297-5300. <https://doi.org/10.1073/pnas.0406627102>
- Poljac, E., de-Wit, L., & Wagemans, J. (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition*, 123(2), 308-312. <https://doi.org/10.1016/j.cognition.2012.01.001>
- Powell, L. (2021). *Adopted utility calculus: Origins of a concept of social affiliation*. PsyArXiv. <https://doi.org/10.31234/osf.io/kuwgf>
- Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41), E3965-E3972. <https://doi.org/10.1073/pnas.1304326110>
- Powell, L. J., & Spelke, E. S. (2018a). Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition*, 170, 31-48. <https://doi.org/10.1016/j.cognition.2017.09.007>
- Powell, L. J., & Spelke, E. S. (2018b). Third-Party Preferences for Imitators in Preverbal Infants. *Open Mind*, 2(2), 61-71. [https://doi.org/10.1162/opmi\\_a\\_00018](https://doi.org/10.1162/opmi_a_00018)
- Poyo Solanas, M., Vaessen, M. J., & de Gelder, B. (2020). The role of computational and subjective features in emotional body expressions. *Scientific Reports*, 10(1), 6202. <https://doi.org/10.1038/s41598-020-63125-1>
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, 36(1), 1-16. [https://doi.org/10.1016/0010-0277\(90\)90051-K](https://doi.org/10.1016/0010-0277(90)90051-K)
- Premack, D., & Premack, A. J. (1997). Infants Attribute Value± to the Goal-Directed Actions of Self-propelled Objects. *Journal of Cognitive Neuroscience*, 9(6), 848-856. <https://doi.org/10.1162/jocn.1997.9.6.848>
- Pun, A., Birch, S. A. J., & Baron, A. S. (2016). Infants use relative numerical group size to infer social dominance. *Proceedings of the National Academy of Sciences of the United States of America*, 113(9), 2376-2381. <https://doi.org/10.1073/pnas.1514879113>
- Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: Configuration enhances detection. *Perception & Psychophysics*, 43(4), 355-366. <https://doi.org/10.3758/BF03208806>
- Quadflieg, S., Gentile, F., & Rossion, B. (2015). The neural basis of perceiving person interactions. *Cortex*, 70, 5-20. <https://doi.org/10.1016/j.cortex.2014.12.020>
- Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Annals of the New York Academy of Sciences*, 1396(1), 166-182. <https://doi.org/10.1111/nyas.13331>
- Quadflieg, S., & Penton-Voak, I. S. (2017). The Emerging Science of People-Watching: Forming Impressions From Third-Party Encounters. *Current Directions in Psychological Science*, 26(4), 383-389. <https://doi.org/10.1177/0963721417694353>

- Ratliff, F., & Zemon, V. (1982). Some New Methods for the Analysis of Lateral Interactions That Influence the Visual Evoked Potential\*. *Annals of the New York Academy of Sciences*, 388(1), 113-124. <https://doi.org/10.1111/j.1749-6632.1982.tb50787.x>
- Rayson, H., Massera, A., Belluardo, M., Ben Hamed, S., & Ferrari, P. F. (2021). Early social adversity modulates the relation between attention biases and socioemotional behaviour in juvenile macaques. *Scientific Reports*, 11(1), Art. 1. <https://doi.org/10.1038/s41598-021-00620-z>
- Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, 20(3), 238-248. [https://doi.org/10.1016/0013-4694\(66\)90088-5](https://doi.org/10.1016/0013-4694(66)90088-5)
- Regan, D., & Heron, J. R. (1969). Clinical investigation of lesions of the visual pathway: A new objective technique. *Journal of Neurology, Neurosurgery, and Psychiatry*, 32(5), 479-483.
- Regan, M. P., & Regan, D. (1988). A frequency domain technique for characterizing nonlinearities in biological systems. *Journal of Theoretical Biology*, 133(3), 293-317. [https://doi.org/10.1016/S0022-5193\(88\)80323-0](https://doi.org/10.1016/S0022-5193(88)80323-0)
- Regan, M. P., & Regan, D. (1989). Objective Investigation of Visual Function Using a Nondestructive Zoom-FFT Technique for Evoked Potential Analysis. *Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques*, 16(2), 168-179. <https://doi.org/10.1017/S0317167100028845>
- Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport*, 15(16), 2553-2555.
- Remland, M. S., Jones, T. S., & Brinkman, H. (1995). Interpersonal Distance, Body Orientation, and Touch: Effects of Culture, Gender, and Age. *The Journal of Social Psychology*, 135(3), 281-297. <https://doi.org/10.1080/00224545.1995.9713958>
- Retter, T. L., Rossion, B., & Schiltz, C. (2021). Harmonic Amplitude Summation for Frequency-tagging Analysis. *Journal of Cognitive Neuroscience*, 33(11), 2372-2393. [https://doi.org/10.1162/jocn\\_a\\_01763](https://doi.org/10.1162/jocn_a_01763)
- Rhodes, M., Hetherington, C., Brink, K., & Wellman, H. M. (2015). Infants' Use of Social Partnerships to Predict Behavior. *Developmental science*, 18(6), 909-916. <https://doi.org/10.1111/desc.12267>
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2(9), 661-670. <https://doi.org/10.1038/35090060>
- Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition*, 15, 322-348. <https://doi.org/10.1080/13506280600590434>
- Roberts, K. L., & Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *NeuroImage*, 52(4), 1541-1548. <https://doi.org/10.1016/j.neuroimage.2010.05.044>
- Robson, A. J., & Kaplan, H. S. (2003). The Evolution of Human Life Expectancy and Intelligence in Hunter-Gatherer Economies. *American Economic Review*, 93(1), 150-169. <https://doi.org/10.1257/000282803321455205>
- Rochat, P., Morgan, R., & Carpenter, M. (1997). Young infants' sensitivity to movement information specifying social causality. *Cognitive Development*, 12(4), 537-561. [https://doi.org/10.1016/S0885-2014\(97\)90022-8](https://doi.org/10.1016/S0885-2014(97)90022-8)

- Rochat, P., Striano, T., & Morgan, R. (2004). Who is Doing What to Whom? Young Infants' Developing Sense of Social Causality in Animated Displays. *Perception*, 33(3), 355-369. <https://doi.org/10.1068/p3389>
- Rossion, B., Retter, T. L., & Liu-Shuang, J. (2020). Understanding human individuation of unfamiliar faces with oddball fast periodic visual stimulation and electroencephalography. *European Journal of Neuroscience*, 52(10), 4283-4344. <https://doi.org/10.1111/ejn.14865>
- Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W. R., & Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience*, 22(11), 2917-2926. <https://doi.org/10.1111/j.1460-9568.2005.04482.x>
- Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret Agents: Inferences About Hidden Causes by 10- and 12-Month-Old Infants. *Psychological Science*, 16(12), 995-1001. <https://doi.org/10.1111/j.1467-9280.2005.01649.x>
- Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, 253(5489), Art. 5489. <https://doi.org/10.1038/253265a0>
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299-309. [https://doi.org/10.1016/S1364-6613\(00\)01506-0](https://doi.org/10.1016/S1364-6613(00)01506-0)
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 42, 9-34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70-76. <https://doi.org/10.1016/j.tics.2005.12.009>
- Skripkauskaitė, S., Mihai, I., & Koldewyn, K. (2022). Attentional bias towards social interactions during viewing of naturalistic scenes. *Quarterly Journal of Experimental Psychology*, 17470218221140880. <https://doi.org/10.1177/17470218221140879>
- Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, 105(2), 809-813. <https://doi.org/10.1073/pnas.0707021105>
- Simpson, E. A., Miller, G. M., Ferrari, P. F., Suomi, S. J., & Paukner, A. (2016). Neonatal imitation and early social experience predict gaze following abilities in infant monkeys. *Scientific Reports*, 6(1), Art. 1. <https://doi.org/10.1038/srep20233>
- Slaughter, V. (2021). Do Newborns Have the Ability to Imitate? *Trends in Cognitive Sciences*, 25(5), 377-387. <https://doi.org/10.1016/j.tics.2021.02.006>
- Slaughter, V., Heron, M., & Sim, S. (2002). Development of preferences for the human body shape in infancy. *Cognition*, 85(3), B71-B81. [https://doi.org/10.1016/S0010-0277\(02\)00111-7](https://doi.org/10.1016/S0010-0277(02)00111-7)
- Slaughter, V., & McConnell, D. (2003). Emergence of joint attention: Relationships between gaze following, social referencing, imitation, and naming in infancy. *The Journal of Genetic Psychology*, 164(1), 54-71. <https://doi.org/10.1080/00221320309597503>
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, 23(2), 196-204. <https://doi.org/10.1177/0956797611422072>

- Spelke, E. S. (1990). Principles of Object Perception. *Cognitive Science*, 14(1), 29-56.  
[https://doi.org/10.1207/s15516709cog1401\\_3](https://doi.org/10.1207/s15516709cog1401_3)
- Spokes, A. C., & Spelke, E. S. (2017). The cradle of social knowledge: Infants' reasoning about caregiving and affiliation. *Cognition*, 159, 102-116. <https://doi.org/10.1016/j.cognition.2016.11.008>
- Stahl, A. E., & Feigenson, L. (2014). Social Knowledge Facilitates Chunking in Infancy. *Child Development*, 85(4), 1477-1490. <https://doi.org/10.1111/cdev.12217>
- Striano, T., & Stahl, D. (2005). Sensitivity to triadic attention in early infancy. *Developmental Science*, 8(4), 333-343. <https://doi.org/10.1111/j.1467-7687.2005.00421.x>
- Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*, 184, 101717. <https://doi.org/10.1016/j.pneurobio.2019.101717>
- Sun, Y., Stein, T., Liu, W., Ding, X., & Nie, Q.-Y. (2017). Biphasic attentional orienting triggered by invisible social signals. *Cognition*, 168, 129-139. <https://doi.org/10.1016/j.cognition.2017.06.020>
- Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4), 901-913.  
<https://doi.org/10.1037/0096-1523.21.4.901>
- Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, 11(1), 3002. <https://doi.org/10.1038/s41467-020-16846-w>
- Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, 137, 47-62. <https://doi.org/10.1016/j.cognition.2014.12.007>
- Tatone, D., Hernik, M., & Csibra, G. (2021). Facilitation of object encoding in infants by the observation of giving. *Scientific Reports*, 11(1), 18305. <https://doi.org/10.1038/s41598-021-97910-3>
- Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. B. M. (2021). Social interaction targets enhance 13-month-old infants' associative learning. *Infancy*, 26(3), 409-422. <https://doi.org/10.1111/infa.12393>
- Thiele, M., Hepach, R., Michel, C., & Haun, D. B. M. (2021a). Infants' Preference for Social Interactions Increases from 7 to 13 Months of Age. *Child Development*, 92(6), 2577-2594. <https://doi.org/10.1111/cdev.13636>
- Thiele, M., Hepach, R., Michel, C., & Haun, D. B. M. (2021b). Observing others' joint attention increases 9-month-old infants' object encoding. *Developmental Psychology*, 57(6), 837-850. <https://doi.org/10.1037/dev0001189>
- Thomsen, L., Frankenhuys, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and Mighty: Preverbal Infants Mentally Represent Social Dominance. *Science (New York, N.Y.)*, 331(6016).  
<https://doi.org/10.1126/science.1199198>
- Tollefsen, D. (2005). Let's Pretend! Children and Joint Action. *Philosophy of the Social Sciences*, 35(1), 75-97.  
<https://doi.org/10.1177/0048393104271925>
- Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development*, 57(6), 1454-1463. JSTOR. <https://doi.org/10.2307/1130423>

- Trevarthen, C. (1974). Conversations with a two-month-old. In J. Raphael-Leff (Éd.), *Parent-Infant Psychodynamics* (1<sup>re</sup> éd., p. 25-34). Routledge. <https://doi.org/10.4324/9780429478154-4>
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior* (Ingle, D.J. et al., p. 549-586). MIT Press.
- van Renswoude, D. R., Raijmakers, M. E. J., Koornneef, A., Johnson, S. P., Hunnius, S., & Visser, I. (2018). Gazepath: An eye-tracking analysis tool that accounts for individual differences and data quality. *Behavior Research Methods*, 50(2), 834-852. <https://doi.org/10.3758/s13428-017-0909-3>
- Vergeer, M., Kogo, N., Nikolaev, A. R., Alp, N., Loozen, V., Schraepen, B., & Wagemans, J. (2018). EEG frequency tagging reveals higher order intermodulation components as neural markers of learned holistic shape representations. *Vision Research*, 152, 91-100. <https://doi.org/10.1016/j.visres.2018.01.007>
- Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, 200, 104270. <https://doi.org/10.1016/j.cognition.2020.104270>
- Vestner, T., Gray, K. L. H., & Cook, R. (2021). Visual search for facing and non-facing people: The effect of actor inversion. *Cognition*, 208, 104550. <https://doi.org/10.1016/j.cognition.2020.104550>
- Vestner, T., Over, H., Gray, K. L. H., & Cook, R. (2022). Objects that direct visuospatial attention produce the search advantage for facing dyads. *Journal of Experimental Psychology: General*, 151(1), 161-171. <https://doi.org/10.1037/xge0001067>
- Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S.-A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000545>
- Victor, J. D., & Conte, M. M. (2000). Two-frequency analysis of interactions elicited by Vernier stimuli. *Visual Neuroscience*, 17(6), 959-973. <https://doi.org/10.1017/S0952523800176151>
- Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, 112, 31-39. <https://doi.org/10.1016/j.neuropsychologia.2018.02.023>
- Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, 198, 296-302. <https://doi.org/10.1016/j.neuroimage.2019.05.027>
- Walbrin, J., Mihai, I., Landsiedel, J., & Koldewyn, K. (2020). Developmental changes in visual responses to social interactions. *Developmental Cognitive Neuroscience*, 42, 100774. <https://doi.org/10.1016/j.dcn.2020.100774>
- Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding Intentions in Social Interaction: The Role of the Anterior Paracingulate Cortex. *Journal of Cognitive Neuroscience*, 16(10), 1854-1863. <https://doi.org/10.1162/0898929042947838>
- Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and Precision: How Data Quality May Influence Key Dependent Variables in Infant Eye-Tracker Analyses. *Infancy*, 19(5), 427-460. <https://doi.org/10.1111/infa.12055>
- Weiner, K. S., & Gomez, J. (2021). Third Visual Pathway, Anatomy, and Cognition across Species. *Trends in Cognitive Sciences*, 25(7), 548-549. <https://doi.org/10.1016/j.tics.2021.04.002>

- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49-63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Wickham, H. (2016). Getting Started with ggplot2. In H. Wickham (Éd.), *Ggplot2: Elegant Graphics for Data Analysis* (p. 11-31). Springer International Publishing. [https://doi.org/10.1007/978-3-319-24277-4\\_2](https://doi.org/10.1007/978-3-319-24277-4_2)
- Wurm, M. F., & Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature Communications*, 10(1), 289. <https://doi.org/10.1038/s41467-018-08084-y>
- Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. *The Journal of Neuroscience*, 37(3), 562-575. <https://doi.org/10.1523/JNEUROSCI.1717-16.2016>
- Yan, X., Liu-Shuang, J., & Rossion, B. (2019). Effect of face-related task on rapid individual face discrimination. *Neuropsychologia*, 129, 236-245. <https://doi.org/10.1016/j.neuropsychologia.2019.04.002>
- Yang, D. Y.-J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews*, 51, 263-275. <https://doi.org/10.1016/j.neubiorev.2015.01.020>
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141.
- Yin, J., Xu, H., Duan, J., & Shen, M. (2018). Object-Based Attention on Social Units: Visual Selection of Hands Performing a Social Interaction. *Psychological Science*, 29(7), 1040-1048. <https://doi.org/10.1177/0956797617749636>
- Zemon, V., & Ratliff, F. (1984). Intermodulation components of the visual evoked potential: Responses to lateral and superimposed stimuli. *Biological Cybernetics*, 50(6), 401-408. <https://doi.org/10.1007/BF00335197>
- Zhang, P., Jamison, K., Engel, S., He, B., & He, S. (2011). Binocular Rivalry Requires Visual Attention. *Neuron*, 71(2), 362-369. <https://doi.org/10.1016/j.neuron.2011.05.035>
- Zhou, C., Han, M., Liang, Q., Hu, Y.-F., & Kuai, S.-G. (2019). A social interaction field model accurately identifies static and dynamic social groupings. *Nature Human Behaviour*, 3(8), 847-855. <https://doi.org/10.1038/s41562-019-0618-2>
- Zieber, N., Kangas, A., Hock, A., & Bhatt, R. S. (2015). Body Structure Perception in Infancy. *Infancy*, 20(1), 1-17. <https://doi.org/10.1111/infa.12064>

# Appendices

## 1. Appendix A – Supplementary material of Chapter 1

### 1.1. Alpha suppression analysis

Researchers have hypothesized that spatial relations between non-social objects with clear anterior-posterior organizations can affect the processing of visual scenes, the same way spatial relations between bodies do. Pointing toward each other, facing non-social objects would enhance attentional processing, by generating attentional hot-spots in-between them (Vestner et al., 2020; 2021; 2022). Such attentional process would explain effects observed with facing bodies.

The present analysis aimed at testing that hypothesis, by measuring the level of attention during the frequency-tagging experiment reported in Chapter 1. In EEG, modulations of attention can be evidenced in the Alpha band, as a suppression of power (Klimesch, 1999; Klimesch et al., 1998; Sauseng et al., 2005). Thus, to test whether facingness enhanced visual attention, next analysis measured whether Alpha suppression was stronger with facing body dyads, as well as facing object pairs, compared to non-facing bodies or objects.

This was made possible by the design of Chapter 1, that allowed measuring broad band effects at the same time with Steady-State Visually Evoked Potentials (SSVEP). Short resting state periods before each trial allowed measuring a baseline activity, which could be used to normalize activity in the Alpha band across trials. Power in the alpha band could then be dissociated from narrow band effects such as SSVEPs, scattered over the spectrum (at F1, F2, harmonics and intermodulation frequencies). Though both types of activities may overlap on some frequencies, analyzing them separately can show different effects, unraveling different processes (Aissani et al., 2011). Thus, broadband power suppression in the Alpha band (8 – 13 Hz), could be isolated from narrowband SSVEPs.

#### Analysis

**Procedure.** Each trial lasted 80 seconds (s), starting with a buffer period of 10 s during which participant could retrieve their position. This buffer period was recorded and could be used as a baseline for a time-frequency analysis. Cumulating up to ~2 minutes, this baseline accounted for slow drifts of impedance along the whole testing, and recorded neural activity equally before each condition. As participants had to fixate the screen during testing, recording was eyes open, a better baseline for visual stimulation.

**Preprocessing.** Trials included the baseline and the stimulation period of the SSVEP analysis (total duration of 78.989 s). Data were down sampled at 100 Hz to reduce computation time (total of 7898 samples). Amplitude was normalized by subtracting each electrode its mean amplitude. Trials of the same category and configuration were averaged regardless of frequency side. Because alpha suppression can be topographically widespread (Klimesch, 1999), following analysis was performed averaging all electrodes.

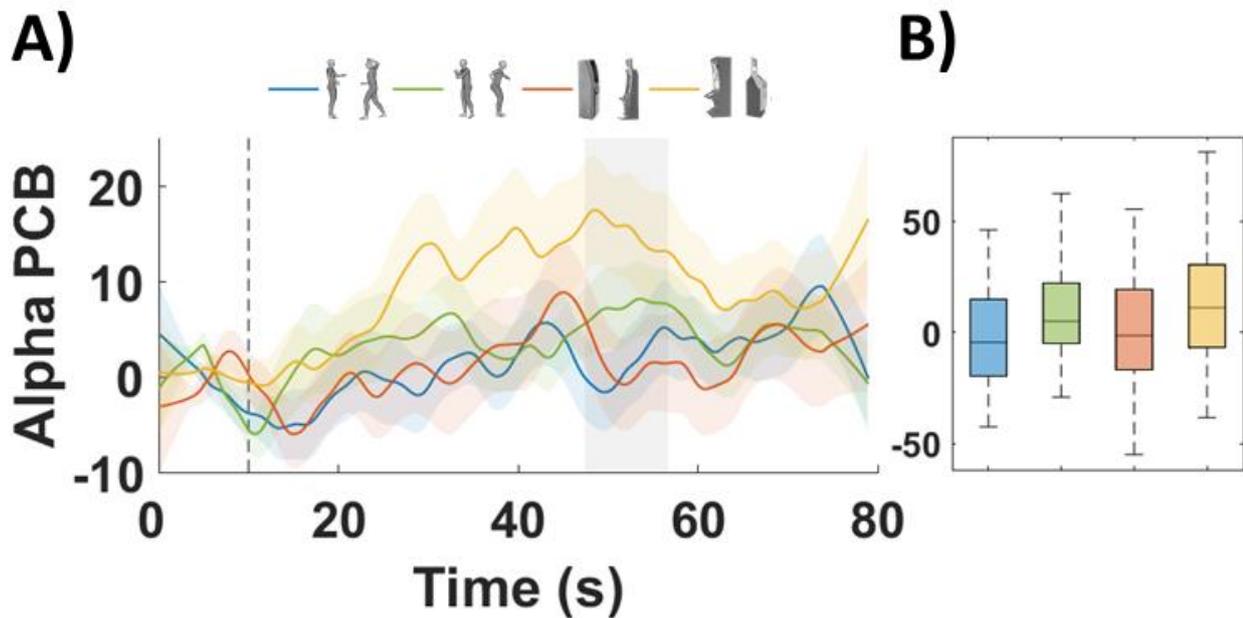
**Alpha suppression.** Trials were much longer compared to experiments investigating event-related Alpha suppression. To account for changes over time, Alpha suppression was analyzed in time-frequency using the Filter-Hilbert method. The following was computed by participant and condition. A linear-phase finite impulse response band-pass filter (8 – 13 Hz) using least-squares error minimization was applied on down-sampled EEG data (Matlab *firls* and *hilbert* functions). Filter kernel was centered on 10 Hz, included at least three cycles of the lowest frequency and had a transition widths of 0.1 Hz. The analytic signal of filtered data was obtained by Hilbert transform. The envelop of the analytic signal provided the amplitude of the Alpha band. Next, amplitude was normalized by computing the percentage change from baseline (PCB). The average amplitude during the baseline was subtracted from the amplitude at each time point, this difference was divided by the baseline average, then multiplied by 100:  $PCB(t) = \frac{activity_t - \overline{baseline}}{\overline{baseline}} * 100$ . The PCB is negative if the amplitude was lower relative to baseline, positive if it was higher. To account for the variability over time, time series of the PCB were smoothed with the *smooth* function of Matlab, over spans of 10 s. Effects of Configuration, Category and the interaction between the two were computed at each time point. The effect of Configuration was computed as the difference between the PCB for face-to-face stimuli and the PCB for back-to-back stimuli. Thus, a negative value indicates stronger Alpha suppression for face-to-face stimuli; a positive value indicates stronger Alpha suppression for back-to-back stimuli. The effect of Category was measured as the difference between the PCB for Bodies and the PCB for Objects. A negative value means stronger Alpha suppression for bodies; a positive value means stronger Alpha suppression for objects. The Category by Configuration interaction was computed as the difference between the effect of Configuration with Objects and the effect of Configuration with Bodies. Significance was assessed by Wilcoxon signed-rank tests against chance (0). The time windows of effects and interaction were determined by testing each time point. To correct for multiple comparisons, clusters of at least two adjacent electrodes showing significant effects ( $\alpha = .05$ ) were submitted to cluster-mass permutation analyses (Maris & Oostenveld, 2007).

## Results

There was a significant difference between the PCB recorded with stimuli face-to-face of both categories, compared to stimuli back-to-back of both categories between 47.53 and 56.72 s ( $p = .0384$ ; Supplementary Figure S1). In this time window, PCB was lower for stimuli face-to-face ( $M = -0.20$ ,  $SD = 22.06$ ) than stimuli back-to-back ( $M = 11.14$ ,  $SD = 21.69$ ). There was no effect of Category and no interaction between the two effects.

## Supplementary Figure S1.

*Time course of Alpha suppression.*



*Note:* A) Time course (x axis) of the Alpha band amplitude's Percent Change from Baseline (PCB, y axis) for each condition. Each curve represents the average PCB of one condition, the corresponding shaded area indicates the standard error from the mean. A dashed vertical line indicates the end of the baseline period and the start of the stimulation, 10 seconds after trial start. The gray rectangle highlights the time window during which the PCB was different with stimuli face-to-face compared to stimuli back-to-back. B) Boxplots represent the distributions of the PCB during that window, averaged by participant, for each condition. Conditions are color-coded: blue = face-to-face bodies, green = back-to-back bodies, orange = face-to-face objects, yellow = back-to-back objects.

### Summary

Alpha rhythms were more suppressed when participants viewed stimuli face-to-face compared to when stimuli were back-to-back, suggesting that participants paid more attention to stimuli face-to-face (Klimesch, 1999; Klimesch et al., 1998; Sauseng et al., 2005). The category of stimulus had no effect.

This pattern is consistent with the hypothesis that objects with a clear directionality generate attentional “hot-spots” (Vestner et al., 2020, 2021, 2022). However, it cannot account for the effects recorded on SSVEPs, as the effect of Configuration on Alpha suppression was insensitive to the category of stimuli and located inside the frequency band where intermodulation was absent.

This suggests the coexistence of two different mechanisms: a perceptual mechanism specifically integrating bodies in the spatial configuration of social interaction, and a general attentional mechanism cued by the direction of objects organized along an anterior-posterior axis.

## 2. Appendix B – Supplementary materials of Chapter 2

### Supplementary Table S1.

*Summary of sample sizes in the literature.*

Study	Sample size <sup>a</sup>	Topic	Age
Baillargeon et al. (1985)	$N = 21-22^b$	Object permanence	4-6 months
Farroni et al. (2005)	$N = 12-17$	Social perception	newborns
Gluckman & Johnson (2013)	$N = 32$	Social perception	6 months
Gwiazda et al. (1997)	$N = \sim 24^c$	Vision	2-8 months
Hamlin et al. (2007)	$N = 12-16$	Social cognition	6-10 months
Hochmann & Papeo (2014)	$N = 14-16$	Speech perception	3-6 months
Jin & Baillargeon (2017)	$N = 16$	Social cognition	17 months
Kelly et al. (2007)	$N = 16$	Social perception	3-6-9-12 months
Kidd et al. (2012)	$N = 30-42$	Attention	7-8 months
Leslie (1982)	$N = 4-7$	Causality	4-8 months
Leslie & Keeble (1987)	$N = 12-17$	Causality	6 months
Lewkowicz & Ghazanfar (2011)	$N = 8$	Vision	6-8-10-12 months
Margoni et al. (2018)	$N = 16$	Social cognition	21 months
Mascaro & Csibra (2012)	$N = 16$	Social cognition	9-12-15 months
Muentener & Carey (2010)	$N = 20-40$	Causality	8 months
Newman et al. (2010)	$N = 12-14$	Social cognition	12 months
Pascalis et al. (2015)	$N = 13-18$	Social perception	6 months
Powell & Spelke (2013)	$N = 16-24$	Social cognition	11 months
Powell & Spelke (2018)	$N = 16-24$	Social cognition	4-5-12 months
Saxe et al. (2005)	$N = 16-20$	Causality	10-12 months
Simion et al. (2008)	$N = 12-18$	Social perception	newborns
Sloane et al. (2012)	$N = 16-18$	Social cognition	19-21 months
Stahl & Feigenson (2014)	$N = 16-32$	Social cognition	16 months
Tatone et al. (2015)	$N = 16$	Social cognition	12 months
Thomsen et al. (2011)	$N = 16$	Social cognition	8-13 months

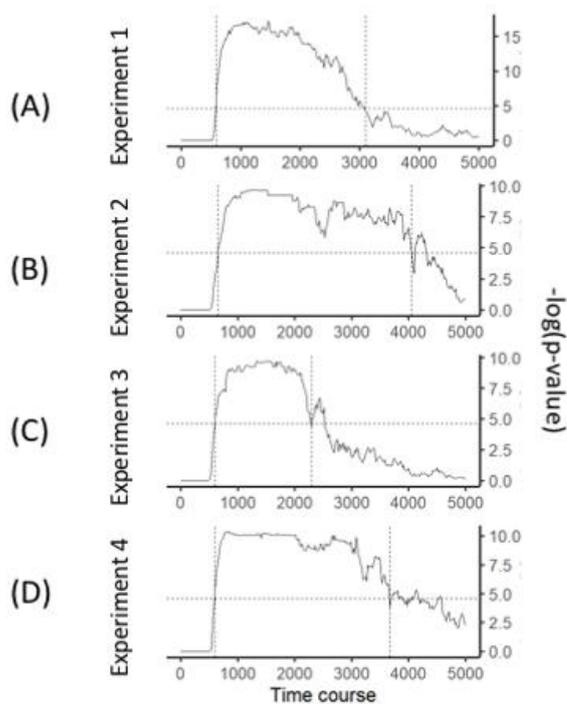
*Note:* <sup>a</sup> maximal and minimal sample sizes of groups in repeated-measures designs, or groups assigned to only one condition in between-subjects designs; <sup>b</sup> range of samples sizes across different experiments of a study; <sup>c</sup> sample sizes are only approximate given the information in the article.

**Supplementary Table S2.***Proportions of first fixations.*

	Contrast	Mean proportions (Standard-Deviation)		Wilcoxon signed-rank test
		Condition 1	Condition 2	
Experiment 1	Facing vs. Non-facing (upright)	.48 (.11)	.52 (.11)	W = 76, p = .46
	Facing vs. Non-facing (inverted)	.45 (.16)	.55 (.16)	W = 69, p = .30
Experiment 2	Facing vs. Single-body	.54 (.19)	.46 (.19)	W = 53, p = .62
	Non-facing vs. Single-body	.65 (.23)	.35 (.23)	W = 113, p = .02
Experiment 3	Facing vs. Non-reciprocal	.49 (.24)	.51 (.24)	W = 61, p = .74
	Non-facing vs. Non-reciprocal	.51 (.16)	.49 (.16)	W = 45, p = 1
Experiment 4	Bodies facing vs. Non-facing plants	.52 (.17)	.48 (.17)	W = 71.5, p = .88

## Supplementary Figure S2.

### *Identification of the Informative Time Window.*



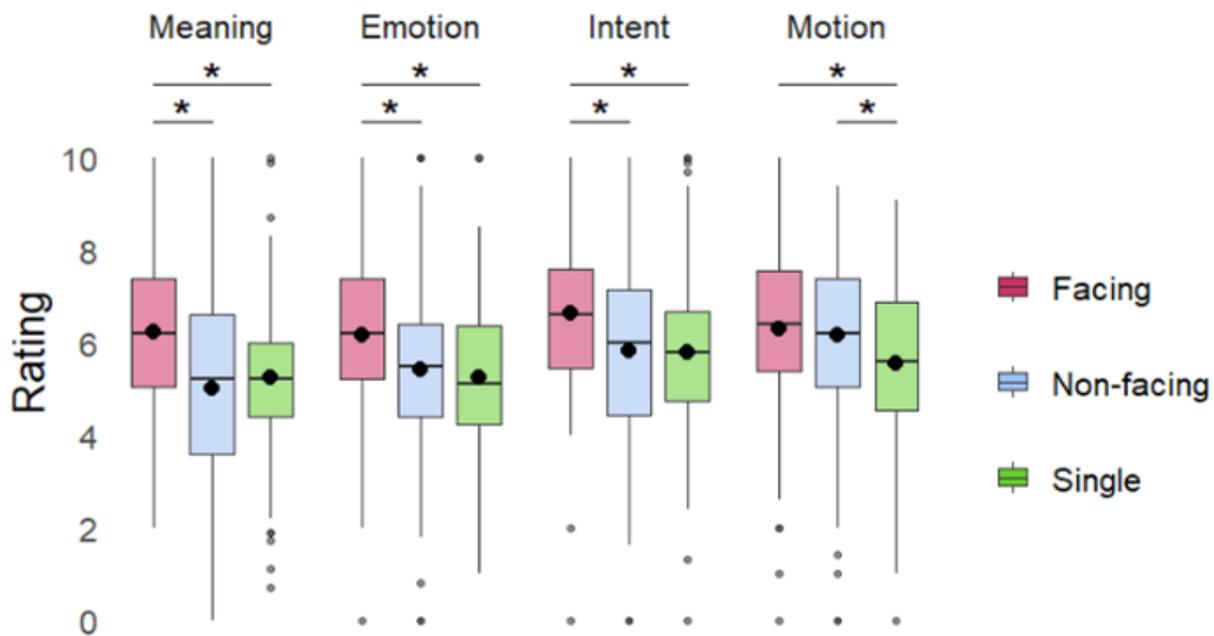
*Note:* Informative Time Window (ITW) for Experiments 1-4. Each curve represents the time course of  $-\log(p\text{-value})$ , where p-value is the result of a Wilcoxon signed rank test comparing the proportion of trials looking at one of the picture to .50. The horizontal dotted line indicates the threshold corresponding to a p-value of .01 ( $-\log(.01) = 4.6052$ ). The vertical black lines mark the boundaries of the ITW for each experiment.

### 3. Appendix C – Supplementary material of Chapter 3

#### 3.1. Supplementary material of Experiment 1b

##### Supplementary Figure S3.

*Ratings of social semantic dimensions.*



*Note:* Results of the rating study evaluating three social semantic dimensions of the stimuli –meaningfulness (meaning), emotional content (emotion) and intentionality (intent)– and one perceptual dimension –implied motion (motion). In box plots, large dots indicate means, thick horizontal bars medians, lower and upper hinges, 1st and 3rd quartiles respectively, whiskers, the span encompassing values largest/smallest than 1.5 times the inter-quartile range, and small dots, values beyond this range. Stars highlight significant pairwise comparisons (\*  $p < .001$ ).

### 3.2. Memory task with human adults

In order to encourage the careful inspection of displays, participants were instructed that they would perform a task after stimulus presentation.

#### Method

*Stimuli:* Sixteen new body postures (eight unique bodies in a lateral view and their mirrored images) were used to create 16 facing dyads and 16 non-facing dyads, following the procedure detailed above. Sixteen dyads of this new set were intermingled with sixteen dyads of the first set (50% dyads facing in each set). Thirty-two displays were generated, showing dyads in the center of the screen.

*Procedure:* The details of the task were given after recording preferential looking times. Participants were instructed that they would see a new series of single images, some of which they had seen before. Half trials showed new images, randomly intermingled with old images. Trials started automatically after central cross fixation. Each trial ( $N=32$ ), participants had to report whether they recognized the image, by pressing one of two keys – “w” or “o” – on a computer keyboard (QWERTY), with the left or right index finger respectively (key mapping to “yes” or “no” was counterbalanced across subjects). Images were displayed for 5 s; response time was unlimited. Accuracy and response times (RT) were recorded.

#### Results

Participants' accuracy was average separately for facing bodies ( $M = 0.81$ ,  $SD = 0.18$ ) and non-facing bodies ( $M = 0.77$ ,  $SD = 0.17$ ). Testing a difference with a two-tailed paired  $t$ -test found no effect  $t(23) = 1.09$ ,  $p = .289$ ,  $d = 0.22$ . Response times were averaged in each condition for accurate response only (facing:  $M = 2555$ ,  $SD = 817$ ; non-facing:  $M = 2570$ ,  $SD = 833$ ). Another two-tailed paired  $t$ -test found no difference either  $t(23) = -0.24$ ,  $p = .815$ ,  $d = 0.05$ .

### 3.3. Macaques rearing

The experimental group consisted in 11 mother-reared (MR: 6 females) and 10 peer-reared (NR: 5 females) monkeys.

#### Method

Subjects were born and raised at the Laboratory of Comparative Ethology at the National Institutes of Health, US. Rearing procedures were approved by the NICHD and the University of Maryland Animal Care and Use Committee, and adhered to the NIH Guide for the Care and Use of Laboratory Animals. Animals were relocated to France at two years of age and housed at the CNRS Primatological Station in Rousset. At this station, animals live together in semi free-ranging conditions, with access to both indoor and outdoor areas. As part of a wider study, subjects were temporarily relocated to the ISC-MJ, Bron each year in small groups. At the ISC-MJ, groups were housed together in an indoor enclosure, enriched to allow for meaningful activities and interactions. Subjects were housed in mixed mother- and peer-reared groups of 5-6 animals. Peer-reared animals were raised in a nursery with access to same-aged peers; see Shannon et al. (1998) for more details. At eight months, mother-reared and peer-reared subjects were placed into a single social group. Small groups temporarily relocated to the ISC-MJ, Bron each year contained both mother-reared and peer-reared animals.

#### Results

About the same number of subjects successfully passed the task in each group (mother-reared:  $N = 9$ , 4 females; peer-reared:  $N = 6$ , 2 females). Groups differed neither in the number of trials included ( $t(13) = -0.57$ ,  $p = .579$ ) (mother-reared:  $M = 6.56$ ,  $SD = 2.79$ ; peer-reared:  $M = 7.33$ ,  $SD = 2.25$ ) nor average looking time on images ( $t(13) = 0.71$ ,  $p = .49$ ) (mother-reared  $M = 1.97$ ,  $SD = 0.87$ ; peer-reared:  $M = 1.70$ ,  $SD = 0.42$ ). Differential looking times between facing and non-facing conspecifics did not change between groups (mother-reared:  $M = 0.21$ ,  $SD = 0.27$ ; peer-reared:  $M = 0.20$ ,  $SD = 0.15$ ;  $t(13) = 0.05$ ,  $p = .960$ ,  $d = 0.03$ ).

- 4. Appendix D – Published version of Visual perception grounding of social cognition in preverbal infants (Goupil, Papeo & Hochmann, 2022)**
- 5. Appendix E – Preprint of Category-specific effects of high-level relations in visual search (Goupil, Kaiser & Papeo, 2023)**

# Visual perception grounding of social cognition in preverbal infants

Nicolas Goupil  | Liuba Papeo  | Jean-Rémy Hochmann 

Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, Centre National de la Recherche Scientifique (CNRS) & Université Claude Bernard Lyon1, Bron, France

## Correspondence

Nicolas Goupil and Liuba Papeo,  
Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, 67 Boulevard Pinel, 69675, Bron France.  
Emails: nicolas.goupil@isc.cnrs.fr;  
liuba.papeo@isc.cnrs.fr

## Funding information

European Research Council, Grant/Award Number: 758473

## Abstract

Social life is inherently relational, entailing the ability to recognize and monitor social entities and the relationships between them. Very young infants privilege socially relevant entities in the visual world, such as faces and bodies. Here, we show that six-month-old infants also discriminate between configurations of multiple human bodies, based on the internal visuo-spatial relations between bodies, which could cue—or not—social interaction. We measured the differential looking times for two images, each featuring two identical bodies, but in different spatial relations. Infants discriminated between face-to-face and back-to-back body dyads (Experiment 1), and treated face-to-face dyads with higher efficiency (i.e., processing speed), relative to back-to-back dyads (Experiment 2). Looking times for dyads in an asymmetrical relation (i.e., one body facing another without reciprocation) were comparable to looking times for face-to-face dyads, and differed from looking times to back-to-back dyads, suggesting general discrimination between the presence versus absence of relation (Experiment 3). Infants' discrimination of images based on relative positioning of items did not generalize to body-object pairs (Experiment 4). Early sensitivity to the relative positioning of bodies in a scene may be a building block of social cognition, preparing the discovery of the keel and backbone of social life: relations.

**KEYWORDS**

body perception, event representation, relations, scene perception, social cognition, spatial cognition

## 1 | INTRODUCTION

Human cognition has been massively shaped by the requirements of social life. Various perceptual adaptations have evolved to serve social life with remarkable efficiency. The social value of objects in the visual world is a guiding principle of human selective attention and perception. Socially relevant entities such as faces and bodies are attended to, detected and recognized with the highest priority, in very young infants (Farroni et al., 2005; Morton & Johnson, 1991; Simion et al., 2008) and throughout the life span (Gluckman & Johnson, 2013; New et al., 2007; Ro et al., 2007; Sun et al., 2017).

Early perceptual tuning for socially relevant entities is a pillar in the development of social cognition (Baron-Cohen et al., 2000), but understanding the social world entails processing social relationships is beyond the social entities. How does the ability to understand social relationships come about? Approaching this question, we sought to identify visual perceptual mechanisms that may lay the foundations for detection and recognition of social interaction in the visual world.

Social relationships such as physical or communicative exchanges involve at least two entities, often close and perceptually accessible to one another, to allow fundamental social processes such as gaze following and shared attention (Baron-Cohen, 1997; Birmingham & Kingstone, 2009; Graziano & Kastner, 2011). While a reciprocal relationship illustrated by face-to-face social agents may constitute the prototypical representation of a social relationship, social relationships can also be non-reciprocal, as long as two social entities play a role in the same structure (e.g., in agent-patient relations, where one acts on another). Under the pressure of the social life to encode social relationships, human vision might have developed efficient coding of perceptual units larger and more complex than single social entities.

Encouraging this hypothesis, research on human adults has shown that two bodies facing toward each other, as if interacting, are detected and recognized more efficiently (i.e., faster and/or more accurately) than the same bodies facing away from each other, when stimuli are presented around perceptual threshold (Papeo et al., 2017), or in visual search through a crowd (Papeo et al., 2019; Vestner et al., 2019). A working hypothesis is that efficient perception of facing dyads is mediated by perceptual grouping, that is, the processing of two bodies as a unitary configuration (as opposed to two independent items), analogous to the processing of facial features in a face, or body parts in a body (Papeo et al., 2019; Vestner et al., 2019; Yin et al., 2018).

Visual sensitivity to facing—seemingly interacting—bodies in adults could be the product of decades of exposure to social events. Alternatively, it could be a signature of early developing perceptual mechanisms that contribute to the construction of cognitive representations toward a mature social cognitive system. The relationship between perceptual mechanisms that manifest early and the development of cognitive abilities is pervasive across domains of cognition (Carey, 2009). For example, perception of cohesion, boundedness and movement through a continuous path supports object representation (Spelke, 1990); sensitivity to spatiotemporal continuity in the motion of two objects (e.g., object A moves toward object B, which starts moving immediately after being contacted by A) triggers representation of cause-effect relationships (Kominsky et al., 2017; Leslie & Keeble, 1987; Leslie, 1982; Michotte, (1946/1963); Muentener & Carey, 2010; Scholl & Tremoulet, 2000) eye-like stimuli (i.e., two dark spots on a brighter background) or

self-propelled motion prompt attribution of agency and intentionality (Gergely & Csibra, 2003; Newman et al., 2010; Saxe et al., 2005).

Along these lines, early discrimination of visuo-spatial relations between social entities could mediate discrimination of basic social events (e.g., conflict or cooperation) and role attribution, promoting rapid enactment of contextually adaptive behaviors and learning about social norms and relationships. In this process, perceptual grouping triggered by certain visuo-spatial relations between human bodies may give an advantage to interacting, over non-interacting agents, which may streamline the processing across body perception and representation of social events.

The ability to represent social events emerges early in life. Within the first year of life, infants discriminate between pro-social and anti-social actions (Hamlin et al., 2007; Jin, & Baillargeon, 2017; Margoni et al., 2018; Sloane et al., 2012), they infer dominance-relationship from physical interaction between two conflicting agents (Mascaro & Csibra, 2012; Thomsen et al., 2011), and an agent's affiliation to a group from her imitative motor behavior (Powell & Spelke, 2013, 2018a, 2018b). The relatively short ontological time course of social inference suggests early sensitivity to basic visual cues of interaction, upon which the representation of visual social events would be built.

In the current study, we first asked whether six-month-old infants could discriminate between multiple-person scenarios based on the mere relative positioning of bodies, which could cue interaction or not (Experiment 1). Discrimination could be revealed by longer looking times to facing dyads, or longer looking times to non-facing dyads. The former effect would be compatible with the preference for socially relevant stimuli (e.g., faces), manifested early on in life (e.g., Farroni et al., 2005; Morton & Johnson, 1991; Simion et al., 2008). In the case of longer looking times toward non-facing dyads, an explanation based on preference would be less likely. The latter effect can rather be predicted based on the abovementioned research on adults, showing that two bodies, whose spatial positioning suggests relationship (e.g., face-to-face), are processed more efficiently (i.e., faster) than two unrelated bodies (Papeo et al., 2017, 2019; Vestner et al., 2020; Yin et al., 2018). Since Experiment 1 showed evidence for the second effect, we carried out a follow-up study to highlight the processing advantage of facing over non-facing dyads. In Experiment 2, we asked whether facing dyads were processed analogously to single units (i.e., single bodies) in terms of looking times. Finally, we addressed whether discrimination based on dyadic visuo-spatial relations extended to non-reciprocal relations involving two bodies (Experiment 3) or body-object pairs (Experiment 4). Across four experiments, eye-tracking was used to measure the relative time that infants looked at each of two images presented simultaneously on the two sides of a screen, and differing by one property only: spatial positioning of two bodies (Experiments 1, 3, 4) or number of bodies (one versus two; Experiment 2). We reasoned that, if the manipulated property is salient to infants, looking times to the two images should differ. With this logic, we examined the ability of preverbal infants to encode one of the most basic perceptual cues of social relationship, the relative positioning of spatially close bodies. As a result of our investigation, we describe a perceptual mechanism that can contribute to get the processing of visual social events off the ground.

## 2 | EXPERIMENT 1—FACING VERSUS NON-FACING

Can infants distinguish two human groups involving identical people but in a different spatial relation? To address this question, 6-months-old infants were presented with two images at a time, showing the same two bodies facing toward and away from each other, respectively (Upright

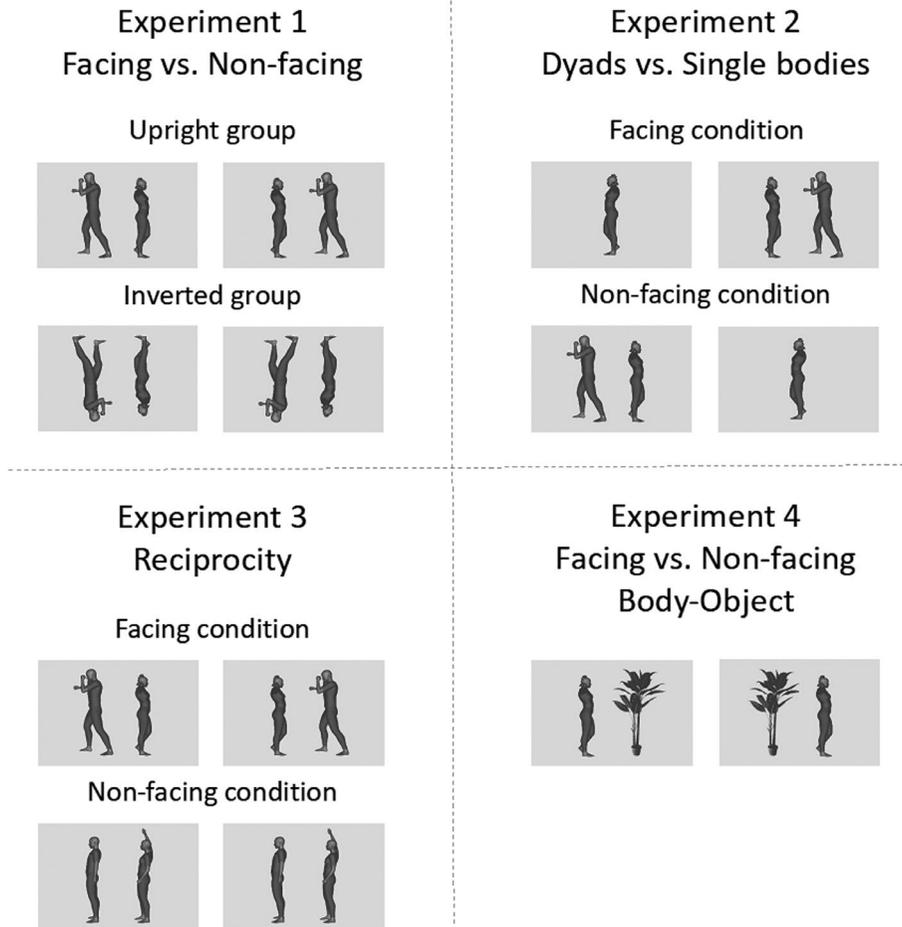


FIGURE 1 Examples of the stimuli used in Experiments 1–4

Group; see Figure 1). Facing and non-facing dyads differed only for the relative spatial positioning of the two bodies. A second group of infants was tested with the same stimuli, but inverted upside-down (Inverted Group; see Figure 1), to control for the effect of possible visual differences, other than body positioning. Inversion disrupts the body structure, while preserving all the low-level visual features of the images (e.g., shapes, luminance, and contrast). By testing inverted body dyads, we could ascertain that any effect observed with upright bodies reflected the processing of bodies, rather than possible low-level visual differences between the two conditions.

## 2.1 | Participants

Two groups of 20 6-months-old infants were tested in Experiment 1 (Upright group: 9 females, 11 males, age range 5 months (m) 21 days (d) – 7 m 16 d; average 6 m 19 d; Inverted group: 6 females, 14 males, age range 5 m 14 d – 7 m 04 d; average 6 m 18 d). Data from one participant in the Inverted group were discarded after analysis, based on the exclusion criteria (see below). Experiment 1 was exploratory with respect to the sample size, although we note that a sample size of 20 corresponds to the average and median sample size of relevant infants'

studies in the field (a meta-analysis of the 25 infants' studies cited in this article shows that sample sizes ranged from 4 to 42, with a mean value of 18.22,  $SD = 7.82$ , and median of 16; see Table S1).

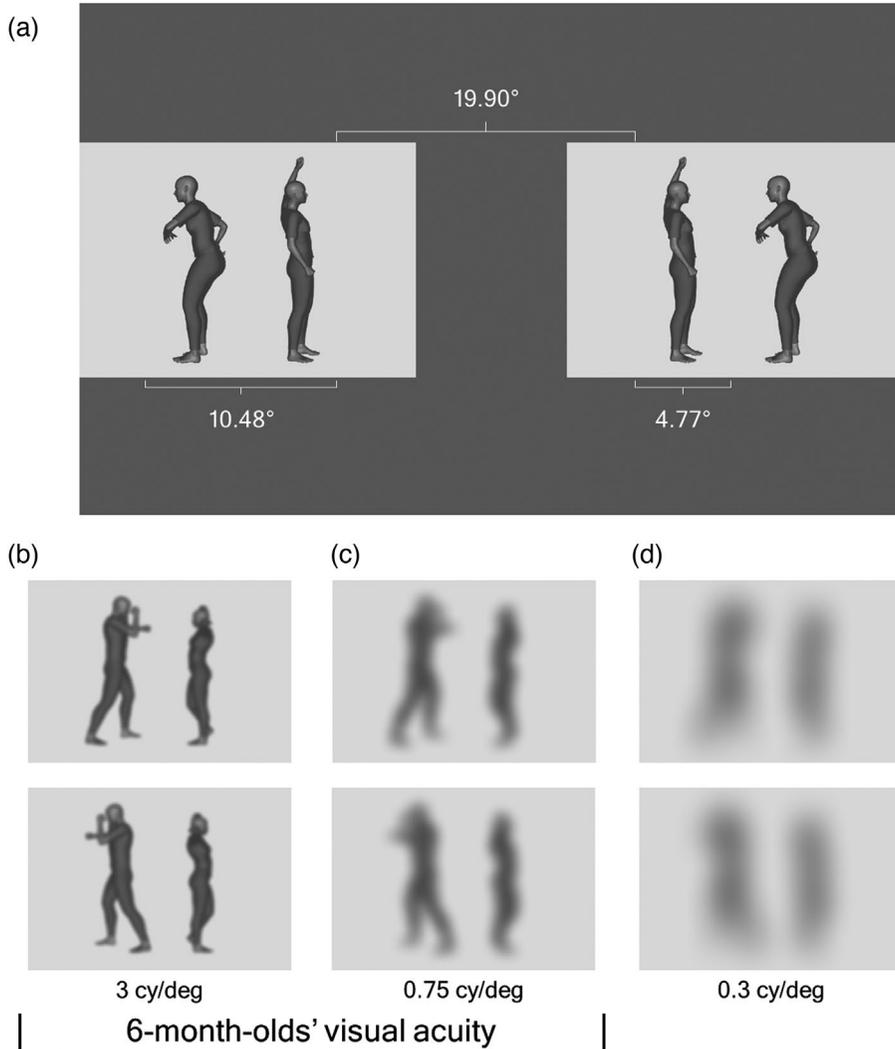
The present and following experiments were conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the local ethics committee (CPP sud-est II), and conducted at the Institut des Sciences Cognitives Marc Jeannerod. Parents were given a 5€ for reimbursement of travel expenses.

## 2.2 | Stimuli

Stimuli were created using gray-scale renders of 16 human bodies (eight unique bodies in a lateral view and their mirrored images), edited with Daz3D (Daz Productions, Salt Lake City, UT) and the Image Processing Toolbox in MATLAB (The MathWorks, Natick, MA). Body poses were all biomechanically possible without fully outstretched limbs in order to have a comparable number of pixels in each half of the figure (the anterior and posterior ends), and to keep bodies at a comparable distance across facing and non-facing dyads. Sixteen unique facing body dyads were created combining the 16 bodies. By randomly combining poses in pairs, dyads did not give rise to any meaningful or familiar interaction. The two bodies in each facing dyad were swapped to create 16 non-facing dyads. Thus, facing and non-facing dyads only differed for the relative positioning of bodies. In Experiment 1, stimuli were 16 displays featuring one facing and the corresponding non-facing dyad, at either side of the screen (the facing dyad was on the left in 50% of displays). In all the displays, dyads were presented inside a rectangular box with a different background color (light gray) relative to the screen background (darker gray), so to create two clearly distinct areas for each dyad on the screen. The two areas were separated by  $\sim 19.9^\circ$ , so that the two bodies of a dyad were much closer to one another than to bodies of the other dyad (Figure 2A). A dyad subtended  $\sim 10.48^\circ$  of visual angle ( $\sim 4.77^\circ$  for a single body). Stimuli of this size are clearly visible at 6 months, when contrast sensitivity peaks between .75 and 3 cy/deg (Gwiazda et al., 1997). Figure 2B–D shows low-pass filtering of an image within a range of .75–3 cy/deg, to illustrate of 6-months-old infants' perception of our stimuli. Dyads were presented in the upright canonical orientation to the Upright group, and rotated by  $180^\circ$ , to the Inverted group.

## 2.3 | Procedure

Infants sat on their parent's lap at a distance of  $\sim 60$  cm from a Tobii T60XL eye-tracker screen, with a sampling rate of 60 Hz, yielding 300 data points per trial. All lights were switched off, except for those coming from the eye-tracker screen. Parents were instructed to close their eyes during the experiment to prevent biasing the child's behavior. Stimulus presentation and recording of eye-tracking data were controlled through PsyScope X (<http://psy.cns.sissa.it/>). In each experiment, infants were presented with 16 trials. Each trial began automatically, when the infant fixated a blinking cross at the center of the screen for at least 200 ms. The cross was then replaced by the stimulus, displayed for 5 s.



**FIGURE 2** Illustration of the stimulation with (A) display layout and (B-D) stimulus visibility at 6 months. (A) Image outlay and dimensions. (B-D) The same exemplar stimulus was filtered by 3 low-pass Gaussian filter. (B) Stimulus low-pass filtered at the highest spatial frequencies of the 6-months-old infants' contrast sensitivity peak range (3 cy/deg;  $\sigma = 7$ ). (C) Stimulus low-pass filtered at the lowest spatial frequencies of the infants' contrast sensitivity peak range (0.75 cy/deg;  $\sigma = 25$ ). (D) Stimulus low-pass filtered at spatial frequencies below the infants' contrast sensitivity peak range (0.3 cy/deg;  $\sigma = 60$ )

## 2.4 | Analyses

### 2.4.1 | Exclusion Criteria

For each participant, missing eye-tracking samples due to technical failure or to the infant looking away, were not interpolated. In Experiments 1–4, for each infant, we computed the cumulative looking time for the two images in each trial. Traditionally, in infant cognition research, individual participants judged as “fussy” are excluded from the final analysis. However, criteria to identify an infant as fussy vary across researchers and studies. Here, we used short looking

TABLE 1 Number of trials included and excluded in the final analyses

Experiment	Total #trial excluded	Total #included	Average # trial excluded per infant (standard deviation)
Experiment 1 Upright	66	254	3.30 (3.05)
Inverted	58	262	2.9 (2.43)
Experiment 2	63	257	3.15 (3.80)
Experiment 3	56	264	2.80 (3.02)
Experiment 4	61	259	3.05 (2.54)

times to identify trials where infants were inattentive, and low overall cumulative looking times to identify infants that were globally inattentive. In doing this, we set criteria to discard the data points for which infants were inattentive, while providing data-driven, objective, and replicable cut-offs, which could keep as much information as possible, in the final dataset.

In particular, trials in which cumulative looking time toward the two images were less than 1 standard deviation (*SD*) from the mean were excluded from further analyses. This resolved into discarding trials with mean duration of one second or less (Experiment 1:  $M = 962$ ,  $SD = 419$ ; Experiment 2:  $M = 1017$ ,  $SD = 476$ ; Experiment 3:  $M = 720$ ,  $SD = 441$ ; Experiment 4:  $M = 1158$  ms,  $SD = 458$ ). In addition, for each experimental group, participants whose cumulative looking time was below 2 *SD* from the group mean were excluded from further analyses (1 in Experiment 1 – inverted condition; 1 in Experiment 2 and 1 in Experiment 3). Table 1 reports the number of trials included for each experiment.

#### 2.4.2 | Time-course analysis and informative time window

The time-course analysis was carried out to detect transient, but statistically reliable effects that might be hidden in the average, when the entire, relatively long, trial duration is taken into account. Here, the trial duration was 5000 ms. However, infants are not expected to attend to the images for 5000 ms. In fact, every trial began with the infant fixating the center of the screen, meaning that in the very first part of the trial the infant did not fixate either stimulus. The first fixation on either stimulus occurred 500 to 1000 ms after the trial onset. Moreover, toward the end of the trial, infants tended to lose interest and look away from the screen. As a result, data at the beginning and at the end of the trial were noisier and less informative, than in the central part of the trial. To account for this, separately for each experiment, we implemented a data-driven approach to identify the *informative time window* (ITW), that is, the most informative interval, during which the majority of participants looked at the screen in most of the trials. For every time point, we computed the proportion of trials, in which each infant looked at either image. We compared these proportions to chance (50%) with a cluster-mass permutation test (Hochmann & Papeo, 2014; Maris & Oostenveld, 2007) using Wilcoxon signed-rank test (one-tail), and defined the ITW as the largest cluster of adjacent time points where the *p*-value of the Wilcoxon test was below .01. While, as we verified for each experiment, there was no significant difference between conditions in the discarded time intervals, excluding those intervals significantly increased the power of our time-course analysis.

Within the defined ITW, for each infant, at each time point, we computed the difference between the number of trials in which the infant looked at one image (e.g., the facing dyad in Experiment 1) and the number of trials in which the infant looked at the other image (e.g., the non-facing dyad in Experiment 1), divided by the total number of trials that the infant contributed.

**TABLE 2** Summary of the average total looking times (and standard deviations) toward the stimuli during the Informative Time Window and during the whole trial (5 seconds), and the mean (and standard deviations) differences in cumulative looking times between conditions, for Experiments 1–4. Time is expressed in ms

Condition	ITW		5 seconds	
	LT <sub>stimuli</sub> (ms)	LT <sub>difference</sub> (ms)	LT <sub>stimuli</sub> (ms)	LT <sub>difference</sub> (ms)
Experiment 1 ( <i>Facing vs. Non-facing</i> )				
Upright	2102 (229)	160 (223)	3320 (496)	−211 (404)
Inverted	1614 (324)	18 (248)	2522 (578)	−28 (386)
Experiment 2 ( <i>Facing or Non-facing vs. Single</i> )				
Facing	2615 (401)	−36 (461)	3322 (456)	−208 (470)
Non-facing	2612 (323)	−264 (380)	3289 (505)	−429 (379)
Experiment 3 ( <i>Facing or Non-facing vs. Non-reciprocal</i> )				
Facing	1415 (165)	4 (312)	2979 (646)	55 (557)
Non-facing	1400 (167)	−154 (249)	2932 (646)	−220 (423)
Experiment 4 ( <i>Facing vs. Non-facing</i> )				
body-plant dyads	2445 (205)	−5 (324)	3323 (384)	−39 (453)

With this method, a positive value indicated higher number of looks toward the facing dyad; a negative difference indicated higher number of looks toward the non-facing dyad. The positive or negative value representing the difference between conditions was compared to chance (0) with a cluster-mass permutation test (relying on a Wilcoxon signed-rank test).

### 2.4.3 | Cumulative looking time analysis

With this analysis, we identified the relative time that infants spent on either type of image, presented in a given trial. In a first analysis, we considered the cumulative looking times for the two stimuli within the ITW, as defined for the above time-course analysis. In a second analysis, we considered the cumulative looking times over the whole trial duration (5000 ms). For each analysis, for each trial, we quantified the infants' ability to discriminate between two body dyads as the difference between the cumulative looking time (LT) to one type of image and the LT to the other type of image divided by the sum of the two values:  $(LT_{\text{facing}} - LT_{\text{non-facing}}) / (LT_{\text{facing}} + LT_{\text{non-facing}})$ . Here, a positive value indicated longer looking time toward the facing dyad; a negative difference indicated longer looking time toward the non-facing dyad. Difference scores were compared to 0. The mean time that infants spent looking at the stimuli and mean difference in looking times between the two stimuli are reported in Table 2.

## 2.5 | Results

### 2.5.1 | Time-course analysis

The ITW of Experiment 1 fell between 578 and 3094 ms (Figure S1A). The cluster-mass permutation test revealed that infants in the Upright group looked longer at the non-facing dyads

between 2040 and 2839 ms ( $p = .009$ ); whereas infants in the Inverted group showed no significant bias toward either stimuli, at any point in time (Figure 3A). Between-group comparison in the time-course analysis showed a trend for a difference between 2176 and 2652 ms, which however, did not reach the significance ( $p = .092$ ), possibly because of limited statistical power to detect small/moderate between-group effects. The between-group difference, however, was statistically reliable in the following cumulative looking time analysis.

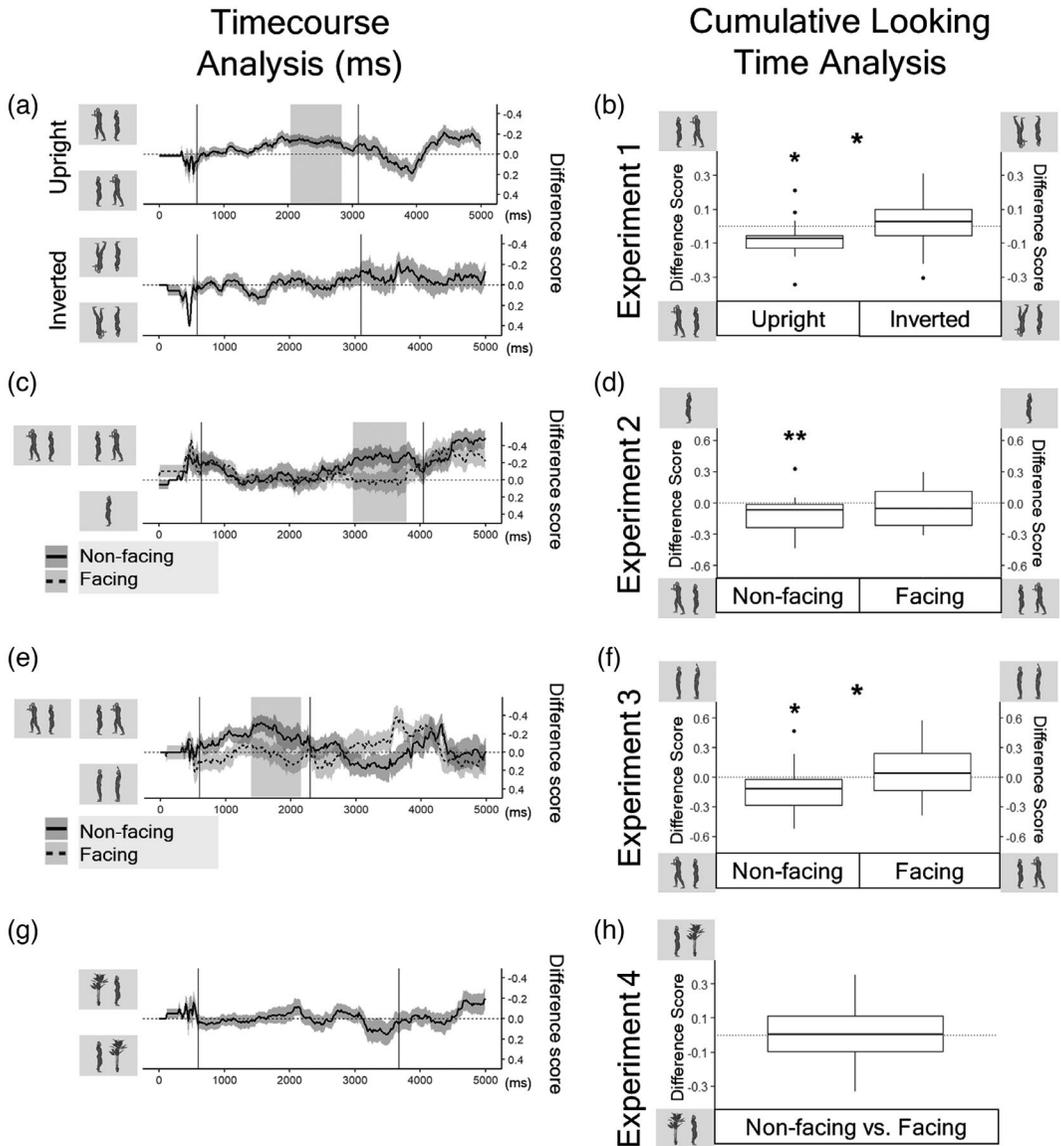
### 2.5.2 | Cumulative looking time analysis

We quantified the infants' ability to discriminate between two body dyads, in terms of normalized difference between the LT toward either type of image. This analysis within the ITW showed an effect of group (Upright vs. Inverted:  $U = 109$ ;  $p = .02$ ;  $r_{\text{Rosenthal}} = .36$ ; Wilcoxon rank-sum test; Figure 3B). In the Upright Group, 17 infants showed a negative difference (longer looking times toward non-facing dyads) and 3 infants showed a positive difference (longer looking times toward facing dyads); in the Inverted Group, 6 infants showed a negative difference and 13 showed a positive difference. The two distributions differed significantly ( $p = .001$ ; Fisher's exact test). Moreover, infants in the Upright Group looked longer at non-facing dyads than at facing dyads ( $M = -.08$ ,  $SD = .11$  corresponding to a difference of  $160 \pm 223$  ms;  $W = 32$ ;  $p = .005$ ;  $r_{\text{Rosenthal}} = .61$ ; Wilcoxon signed-rank test); while no difference between the two types of dyads was found in the Inverted group ( $M = .01$ ,  $SD = .15$  corresponding to a difference of  $18 \pm 248$  ms;  $W = 112$ ;  $p = .52$ ;  $r_{\text{Rosenthal}} = .16$ ; Wilcoxon signed-rank test).

Considering the cumulative looking times over the whole trial duration, pairwise comparisons within each group confirmed the above results. In particular, infants looked longer at non-facing dyads than facing dyads in the upright group ( $M = -.07$ ,  $SD = .12$  corresponding to a difference of  $211 \pm 404$  ms;  $W = 46$ ,  $p = .03$ ;  $r_{\text{Rosenthal}} = .49$ ), but not in the inverted group ( $M = -.01$ ,  $SD = .17$  corresponding to a difference of  $28 \pm 386$  ms;  $W = 94$ ,  $p = .98$ ;  $r_{\text{Rosenthal}} = .01$ ). However, the difference between groups was not significant ( $U = 138$ ,  $p = .15$ ;  $r_{\text{Rosenthal}} = .23$ ), and neither was the difference between the two distributions (Upright Group: 15 infants looked longer toward non-facing dyads and 5 infants looked longer toward facing dyads; Inverted Group: 9 infants looked longer toward non-facing dyads and 10 looked longer toward facing dyads;  $p = .11$ ; Fisher's exact test).

In summary, 6-months-old infants could discriminate between two identical images that differed only for the relative spatial positioning of the two bodies in the scene: facing toward versus away from each other. Inversion canceled the effect, suggesting that the difference between upright body dyads in the Upright Group resulted from processing of bodies and their spatial relations, rather than from other (unforeseen) visual differences between the images, which were all preserved in the inverted images.

These results show that infants encode the spatial relation between bodies in a visual scene. However, while differential looking times imply discrimination between two stimuli (Aslin, 2007), longer looking times to a given stimulus are ambiguous, in that they could index preference for that stimulus, greater complexity relative to the other stimulus (Kidd et al., 2012), or surprise (e.g., Baillargeon et al., 1985). Which one of these cognitive processes accounts for the differential looking times between facing and non-facing dyads? Recent studies on human adults have shown that face-to-face human bodies are processed more efficiently than independent bodies, possibly because they form a structured ensemble, or a group (Papeo, 2020; Papeo et al., 2017, 2019; Vestner et al., 2019; Yin et al., 2018). In Experiment 2, we considered the possibility



**FIGURE 3** Results of the time-course analyses (A,C,E,G) and of the cumulative looking time analyses inside the Informative Time Window (B,D,F,H) for Experiments 1–4. (A,C,E,G) Each curve represents the time course of the difference scores, measuring the tendency to look toward one or the other type of stimuli, by condition. The Horizontal dotted line represents chance (0). Negative values on the Y-axis are plotted upward. Vertical lines mark the boundaries of the informative time window for each experiment. Gray areas highlight the time period of significant differences from chance (Experiments 1 and 4) or significant differences between two conditions (Experiments 2–3), as identified with a cluster-mass permutation test. Light-gray areas indicate standard deviations from the mean. (B,D,F,H) Boxplot of the cumulative looking times analysis within the informative time window for each experiment. The thick horizontal bar represents the median. The lower and upper hinges correspond to the first and third quartiles. Whisker extends from the hinges to the largest/smallest values no further than 1.5 times the inter-quartile range. Single dots correspond to values beyond this range. The dotted horizontal line represents the chance level (0). \*  $p < .05$ ; \*\*  $p < .01$

that results of Experiment 1 implied higher efficiency in processing facing versus non-facing dyads.

### 3 | EXPERIMENT 2 – 1 VERSUS 2

Longer looking times to non-facing dyads could index, against all odds, preference for non-interacting over seemingly interacting scenarios. Here, we addressed the alternative hypothesis that longer looking times to non-facing dyads reflected an effect of number, that is, the difference in the time required to process two independent items (two non-facing bodies) versus one structured unit (a facing dyad).

It has been shown that adults took longer to evaluate images with two bodies back-to-back, than with two bodies face-to-face; while the time to evaluate facing dyads was rather comparable to the time spent on single bodies (Papeo et al., 2017). In a similar vein, research on infants has shown that after witnessing a social interaction between two dolls, 16-months-old represent the two as one chunk in working memory (Stahl & Feigenson, 2014). Those results encourage the hypothesis that infants, just like adults, can process bodies that appear related to one another, more efficiently than independent bodies, and possibly analogously to one single unit.

To test this, in Experiment 2, we presented 6-months-old infants with displays featuring a dyad and a single body (Figure 1). On each display, the dyad could feature either facing bodies (Facing condition) or non-facing bodies (Non-facing condition). If in Experiment 1 longer looking times to non-facing (vs. facing) dyads reflected an effect of number, looking times should differ more between single bodies and dyads when dyads involved non-facing (vs. facing) bodies.

#### 3.1 | Participants

Experiment 2 involved a new group of 20 infants (14 females, 6 males, age range 5 m 14 d – 6 m 27 d; average 6 m 04 d). In Experiment 2 (and in the following experiments), we tested 20 infants like in each group of Experiment 1, which satisfied the minimal sample size ( $N = 18$ ) to obtain an effect of positioning comparable to that of Experiment 1 (i.e., longer looking times for non-facing versus facing dyads;  $t(19) = -3.18$ ;  $p = .005$ ;  $d_{\text{Cohen}} = -.71$ ,  $\beta = .80$ ,  $\alpha = .05$ ; GPower 3.1 package). Data from one participant were discarded based on the exclusion criteria (see Analyses of Experiment 1), yielding a final sample size of 19.

#### 3.2 | Stimuli and procedure

The 16 displays used as stimuli in Experiment 1 were modified for Experiment 2, by replacing one of the two dyads (either the facing or the non-facing one) with a single body. Thus, stimuli of Experiment 2 were 16 displays, each featuring a facing or a non-facing dyad on one side, and a single body on the other side (the dyad was on the left in 50% of displays). Stimulus presentation and every other aspect of the experimental procedure were identical to Experiment 1.

### 3.3 | Data analyses

Data were analyzed as in Experiment 1. In the difference scores, computed separately for the two conditions  $(LT_{\text{single}} - LT_{\text{dyad}})/(LT_{\text{single}} + LT_{\text{dyad}})$ , positive values indicated longer looking times toward the single body; negative values indicated longer looking times toward the dyad (facing or non-facing). The comparison between conditions was one-tailed, given that the direction of the effect (i.e., longer looking times to dyads in both conditions, and larger differences in the non-facing condition) was predicted based on the results of Experiment 1.

### 3.4 | Results

#### 3.4.1 | Time-course analysis

The ITW fell between 646 and 4046 ms (Figure S1B). The time-course analysis showed that the two conditions differed between 2907 and 3791 ms ( $p = .03$ ; one-tail, cluster-mass permutation test). Within this interval, infants looked toward the dyad more often in the non-facing condition than in the facing condition (Figure 3C). In the non-facing condition, the difference in looking time toward single bodies versus dyads was significant between 2958 and 3621 ms ( $p = .03$ ; one-tail, cluster-mass permutation test). No significant difference was found at any point in time between single bodies and dyads, in the facing condition.

#### 3.4.2 | Cumulative looking time analysis

Consistent with the above findings, within the ITW, infants spent significantly longer in fixating the dyad than the single body, in the non-facing condition ( $M = -.12$ ,  $SD = .19$  corresponding to a difference of  $264 \pm 380$  ms;  $W = 25$ ;  $p = .003$ ;  $r_{\text{Rosenthal}} = .65$ ), but not in the facing condition ( $M = -.05$ ,  $SD = .19$  corresponding to a difference of  $36 \pm 461$  ms;  $W = 64$ ;  $p = .23$ ;  $r_{\text{Rosenthal}} = .29$ ; Figure 3D), although the difference between the two conditions (facing vs. non-facing) did not reach significance ( $W = 129$ ;  $p = .09$ ;  $r_{\text{Rosenthal}} = .31$ ; one-tail Wilcoxon signed-rank test). The analysis over the whole trial duration showed that infants looked longer at dyads than single bodies, in both the non-facing condition ( $M = -.16$ ,  $SD = .16$  corresponding to a difference of  $429 \pm 379$  ms;  $W = 16$ ,  $p = .001$ ;  $r_{\text{Rosenthal}} = .73$ ) and the facing condition ( $M = -.10$ ,  $SD = .20$  corresponding to a difference of  $208 \pm 470$  ms;  $W = 45$ ,  $p = .04$ ;  $r_{\text{Rosenthal}} = .46$ ). Again, the two conditions did not differ significantly ( $W = 129$ ,  $p = .09$ ;  $r_{\text{Rosenthal}} = .31$ ; one-tailed).

Experiment 2 confirmed and extended the results of Experiment 1. Infants processed differently two images that were identical except for the relative positioning of bodies. The difference between facing and non-facing dyads here was highlighted in relation to a third stimulus, the single body: although dyads were overall more attractive than single bodies (especially at the beginning of the trial), this effect was consistently found across the three analyses, only when dyads depicted non-facing bodies.

As for the difference between the two conditions, results of the cumulative looking time analysis are not conclusive, though the stronger effect is observed when single bodies were compared with non-facing dyads. More conclusive are the results of the time-course analysis, which captured a significant difference between the facing and the non-facing conditions between 2907 and 3791 ms. Interestingly, despite the fact that facing dyads contain twice more visual information

than single bodies (twice more bodies, twice more pixels), no significant difference between facing dyads and single bodies was found in any time interval.

In summary, Experiment 2 revealed a consistent effect of number (longer looking times to dyads than single bodies) when single bodies were compared with non-facing dyads, and only an inconsistent effect when single bodies were compared with facing dyads. These results lend credit to the hypothesis that, like adults, infants as young as 6 months may process facing –seemingly interacting– bodies more similarly to a single percept than to two independent items (i.e., two unrelated bodies).

Results of Experiments 1–2 might capture the infants' discrimination between a dyadic relation with two mutually accessible (face-to-face) bodies versus the absence of such relation. Alternatively, they might imply a more general discrimination between any relation versus no relation. Experiment 3 addressed the categorical distinction underlying the above discrimination between facing and non-facing dyads in infants.

## 4 | EXPERIMENT 3

### 4.1 | Reciprocity

In Experiment 3, we introduced a new set of dyads, in which one body faced another, who faced away, in a standing pose (Figure 1). We labeled this type of dyad “non-reciprocal,” as the relative positioning and the passive pose of the facing-away body imply a unidirectional relation with one body addressing the other and no sign of reciprocation. A group of 6-months-old infants saw two types of displays: displays with a facing dyad and a non-reciprocal dyad (Facing condition) and displays with a non-facing (back-to-back) dyad and a non-reciprocal dyad (Non-facing condition).

We reasoned that, if infants represent the broad distinction between presence of relation (reciprocal or not) versus no relation, they should discriminate between non-facing and non-reciprocal dyads, but not between facing dyads and non-reciprocal dyads, as the last two would both implement a relation. In contrast, if infants can also represent the narrower distinction between reciprocal and non-reciprocal relations, they should discriminate between facing and non-reciprocal dyads.

### 4.2 | Participants

Experiment 3 involved a new group of 20 infants (9 females, 11 males, age range 6 m 0 d – 7 m 08 d; average 6 m 26 d). One additional infant was tested, but data were not analyzed because of an abrupt interruption (by the infant's sister) of the experiment. Data from one participant were discarded based on the exclusion criteria (see Analyses of Experiment 1), yielding a final sample size of 19.

### 4.3 | Stimuli, procedure, and data analyses

The same displays of Experiment 2 were used in Experiment 3, except that the single body was replaced by non-reciprocal dyads, in which one body faced the other who faced away

(the non-reciprocal dyad was on the left in 50% of displays). The 16 non-reciprocal dyads were created by pairing each of the 16 bodies with the same body facing away, in a standing pose. The choice of a passive pose was made to emphasize the representation of one body acting on another, with no reciprocation. Procedures and data analyses were identical to Experiment 1. Difference scores were computed as  $(LT_{\text{non-reciprocal}} - LT_{\text{facing/non-facing}}) / (LT_{\text{non-reciprocal}} + LT_{\text{facing/non-facing}})$ . Thus, positive difference values indicated longer looking time to non-reciprocal dyads; negative values indicated longer looking time to facing or non-facing dyads.

## 4.4 | Results

### 4.4.1 | Time-course analysis

Within the ITW, identified for this experiment between 595 and 2295 ms (Figure S1C), a difference between the two conditions (facing vs. non-reciprocal and non-facing vs. non-reciprocal) emerged between 1394 and 2159 ms ( $p = .01$ , cluster-mass permutation test)<sup>1</sup>. Comparisons within each condition showed that infants looked longer at non-facing dyads relative to non-reciprocal dyads between 952 and 1887 ms ( $p = .01$ , cluster-mass permutation tests); while at no point in time they showed a bias to either facing or non-reciprocal dyads (Figure 3E).

### 4.4.2 | Cumulative looking time analysis

The analysis of cumulative looking times over the whole trial duration showed no significant difference between the two conditions ( $W = 130$ ,  $p = .17$ ;  $r_{\text{Rosenthal}} = .32$ ). Within each condition, there was no difference between non-reciprocal and facing dyads ( $M = -.006$ ,  $SD = .27$  corresponding to a difference of  $55 \pm 557$  ms;  $W = 95$ ,  $p = 1$ ;  $r_{\text{Rosenthal}} = 0$ ), and only a trend for the difference between non-reciprocal and non-facing dyads ( $M = -.10$ ,  $SD = .22$  corresponding to a difference of  $220 \pm 423$  ms;  $W = 49$ ,  $p = .07$ ;  $r_{\text{Rosenthal}} = .43$ ).

The analysis of cumulative looking times constrained by the ITW proved more sensitive to the effects of conditions. Here we found a significant difference between the two conditions ( $W = 162$ ;  $p = .005$ ;  $r_{\text{Rosenthal}} = .62$ ; Figure 3F), whereby infants looked longer at non-facing than non-reciprocal dyads ( $M = -.18$ ,  $SD = .29$ ; corresponding to a difference of  $154 \pm 249$  ms;  $W = 32$ ,  $p = .009$ ;  $r_{\text{Rosenthal}} = .58$ ), but they spent a comparable amount of time on facing and non-reciprocal dyads ( $M = .05$ ,  $SD = .27$  corresponding to a difference of  $4 \pm 312$  ms;  $W = 109$ ;  $p = .59$ ;  $r_{\text{Rosenthal}} = .13$ ).

Overall, the results of Experiment 3 highlighted stronger discrimination between non-reciprocal and non-facing dyads than between non-reciprocal and facing dyads. We found no evidence that infants could discriminate facing and non-reciprocal dyads. Our results show that infants could distinguish non-facing (back-to-back) dyads possibly cueing absence of relationship, from facing and non-reciprocal dyads, possibly cueing presence of relationship, but they

<sup>1</sup>Visual inspection of the time course shows an apparent difference between conditions, in a late time interval outside the ITW. To test the significance of this effect, we carried out a cluster-mass permutation test over the whole trial duration, which confirmed the effect found within the ITW ( $p = .035$ ), and showed no significant difference at the later time interval ( $Ps > .77$ ).

could not distinguish between two different types of relationships: facing/reciprocal and non-reciprocal. Thus, reciprocity, or the mutual perceptual accessibility of the two bodies held in facing dyads, is not the key feature for infants to discriminate between different spatial relations. Results of Experiment 3 rather suggest that infants may hold a general representation with a body facing toward another, whether the other faces back or not. Such general representation could virtually apply to social scenes (i.e., scenes involving multiple people such as the current dyads), as well as to non-social contexts, for instance, involving human-object relationships. Experiment 4 addressed this possibility.

## 5 | EXPERIMENT 4—BODY-OBJECT DYADS

Experiment 3 suggests that infants can discriminate between scenes in which a body faces toward another (who does or does not reciprocate) versus scenes in which two bodies face away. Can infants perform such discrimination only in the context of social scenes? Experiment 4 addressed whether infants can discriminate between scenes in which a body faces toward another versus scenes in which a body faces away from another, when the other is a non-body object. A new group of 6-months-old infants saw displays featuring two images with a body facing toward and away from a non-body object (i.e., a plant), respectively. On each display, the two images differed only for the relative positioning of the two items.

### 5.1 | Participants

Experiment 4 involved a new group of 20 infants (6 females, 14 males, age range 5 m 23 d – 7 m 00 d; average 6 m 20 d). One additional infant was tested in Experiment 4, but data were discarded because a technical failure stopped the experiment. All the remaining participants in Experiment 4 met the inclusion criterion, yielding a final sample size of 20.

### 5.2 | Stimuli, procedure and analyses

In Experiment 4, 16 new facing pairs were created by coupling each of the 16 single bodies with one of four plants created with the same procedures and software packages as the bodies. By swapping the body and the plant in each facing pair, 16 non-facing pairs were created. Sixteen displays were created with one facing and one non-facing pair showing the same body and plant (the facing pair was on the left in 50% of displays). Procedures and analyses were identical to Experiment 1.

### 5.3 | Results

#### 5.3.1 | Time-course analysis

The ITW covered an interval between 595 and 3672 ms (Figure S1D). The time-course analysis did not reveal any bias toward facing or non-facing body-plant pairs, at any point in time (Figure 3G).

### 5.3.2 | Cumulative looking time analysis

Difference scores did not differ from chance, when considering the cumulative looking times within the ITW ( $M = .004$ ,  $SD = 0.15$  corresponding to a difference of  $5 \pm 324$  ms;  $W = 111$ ;  $p = .84$ ;  $r_{\text{Rosenthal}} = .05$ ; Figure 3H), or over the whole trial duration ( $M = -0.01$ ,  $SD = 0.16$  corresponding to a difference of  $39 \pm 453$  ms;  $W = 100$ ,  $p = .87$ ,  $r_{\text{Rosenthal}} = .04$ ).

The cumulative looking time analyses and the time-course analysis failed to identify any significant difference between the infants' looking behavior toward facing and non-facing body-plant dyads. Experiments 1–3 had shown that infants discriminate between scenes in which a body faces toward a second body from scenes in which two bodies face away from each other. Since the directionality of the second body did not seem to be important for the above discrimination (i.e., the second body could reciprocate, facing back to the first, or not), in Experiment 4, we asked whether the identity of the second object (i.e., the fact that it was a body) mattered, or whether we could obtain a similar effect (discrimination between facing toward vs. facing away) with non-social scenes, where a body faces toward or away from a non-body object. The null effect of Experiment 4 suggests that infants are especially sensitive to visuo-spatial relations between multiple items, in the context of social (multiple-body) scenarios. Thus, rather than a general action scheme (*someone acting over anything*), the effects reported here imply representation of a social scheme with *someone acting with and/or over someone else*.

## 6 | DISCUSSION

Across four experiments, we have shown that the visuo-spatial relation between two bodies is a distinctive feature of a scene, to which humans are sensitive early on in life. The relative positioning of two bodies affected the time that 6-months-old spent looking at an image (Table 3). If the looking time is a measure of processing time, then infants processed face-to-face bodies faster than the same two bodies appearing as unrelated (i.e., back-to-back; Experiment 1). Moreover, the difference in looking times was more pronounced between non-facing dyads and single bodies than between facing dyads and single bodies, with the longest looking times toward non-facing dyads (Experiment 2). Such behavior encourages the thinking that, as for adults, also for infants, spatial relations that cue relationship favor the parsing of complex (multi-body) scenarios, producing a perceptual advantage reflected in more efficient (e.g., faster) processing of the stimuli. Infants also discriminated between two bodies back-to-back and a dyad with a body facing toward another, who faced away (non-reciprocal dyads; Experiment 3). Discrimination between non-facing and non-reciprocal dyads was stronger than discrimination between facing and non-reciprocal dyads. Finally, the effect of positioning (facing toward vs. away from) was found for visual scenes that involved dyads of bodies, but not human-object pairs (Experiment 4). While acknowledging the relatively small sample size of each individual experiment ( $N=20$ ) as a limitation of the study, we point out that Experiments 1–3 provide three conceptual replications of the same effect, which emerged most clearly from the time-course analysis (see Table 3). In particular, across Experiments 1–3, the time-course analysis consistently showed longer looking times toward non-facing (back-to-back) dyads: infants looked longer at non-facing than at facing (upright) dyads in Experiment 1; they looked longer at non-facing dyads than at single bodies in Experiment 2 (where the difference was significantly smaller between facing dyads and single bodies); and they looked longer at non-facing dyads than at non-reciprocal dyads in Experiment 3 (where the difference was smaller between facing dyads and non-reciprocal dyads). While most

TABLE 3 Summary of results. Difference significance and time intervals are reported for Experiments 1–4

Condition	ITW	Time course	Cumulative LTs in ITW
Experiment 1 ( <i>Facing vs. Non-facing</i> )			
Upright	578–3094	Non-facing > Facing (2040–2839, $p = .009$ )	Non-facing > Facing ( $p = .005$ ) <sup>a</sup>
Inverted		n.s.	n.s.
Upright versus Inverted		Upright > Inverted (2176–2652, $p = .09$ )	Upright > Inverted ( $p = .02$ ) <sup>b</sup>
Experiment 2 ( <i>Facing or Non-facing vs. Single</i> )			
Facing	646–4046	n.s.	n.s.
Non-facing		Non-facing > Single (2958–3621, $p = .03$ )	Non-facing > Single ( $p = .003$ ) <sup>a</sup>
Facing versus Non-facing		Non-facing > Facing (2907–3791, $p = .03$ )	Non-facing > Facing ( $p = .09$ ) <sup>a</sup>
Experiment 3 ( <i>Facing or Non-facing vs. Non-reciprocal</i> )			
Facing	595–2295	n.s.	n.s.
Non-facing		Non-facing > Non-reciprocal (952–1887, $p = .01$ )	Non-facing > Non- reciprocal ( $p = .009$ ) <sup>a</sup>
Facing vs. Non-facing		Non-facing > Facing (1394–2159, $p = .01$ )	Non-facing > Facing ( $p = .005$ ) <sup>a</sup>
Experiment 4 ( <i>Facing vs. Non-facing</i> )			
body-plant dyads	595–3672	n.s.	n.s.

<sup>a</sup>Within group comparisons were assessed through Wilcoxon signed-rank tests.

<sup>b</sup>Between-group comparison was assessed through a Wilcoxon rank-sum test.  $n.s. = p > .10$ .

effects were corroborated by the analysis of difference scores based on the cumulative looking times, the time-course analysis proved particularly useful to characterize the infants' behavior over time, detecting transient effects that might go lost when considering the average behavior over a quite long time period (i.e., the whole trial duration).

Individual body postures involved in the dyads (as opposed to the whole dyad) cannot explain the current effects of body positioning. First, the very same body postures were used in facing and non-facing dyads. Thus, the specific body postures cannot account for the difference between facing and non-facing dyads, when contrasted against each other directly (Experiments 1 and 4) or against another (common) stimulus type (single bodies in Experiment 2 and non-reciprocal dyads in Experiment 3). Moreover, in Experiment 1, the inversion of bodies abolished the difference between facing and non-facing dyads, thus ruling out the role of low-level features that could differ between conditions. Another feature of our stimuli, which could in principle play a role in the current effects of body positioning, is the face/head/body/gaze direction. Gaze/head/body direction is known to be a powerful cue to orient attention (Driver et al., 1999; Friesen & Kingstone, 1998; Frischen et al., 2007; Langton et al., 2000; Vestner et al., 2020), to which six-months-old may already be sensitive (Senju, Csibra & Johnson, 2008; Reid et al., 2004). However, the effects that could be predicted based on body direction are opposite to what we observed. The two bodies in a facing dyad would direct the attention toward each other, thus keeping attention

on the site (i.e., in between them). Such mechanism predicts that infants would look longer at the facing than at the non-facing dyad; but this is not what we observed. The direction of bodies in non-facing dyads would also bias attention toward the facing dyad, since the two bodies looked away from one another and one of them (the one closer to the central fixation) looked toward the facing dyads. But, again, an attentional bias toward facing dyads is not what we observed. Related to body direction, recent research suggests that, in a dyad, two heads/bodies oriented toward each other can create an attentional *hot spot* that captures the observer's attention more rapidly than dyads in other (non-facing) spatial relations (Vestner et al., 2020). As a result, facing dyads would be looked at, and processed faster, than non-facing dyads. To assess whether this mechanism could explain the current pattern of results, we considered infants' first fixation in each trial, which reflects attentional capture more reliably than looking times. This analysis, reported as supplementary material (Table S2), showed no reliable effect, ruling out the role of an unbalanced capture of attention in the infants' looking behavior reported here.

In the current study, back-to-back body dyads elicited the longest visual inspection, relative to all other stimuli considered here. One common interpretation of looking times is preference: between two stimuli, infants would look longer toward the preferred one. On this rationale, a preference for socially relevant stimuli (e.g., faces) has been repeatedly shown in infants (Farroni et al., 2005; Simion et al., 2008). Such preference, however, was found when a social stimulus was compared with a non-social stimulus (e.g., a face vs. a scrambled face). In our study, both conditions (facing and non-facing dyads) had social value (they both involved human bodies) and, as such, they were generally attractive for infants. In this context, attention is likely to be attracted by the weird—that is, less familiar or less-expected—social stimulus, more than by the normal—that is, familiar or expected—social stimulus (Lewkowicz & Ghazanfar, 2012). Thus, one possible explanation is that longer looking times toward non-facing dyads indicate that those stimuli violated infants' expectations about the positioning of two nearby bodies. Infants may have an expectation for sociality such that, when two bodies come spatially close, they would engage in some sort of (reciprocal or non-reciprocal) exchange or relationship. Thus, while the current results directly demonstrate the infants' sensitivity to different spatial configurations of multiple bodies, they open to the possibility that six-months-olds are indeed sensitive to the different representational contents implied by facing and non-facing bodies, showing surprise when two spatially close social agents are non-facing as if ignoring each other, rather than interacting.

On another explanation, if looking times index processing times (see Aslin, 2007; Kidd et al., 2012), non-facing dyads might be more complex than the other stimuli because they consist of two perceptual units that are processed independently from one another. This explanation takes into account a growing corpus of studies on adults, showing that multiple bodies, whose spatial positioning suggests relationship, are processed more efficiently (i.e., faster and/or more accurately) than unrelated bodies (Papeo et al., 2017; Papeo et al., 2019; Vestner et al., 2019; see also Kaiser et al., 2019). Behavioral effects (and related neural effects; see Abassi & Papeo, 2020; Bellot et al., 2021; Walbrin & Koldewyn, 2019; Walbrin et al., 2020; Wurm & Caramazza, 2019) concur to demonstrate efficient coding of human groups in visual perception, which might reflect grouping, the representation of objects as a unitary whole, as opposed to multiple individual items. Grouping would increase efficiency by reducing the representational complexity of the stimuli and the competition for neural representation and processing resources (Kaiser et al., 2019; see also Desimone & Duncan, 1995).

We shall note that the two explanations—the one based on expectation of *facingness* and/or interaction for two nearby bodies and the one based on grouping of seemingly related bodies—are not mutually exclusive and may actually be two sides of the same coin: a tradition of research

on perceptual efficiency has indeed shown that expected stimuli are often the easiest to process (Bukach et al., 2006).

The current results on infants add to the extant literature on adults, suggesting that the visual discrimination of multiple-body configurations does not require extensive experience in the social world; it rather appears early on in life, implying a perceptual adaptation that precedes the expression of social inference in infancy.

Early perceptual adaptations act as powerful learning mechanisms, biasing attention and expectations about the environment (Carey, 2009). In the social domain, high visual sensitivity to certain spatial configurations accounts for the infants' attentional bias toward faces, a mechanism that mediates learning about people in the infants' environment and discrimination across human groups (e.g., other-species and other-race effects; Kelly et al., 2007; Pascalis et al., 2005). Early acoustic sensitivity to speech sounds participates in infants' discovery of the rhythm (Nazzi et al., 1998) and phonology (Werker & Tees, 1984) of their native language. In this spirit, early discrimination of spatial relations between bodies might contribute to encoding of social events, and learning about aspects of social interaction such as social norms and roles (e.g., who is the dominant party, when and how to show subordination/dominance, to reciprocate and so on), which partly but reliably correlate with visuo-spatial relations.

Facing dyads embody the prototypical social exchange, perhaps because *facingness* favors a number of basic social processes such as shared attention and gaze following. While adults' vision seems to be especially attuned to face-to-face bodies (Papeo & Abassi, 2019), infants do not seem to distinguish between face-to-face (e.g., reciprocal) and non-reciprocal dyads. Other research has shown that non-reciprocal events such as *giving*, with an agent and a recipient, are interpreted as inherently social by 12-months-old infants (Tatone et al., 2015). The representation of social (i.e., human-human) relationship in infancy can start off with a single broad category that embeds both reciprocal and non-reciprocal schemas. Further research shall address the possibility of a perceptual development yielding the visual specialization for the relation *facingness* seen in adults. Moreover, Experiment 4 with body-plant pairs addressed whether having a body facing anything (another body or an object) was sufficient to observe the effects reported in Experiments 1–3. While results showed that infants are especially sensitive to visuo-spatial relations in social (two-body) dyads, and to a lesser extent, to relations in non-social body-object pairs, further research shall address infants' sensitivity to non-social (body-object or even object-object) pairs, when non-body objects have a more unambiguous antero-posterior organization (i.e., a front and a back), and/or evoke stronger representations of interaction, than plants. Building on current findings, it shall also be studied what infants represent when they see a facing dyad, and through which transformations a dyadic schema with mutually accessible body-shapes becomes the representation of social interaction. Here, we manipulated spatial relations under the hypothesis that they can cue social relationship. The specific representational content evoked by two facing or two non-facing bodies, in infants, but also in adults, remains a fascinating question for further research.

Past research has exposed perceptual adaptations in young infants, which define expectations about the world, guide perception and shape cognition. Here, we report evidence for early developing discrimination of perceptual proxies of dyadic social interactions, based on internal spatial relations between participants in a visual scene (i.e., individual bodies). Infants' discrimination resolves in the broad distinction between implied presence and absence of relationship between two people. The mechanism revealed here may contribute to explain how infants' perception prepares the discovery of the keel and backbone of social cognition: relations.

## ACKNOWLEDGMENTS

This work was funded by a Fyssen Foundation grant (2014) awarded to JRH and a European Research Council Starting Grant awarded to L.P. (Project: THEMPO, Grant Agreement 758473). The authors declare no conflicts of interest with regard to the funding source for this study.

## ORCID

Nicolas Goupil  <https://orcid.org/0000-0002-1279-0199>

Liuba Papeo  <https://orcid.org/0000-0003-3056-8679>

Jean-Rémy Hochmann  <https://orcid.org/0000-0002-4613-1378>

## REFERENCES

- Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience*, *40*(4), 852–863. <https://doi.org/10.1523/JNEUROSCI.1378-19.2019>
- Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. <https://doi.org/10.1111/j.1467-7687.2007.00563.x>
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208. [https://doi.org/10.1016/0010-0277\(85\)90008-3](https://doi.org/10.1016/0010-0277(85)90008-3)
- Baron-Cohen, S. (1997). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *The maladapted mind: Classic readings in evolutionary psychopathology*, (pp. 207–239).
- Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., & Williams, S. C. R. (2000). The amygdala theory of autism. *Neuroscience & Biobehavioral Reviews*, *24*(3), 355–364. [https://doi.org/10.1016/S0149-7634\(00\)00011-7](https://doi.org/10.1016/S0149-7634(00)00011-7)
- Bellet, E., Abassi, E., & Papeo, L. (2021). Moving Toward versus Away from Another: How Body Motion Direction Changes the Representation of Bodies and Actions in the Visual Cortex. *Cerebral Cortex*, *31*(5), 2670–2685. <https://doi.org/10.1093/cercor/bhaa382>
- Birmingham, E., & Kingstone, A. (2009). Human social attention. *Annals of the New York Academy of Sciences*, *1156*, 118–140. <https://doi.org/10.1111/j.1749-6632.2009.04468.x>
- Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166. <https://doi.org/10.1016/j.tics.2006.02.004>
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509–540. <https://doi.org/10.1080/135062899394920>
- Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, *102*(47), 17245–17250. <https://doi.org/10.1073/pnas.0502205102>
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. <https://doi.org/10.3758/BF03208827>
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694. <https://doi.org/10.1037/0033-2909.133.4.694>
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292.
- Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*, 527. <https://doi.org/10.3389/fpsyg.2013.00527>
- Graziano, M. S., & Kastner, S. (2011). Human consciousness and its relationship to social neuroscience: a novel hypothesis. *Cognitive Neuroscience*, *2*(2), 98–113. <https://doi.org/10.1080/17588928.2011.565121>
- Gwiazda, J., Bauer, J., Thorn, F., & Held, R. (1997). Development of Spatial Contrast Sensitivity from Infancy to Adulthood : Psychophysical Data. *Optometry and Vision Science*, *74*(10), 785–789.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. <https://doi.org/10.1038/nature06288>

- Hochmann, J.-R., & Papeo, L. (2014). The invariance problem in infancy: A pupillometry study. *Psychological Science*, 25(11), 2038–2046. <https://doi.org/10.1177/0956797614547918>
- Jin, K. S., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences*, 114(31), 8199–8204. <https://doi.org/10.1073/pnas.1706286114>
- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, 23(8), 672–685. <https://doi.org/10.1016/j.tics.2019.04.013>
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, 18(12), 1084–1089. <https://doi.org/10.1111/j.1467-9280.2007.02029.x>
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, 7(5). <https://doi.org/10.1371/journal.pone.0036399>
- Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science*, 28(11), 1649–1662. <https://doi.org/10.1177/0956797617719930>
- Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2), 50–59. [https://doi.org/10.1016/S1364-6613\(99\)01436-9](https://doi.org/10.1016/S1364-6613(99)01436-9)
- Leslie, A. M. (1982). The perception of causality in infants. *Perception*, 11(2), 173–186. <https://doi.org/10.1068/p110173>
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3), 265–288. [https://doi.org/10.1016/S0010-0277\(87\)80006-9](https://doi.org/10.1016/S0010-0277(87)80006-9)
- Lewkowicz, D. J., & Ghazanfar, A. A. (2012). The development of the uncanny valley in infants. *Developmental Psychobiology*, 54(2), 124–132. <https://doi.org/10.1002/dev.20583>
- Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences*, 115(38), E8835–E8843. <https://doi.org/10.1073/pnas.1801677115>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, 109(18), 6862–6867. <https://doi.org/10.1073/pnas.1113194109>
- Michotte, A. (1946/1963). *The perception of causality*.
- Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review*, 98(2), 164. <https://doi.org/10.1037/0033-295X.98.2.164>
- Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, 61(2), 63–86. <https://doi.org/10.1016/j.cogpsych.2010.02.001>
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756.
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, 104(42), 16598–16603. <https://doi.org/10.1073/pnas.0703913104>
- Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences*, 107(40), 17140–17145. <https://doi.org/10.1073/pnas.0914056107>
- Papeo, L. (2020). Twos in human visual perception. *Cortex*, 132, 473–478. <https://doi.org/10.1016/j.cortex.2020.06.005>
- Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human–human interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 45(7), 877.
- Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual search for people among people. *Psychological Science*, 30(10), 1483–1496. <https://doi.org/10.1177/0956797619867295>
- Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, 28(3), 369–379. <https://doi.org/10.1177/0956797616685769>
- Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences*, 102(14), 5297–5300. <https://doi.org/10.1073/pnas.0406627102>

- Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, *110*, E3965–E3972. <https://doi.org/10.1073/pnas.1304326110>
- Powell, L. J., & Spelke, E. S. (2018). Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition*, *170*, 31–48. <https://doi.org/10.1016/j.cognition.2017.09.007>
- Powell, L. J., & Spelke, E. S. (2018). Third party preferences for imitators in preverbal infants. *Open Mind*, *2*, 61–71. [https://doi.org/10.1162/opmi\\_a\\_00018](https://doi.org/10.1162/opmi_a_00018)
- Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition*, *15*(3), 322–348. <https://doi.org/10.1080/13506280600590434>
- Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*(12), 995–1001. <https://doi.org/10.1111/j.1467-9280.2005.01649.x>
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. [https://doi.org/10.1016/S1364-6613\(00\)01506-0](https://doi.org/10.1016/S1364-6613(00)01506-0)
- Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809–813. <https://doi.org/10.1073/pnas.0707021105>
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, *23*(2), 196–204. <https://doi.org/10.1177/0956797611422072>
- Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*(1), 29–56. [https://doi.org/10.1207/s15516709cog1401\\_3](https://doi.org/10.1207/s15516709cog1401_3)
- Stahl, A. E., & Feigenson, L. (2014). Social knowledge facilitates chunking in infancy. *Child Development*, *85*(4), 1477–1490. <https://doi.org/10.1111/cdev.12217>
- Sun, Y., Stein, T., Liu, W., Ding, X., & Nie, S. P. (2017). Biphasic attentional orienting triggered by invisible social signals. *Cognition*, *168*, 129–139. <https://doi.org/10.1016/j.cognition.2017.06.020>
- Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*, 47–62.
- Thomsen, L., Frankenhuys, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science*, *331*(6016), 477–480. <https://doi.org/10.1126/science.1199198>
- Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, *200*, 104270. <https://doi.org/10.1016/j.cognition.2020.104270>
- Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S. A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*, *148*(7), 1251–1268.
- Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, *198*, 296–302. <https://doi.org/10.1016/j.neuroimage.2019.05.027>
- Walbrin, J., Mihai, I., Landsiedel, J., & Koldewyn, K. (2020). Developmental changes in visual responses to social interactions. *Developmental Cognitive Neuroscience*, *100774*. <https://doi.org/10.1016/j.dcn.2020.100774>
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866–1878. <https://doi.org/10.1121/1.390988>
- Wurm, M. F., & Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature Communications*, *10*(1), 1–10. <https://doi.org/10.1038/s41467-018-08084-y>
- Yin, J., Xu, H., Duan, J., & Shen, M. (2018). Object-based attention on social units: visual selection of hands performing a social interaction. *Psychological Science*, *29*(7), 1040–1048. <https://doi.org/10.1177/0956797617749636>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Goupil, N., Papeo, L., & Hochmann, J.-R. (2022). Visual perception grounding of social cognition in preverbal infants. *Infancy*, *27*, 210–231. <https://doi.org/10.1111/inf.12453>

# Category-specific effects of high-level relations in visual search

Nicolas Goupil\*, Daniel Kaiser & Liuba Papeo\*

*Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, Centre National de la Recherche Scientifique (CNRS) & Université Claude Bernard Lyon1, 67 Bd. Pinel, 69675, Bron, France.*

*Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-University Gießen, Germany;  
Center for Mind, Brain and Behavior (CMBB), Philipps-University Marburg and Justus-Liebig-University Gießen, Germany*

\* **Correspondence to:** N.G. CNRS, Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, 67 Boulevard Pinel, 69675, Bron, France; Phone: +33 043791 1266; E-mail: [goupilnicolas1@gmail.com](mailto:goupilnicolas1@gmail.com); L.P. CNRS, Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, 67 Boulevard Pinel, 69675, Bron, France; Phone: +33 043791 1266; E-mail: [liuba.papeo@isc.cnrs.fr](mailto:liuba.papeo@isc.cnrs.fr)

Declarations of interest: none

## **Abstract**

Recent empirical findings demonstrate that, in visual search for a target in an array of distractors, observers exploit information about object relations to increase search efficiency. We investigated how people searched for interacting people in a crowd, and how the eccentricity of the target affected the search (Experiments 1-3). Participants briefly viewed crowded arrays and had to search for an interacting dyad (two bodies face-to-face) among non-interacting dyads (back-to-back distractors) or *vice versa*, with the target presented in the attended central location or at peripheral locations. With central targets, we found a search asymmetry, whereby interacting people among non-interacting people were detected better than non-interacting people among interacting people. With peripheral targets, non-interacting targets were detected better than interacting targets. In Experiment 4, we asked whether these asymmetries generalized to object pairs whose spatial relations did or did not form functionally interacting sets (computer screen above keyboard). Results showed that non-interacting targets were detected better than interacting targets, whether presented in central or peripheral locations. Thus, the effect of relational information on visual search is contingent on both stimulus category and attentional focus. Across both stimulus categories (bodies and objects), search is facilitated when individual distractor-items can be organized in larger structured units (social interaction or functional set), effectively reducing the number of distractors. The presentation of social interaction at the attended (central) location breaks this search pattern by readily capturing an individual's attention.

**Keywords:** Visual search, asymmetry, social interaction, attention, eccentricity, spatial relation

## Introduction

Each moment in the visual world is a cluttered scene, where structure emerges from the spatial relations between things. Spatial relations can also indicate social relationship. Proximity and *facingness*, the mutual perceptual accessibility of two people, are reliable cues of social interaction, which individuals readily exploit to detect and decipher social events (Papeo et al., 2017; Papeo et al., 2019; Quadflieg & Koldewyn, 2017; Vestner et al., 2019).

Visual search experiments have begun to unravel how the human visual system takes advantage of spatial relations between people to parse crowded scenarios (Papeo et al., 2019; Vestner et al., 2019, 2020; 2021). Visual search asymmetries are traditionally used to document the efficient processing of a class of stimuli (A), relative to another class (B). Finding a target A among a set of distractors Bs may be faster and/or easier than finding a target B among distractors As. When this kind of asymmetry is observed, it can be inferred that A carries features that makes it more salient to visual attention than B (Treisman & Souther, 1985; Wolfe et al., 2001). Using this test of search efficiency, Papeo et al. (2019) showed that, with fast presentation of crowded search arrays, participants were more accurate in reporting the target when it was a dyad of face-to-face –seemingly interacting– bodies (interacting dyad) among a set of non-interacting (back-to-back) dyads (distractors) than a non-interacting dyad among a set of interacting dyads. This effect was independent from set size (*i.e.*, number of distractors in the array), suggesting that interacting dyads *pop out* of the display, involving automatic covert attention (Treisman & Sato, 1990; Wolfe, 1994). The results of Papeo et al. suggested for the first time that two bodies in a spatial relation that cue interaction capture attention more strongly than the same bodies presented as unrelated (see also Vestner et al., 2019; Skripkauskaitė et al., 2022).

This phenomenon has raised questions concerning its selectivity for social stimuli and whether facing dyads are indeed efficiently detected due to their ability to capture attention (Vestner et al., 2021; 2022).

Do interacting people truly capture attention? Papeo et al. (2019) showed a search asymmetry in favor of interacting dyads when the target was always presented at eight predictable locations (out of 16 locations in the array). This was compatible with an interpretation of the effect based on attentional capture, as capture is promoted when attention can be deployed to a limited set of locations in the array (Wolfe, 2007; Neider & Zelinsky, 2008). When, however, the task increased the need for searching through the distractors, to the detriment of immediate attentional capture, the opposite effect was found: more efficient search for non-facing dyads among facing dyads than *vice versa* (Papeo et al., 2019; Experiment 3). The latter pattern was more akin to what was observed with non-social relational stimuli. In particular, Kaiser et al. (2014) first reported that a target object was found more accurately and efficiently when distractor objects were presented in their typical spatial and functional arrangement, *e.g.*, a computer screen above a keyboard, a mirror above a sink, or a lamp above a table, as opposed to a computer screen below a keyboard, a mirror below a sink, or a lamp below a table (see also Kaiser et al., 2015; Kaiser et al., 2019; Stein et al., 2015). This effect was interpreted as the result of automatic grouping of two related objects into a unitary percept, which effectively reduced the number of units to process as distractors, and therefore improved search efficiency.

In sum, at the current state of knowledge, visual search through body dyads gives rise to seemingly incompatible effects: attentional capture yielding search asymmetry in favor of interacting dyads, and more efficient search for non-interacting dyads among interacting dyads than *vice versa*. Moreover, it is unknown to what extent the effects of social relations generalize to non-social relations, as search asymmetry has so far not been probed with object pairs. The current study sheds light on these questions.

Given that, in Papeo et al. (2019), interacting versus non-interacting dyads evoked one effect (search asymmetry in favor of interacting dyads) or the other (more efficient search for non-interacting dyads among interacting dyads than *vice versa*), we hypothesized that performance with body dyads could depend on how much the task favored immediate capture of attention *versus* how much it required search through the distractors. To test this, we manipulated the location of the target, which could appear at the center *versus* the periphery of the search array. When the target is presented around central fixation, few (or no) distractors need to be processed to get to the target, and the resulting performance emphasizes differences in the capacity of the targets (interacting vs. non-interacting dyad) to capture attention. In contrast, interacting and non-interacting dyads are less visible and discriminable at the periphery, as eccentricity (i.e., the distance of a stimulus from central vision) reduces visual acuity (Carrasco et al., 1995; Carrasco & Frieder, 1997) and promotes visual crowding (Bouma, 1970; Whitney & Levi, 2011). Thus, when presented at peripheral locations, the capacity of a target to capture attention is reduced, and the visual search performance depends more heavily on how efficiently the participant can go through the array and reject the distractors. As a result, non-interacting dyads among interacting dyads may be found more efficiently than *vice versa*. The manipulation of configuration (interacting vs. non-interacting) and target eccentricity (central vs. peripheral) was repeated with interacting vs. non-interacting body dyads and with functionally interacting vs. non-interacting object pairs, to test whether the effects of social relations in visual search apply to other (non-social) higher-level relations.

This research was developed in four experiments. In Experiment 1, we studied the effects of target location (central vs. peripheral) in the search asymmetry between interacting (facing) and non-interacting (non-facing) dyads. The target could appear at one of eight locations, four around central fixation and four at more peripheral locations. To verify that our manipulation of target eccentricity (central vs. periphery) was effective, we used eye-tracking to make sure that participants maintained central fixation at the beginning of the trial. We studied whether the advantage for interacting dyads emerged selectively when the target appeared around central fixation, or it could also be found for targets at peripheral locations. To preview, results showed that interacting dyads were detected better than non-interacting dyads only when presented at central locations. The asymmetry was reversed in the direction of an advantage for non-interacting dyads when targets appeared at peripheral locations. Experiments 2-3 provided a replication of the results in Experiment 1 with small variations of the original task. In Experiment 4, participants were tested on the same task, also with functionally interacting versus non-interacting object pairs (e.g., computer above keyboard vs. computer below keyboard). Results showed that, irrespective of eccentricity, non-interacting targets were always found better than interacting targets. We propose that higher-level (social or non-social) relations between distractors generally benefit visual search by enabling grouping of single distractor items in larger meaningful units (social interactions for bodies and functional sets for objects). In addition, an eccentricity-dependent effect unique for social relations demonstrates that people readily attract attention when interacting.

## Experiment 1

Interacting (*i.e.*, face-to-face) bodies in a crowded array are detected better than non-interacting bodies (Papeo et al., 2019; Vestner et al., 2019). This advantage has been interpreted as evidence for an automatic recruitment of attention by social interaction. Here we addressed how target eccentricity modulates this effect, by systematically varying the target location between center and periphery. Eye-tracking was used to control that the participants maintained central fixation when the trial began.

### **Methods**

#### ***Participants***

Twenty-four healthy participants (12 females; age  $M = 25$ ,  $SD = 3.76$ ) were recruited as paid volunteers. Participants had normal or corrected-to-normal vision and reported no psychiatric or neurological history or current medication. The sample size of 24 was estimated *a priori* with the package *BUCSS* (Anderson & Kelley, 2020), based on the results in Papeo et al. (2019; configuration by distractor orientation interaction in Experiment 1:  $F = 8.43$ ;  $\alpha = .05$ ;  $\beta = 0.80$ ). All experiments were approved by the local ethics committee (Comité de Protection des Personnes Sud-Est II). Participants gave informed written consent before participation and were paid 10 euros at the end of the study.

#### ***Stimuli***

Stimuli were search arrays displaying dyads of interacting and non-interacting bodies on a white background. Thirty-three dyads were created starting from 10 bodies in different biomechanically possible poses seen in left or right profile, for a total of 20 bodies. Bodies were gray-scale models created with Daz3D (Daz Productions, Salt Lake City, UT) and the MATLAB image-processing toolbox (The MathWorks, Natick, MA). Bodies were randomly combined in dyads. Each dyad included one body oriented leftward and one body oriented rightward, which could face toward (interacting dyad; Fig. 1A, left) or away from each other (non-interacting dyad; Fig. 1A, right). Distances between the two bodies were matched across interacting and non-interacting dyads, in terms of distance between the centers of the two bodies (interacting:  $M = 210$  pixels,  $SD = 1.47$ ; non-interacting:  $M = 210$  pixels,  $SD = 2.37$ ;  $t(64) = 0.49$ ,  $p > .250$ ), and distance between the two closest extremities of the two bodies (interacting:  $M = 62.57$  pixels,  $SD = 13.27$ ; non-interacting:  $M = 62$  pixels,  $SD = 13.43$ ;  $t(64) = 0.17$ ,  $p > .250$ ).

Each array was composed of two symmetrical halves separated by a central fixation cross (Fig. 1C). Each half was divided into 8 equally sized cells (two columns of 4 cells each), with slightly shifted onsets along the vertical and horizontal axes. Four dyads, all interacting or all non-interacting, appeared on one side of the array. The other side was the mirror version of the first. The two halves differed by only one cell: on either half, this cell featured one interacting dyad (the target) when all other cells featured non-interacting dyads (the distractors) (50% of trials), or one non-interacting dyad (the target) when all other cells displayed interacting dyads (the distractors) (50% of trials). In each array, the distractors could appear in any of the 16 cells, whereas the target could appear in only one of the eight cells on the two middle rows (see gray area in Fig. 1C). Within these eight locations, eccentricity was defined in terms of proximity to the central fixation cross: central locations corresponded to the four cells flanking the fixation cross; peripheral locations corresponded to the four cells at the lateral edges of the array. For each participant, we created a unique set of stimuli that contained 200 arrays with

one interacting target among seven non-interacting distractors (interacting condition; Fig. 1C, left), and 200 arrays with one non-interacting target among seven interacting distractors (non-interacting condition, Fig. 1C, right). Targets occurred at central (50% of trials) or peripheral locations (50% of trials). At a distance of 60 cm, individual dyads subtended approximately  $1.86^\circ \times 2.10^\circ$  of visual angle ( $\sim 0.76^\circ \times \sim 2.01^\circ$  for a single body) and were separated by  $\sim 4.30^\circ$  of visual angle. Thus, the distance between cells was about ten times the distance between bodies within a dyad, preventing the possibility of a dyad spanning two horizontally aligned cells. Central locations were within  $\sim 5.25^\circ$  around the fixation cross; peripheral locations were the outer  $\sim 5.25^\circ$  of the array. Arrays did not exceed  $10.66^\circ \times 10.29^\circ$  of visual angle.

## **Procedures**

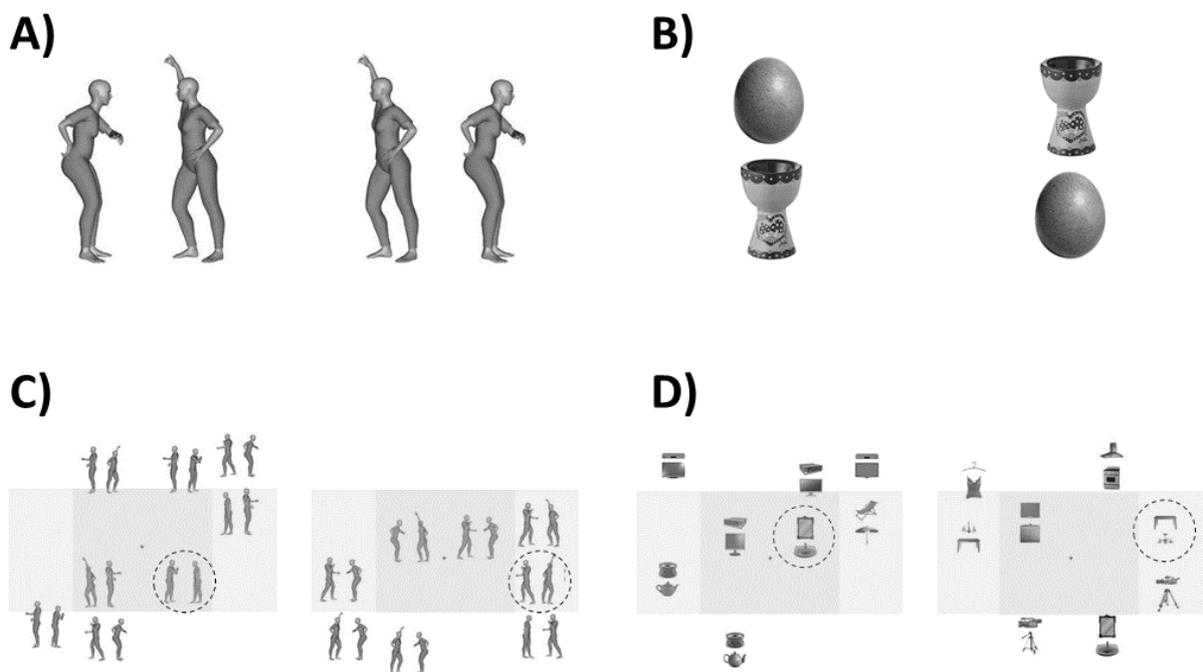
Participants sat on a footstool in a dark soundproof booth, at a distance of  $\sim 60$  cm from a Tobii T60XL eye-tracker screen (60Hz sampling rate). The experiment began with the eye-tracker calibration, which involved fixating a series of crosses, as they appeared on the screen. After the calibration, participants were instructed to fixate the cross at the center of the screen to search for the only interacting dyad (i.e., the facing dyad) among non-interacting dyads (i.e., back-to-back dyads) or, in a separate block, for the only non-interacting dyad among interacting dyads. They had to report whether the target was on the left or right of the central fixation cross. Participants were instructed to fixate the central cross in order to begin the trial. The trial could not begin unless they fixated on the cross. This ensured that peripheral targets were indeed in the visual periphery when the trial started. Each trial began with a central fixation cross (200 ms) followed by a blank screen (700 ms) and a search array (800ms). After the search array disappeared, a blank screen was shown until the participant responded. The next trial began after 1400 ms. Participants were instructed to respond by pressing one of two keys (the “Z” key to respond “left” and the “1” key to respond “right”) on the computer keyboard, with their left or right index finger, respectively. The key assignment (responding “left” by pressing the left key with the left index finger and “right” with the right index finger on the right key) was the same for all participants to avoid stimulus-response incongruence. Participants were invited to take a break every 40 trials and in the interval between the two blocks. Each block began with a familiarization including 16 stimuli, 2 stimuli for each of the eight experimental conditions. The order of blocks (interacting target first or non-interacting target first) was alternated across participants. Stimulus presentation and recording of eye-tracking data were controlled through PsyScopeX (<http://psy.cns.sissa.it>). The experiment lasted  $\sim 30$  min.

## **Behavioral data analysis**

We used R 4.0.2 (R Core Team, 2020) for running analysis of variance (ANOVA; e.g., Lawrence, 2016), and for data visualization (*ggplot2*; Wickham, 2016). Participants with mean accuracy and mean response times (RTs)  $2.5$  *Standard Deviations* (*SD*) away from the group mean, were excluded from later analyses. For the remaining participants, accuracy and RTs were averaged by condition and analyzed with repeated-measures ANOVAs. Pairwise comparisons between critical conditions were performed with *t*-tests. Significance tests were two-tailed, unless *a priori* hypotheses justified one-tailed tests; this was the case for the search asymmetry between interacting non-interacting dyads. The same analytic approach was used in all experiments. All data and scripts for analysis are available on <https://osf.io/vs7m6/>.

## Eye-tracking data analysis

The purpose of eye tracking was to assure that participants had their eyes in central fixation when the trial began, and therefore that our manipulation of the target eccentricity (central vs. peripheral) was effective. A trial began automatically when the participant fixated the cross. We also measured how long during a trial the participants fixated the central location, by computing central dwell times (*i.e.*, the proportion of looking time on the center of arrays, computed as the number of eye-tracking samples recorded on central locations, divided by the total number of eye-tracking samples) by participant and condition. Other measures such as dwell times on target (proportion of looking time on the target's side), number of first fixations on targets, and mean onset times of first fixations (first fixation time-to-onset) were also analyzed.



**Figure 1. Stimuli of Experiments 1-4.** **A)** Examples of interaction and non-interacting body dyads used in Experiments 1-4. **B)** Examples of functionally interacting and non-interacting object pairs used in Experiments 4. **C)** Examples of body-dyad arrays used in Experiments 1-4. Left: condition with an interacting dyad (target) among non-interacting dyads (distractors); right: condition with a non-interacting dyad (target) among interacting dyads (distractors). **D)** Examples of object-pair arrays used in Experiment 4. Left: condition with an interacting pair (target) among non-interacting pairs (distractors); right: condition with a non-interacting pair (target) among interacting pairs (distractors). Distractors could appear at any of sixteen possible locations. The target could appear at one of the eight locations highlighted with grey rectangle: four around central fixations (central locations, darker grey) and four, next to the edges of the arrays (peripheral locations, lighter grey). Targets are indicated by black circles. Neither the rectangles nor the circles were shown in the experiments.

## Results

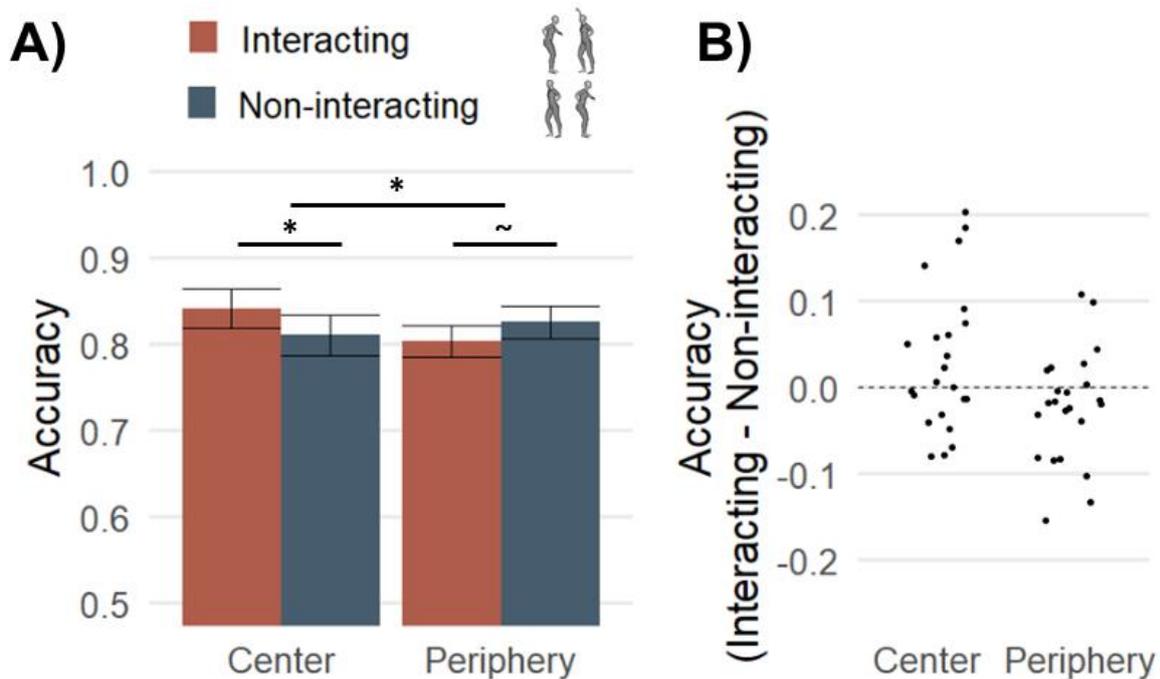
One participant had a mean accuracy 2.5 SD below the group mean and was excluded from further analysis. All other participants had mean RTs within 2.5 SD from the group mean. Respectively, seven and two trials were missed for two participants due to a technical failure of the eye-tracker. These two participants were however included in the analysis. Accuracy

and RTs were analyzed in 2 (target configuration: interacting, non-interacting)  $\times$  2 (target eccentricity: central, peripheral) repeated-measures ANOVAs.

**Accuracy.** Analyzing accuracy (Fig. 2) showed a significant interaction between target configuration and eccentricity,  $F(1, 22) = 5.87, p = .024, \eta_p^2 = 0.21$ , with no main effect of configuration,  $F(1, 22) = 0.17, p = .687, \eta_p^2 = 0.01$ , and no main effect of eccentricity,  $F(1, 22) = 0.97, p = .336, \eta_p^2 = 0.05$ . Pairwise comparisons showed a search asymmetry between interacting and non-interacting targets at central locations,  $t(22) = 1.81, p = .042, d = 0.38$ , where interacting targets were found significantly better than non-interacting targets ( $M_{interacting} = 0.84, SD = 0.11; M_{non-interacting} = 0.81, SD = 0.11$ ). The asymmetry tended to reverse with targets at peripheral locations,  $t(22) = -1.66, p = .056, d = -0.35$ , where accuracy was higher for non-interacting targets ( $M = 0.83, SD = 0.09$ ) than interacting targets ( $M = 0.80, SD = 0.09$ ). Moreover, interacting-dyad targets were more likely to be detected in central (vs. peripheral) locations,  $t(22) = 2.07, p = .050, d = 0.43$ , while there was no effect of eccentricity for non-interacting targets,  $t(22) = -1.15, p = .262, d = -0.24$ .

**RTs.** There was a significant effect of eccentricity,  $F(1, 22) = 7.52, p = .012, \eta_p^2 = 0.42$ , as participants were faster with targets at central than peripheral locations ( $M_{center} = 1267, SD = 266; M_{periphery} = 1337, SD = 249$ ). There was no main effect of configuration,  $F(1, 22) = 0.25, p = .625, \eta_p^2 = 0.06$ , and no interaction  $F(1, 22) = 0.67, p = .421, \eta_p^2 = 0.03$ .

**Eye-tracking results.** A trial began when the participant fixated the central fixation cross. Eye-tracking data analysis showed that the participants' eyes remained in the central area for most of the trial duration, with a mean proportion of central (vs. peripheral) dwell time of 0.74 ( $SD = 0.07$ ). A repeated-measures ANOVA on central dwell times, with a 2 (target configuration: interacting, non-interacting)  $\times$  2 (target eccentricity: central, peripheral) design showed an effect of eccentricity,  $F(1, 22) = 388.56, p < .001, \eta_p^2 = 0.96$ , as participants' look departed from central fixation more often when the target was at the periphery ( $M = 0.67, SD = 0.07$ ) than in the center ( $M = 0.80, SD = 0.08$ ). The effect of configuration,  $F(1, 22) = 1.95, p = .176, \eta_p^2 = 0.42$ , and the interaction,  $F(1, 22) = 0.06, p = .814, \eta_p^2 = 0.00$ , were not significant. The analyses of dwell times on target, number of first fixations on targets, and mean onset times of first fixations showed no effects (see Supplementary material 1), meaning that the above effects on accuracy and RTs were independent from overt eye movements.



**Figure 2. Results of Experiment 1. (A) Mean Accuracy as a function of target configuration (interacting/non-interacting) and target eccentricity (center/periphery).** Bar plots represent the mean group accuracy with interacting dyads and non-interacting dyads; error bars are standard errors of the mean. \* denote significant effects ( $p < 0.05$ ); ~ denote trends ( $p < .10$ ). **(B) Individual mean accuracy differences between interacting and non-interacting dyads.** Positive values indicate higher accuracy with interacting dyads, negative values, higher accuracy with non-interacting dyads; dots indicate individual participants.

**Summary.** Consistent with previous reports (Papeo et al., 2019), accuracy rates were more sensitive than RTs to experimental manipulations in the current task. Participants were more likely to correctly report an interacting dyad (target) among non-interacting dyads (distractors), when the target appeared around central fixation than at the periphery. However, the lack of search asymmetry at the periphery does not mean that participants did not detect the target at the periphery: performance with targets at peripheral locations was well above chance ( $M = 0.81$ ,  $SD = 0.08$ ,  $t(22) = 17.90$ ,  $p < .001$ ,  $d = 3.73$ ) and participants gazed away from central fixation more often when the target was at a peripheral location. These results suggest that the asymmetry, with the advantage for interacting dyads, only emerged when the target appeared at the attended central location. However, when the task required searching through the array (*i.e.*, when the target was at peripheral locations outside fixation), the effect tended to reverse, as non-interacting targets were reported more accurately than interacting target.

## Experiment 2

Experiment 1 showed an advantage for interacting over non-interacting targets in the center of array, but not at the periphery. Experiment 2 sought to replicate the results of Experiment 1 with two changes. First, Experiment 2 was carried out online, without eye-tracking. Second, we varied the set size, including arrays of both four and eight dyads (or eight and 16 bodies). The latter manipulation was introduced to test to what extent performance was affected by the number of distractors, taking the effect of set size (*i.e.*, slower RTs for the larger set size) as an indication of serial search through the array (Treisman & Gelade, 1980).

## Methods

### Participants

Experiment 2 was implemented on Testable.org ([www.testable.org/](http://www.testable.org/)) (Rezlescu et al., 2020), a platform for stimulus presentation and response recording, with built-in procedures for taking care of differences in screen resolution among different private devices. Twenty-four healthy participants (4 females; age:  $M = 24$  years,  $SD = 4.19$ ) were recruited as paid volunteers. Participants were included provided that they had a track record of “correct participants” in 90% of the online studies in Testable.org, and gave informed consent.

### Stimuli and procedures

Due to limited online storage capacity of the online platform, we created only two lists of 400 search arrays. Participants were randomly assigned to one of the two lists. In each list, half of the arrays contained an interacting target among three (100 trials) or seven (100 trials) non-interacting distractors; the other half contained a non-interacting target among three (100 trials) or seven (100 trials) interacting distractors. Targets appeared randomly at central (50% trials) or peripheral locations.

Participants accessed the experiment through a link shared on Testable.org. Before the experiment, they were invited to follow the instructions to calibrate the size of the display, in order to ensure that the image size was similar across different monitors. Next, participants were asked to set their room in a dim light, install the computer on a stable table, sit on a stable seat, align their eyes to the center of the computer’s screen at a distance of 60cm (arm length), and turn off the sound. Task instructions were identical to Experiment 1, except that here participants were asked to use the “e” key or the “o” key, to respond “left” or “right”, respectively. Familiarization, stimulus presentation, and trial structure were identical to Experiment 1.

## Results

All participants had mean accuracy and RTs within 2.5  $SD$  from the group average and were included in following analyses. Mean accuracy and RTs were analyzed in 2 (target configuration: interacting, non-interacting)  $\times$  2 (target eccentricity: central, peripheral)  $\times$  2 (set size: set 4, set 8) repeated-measures ANOVAs.

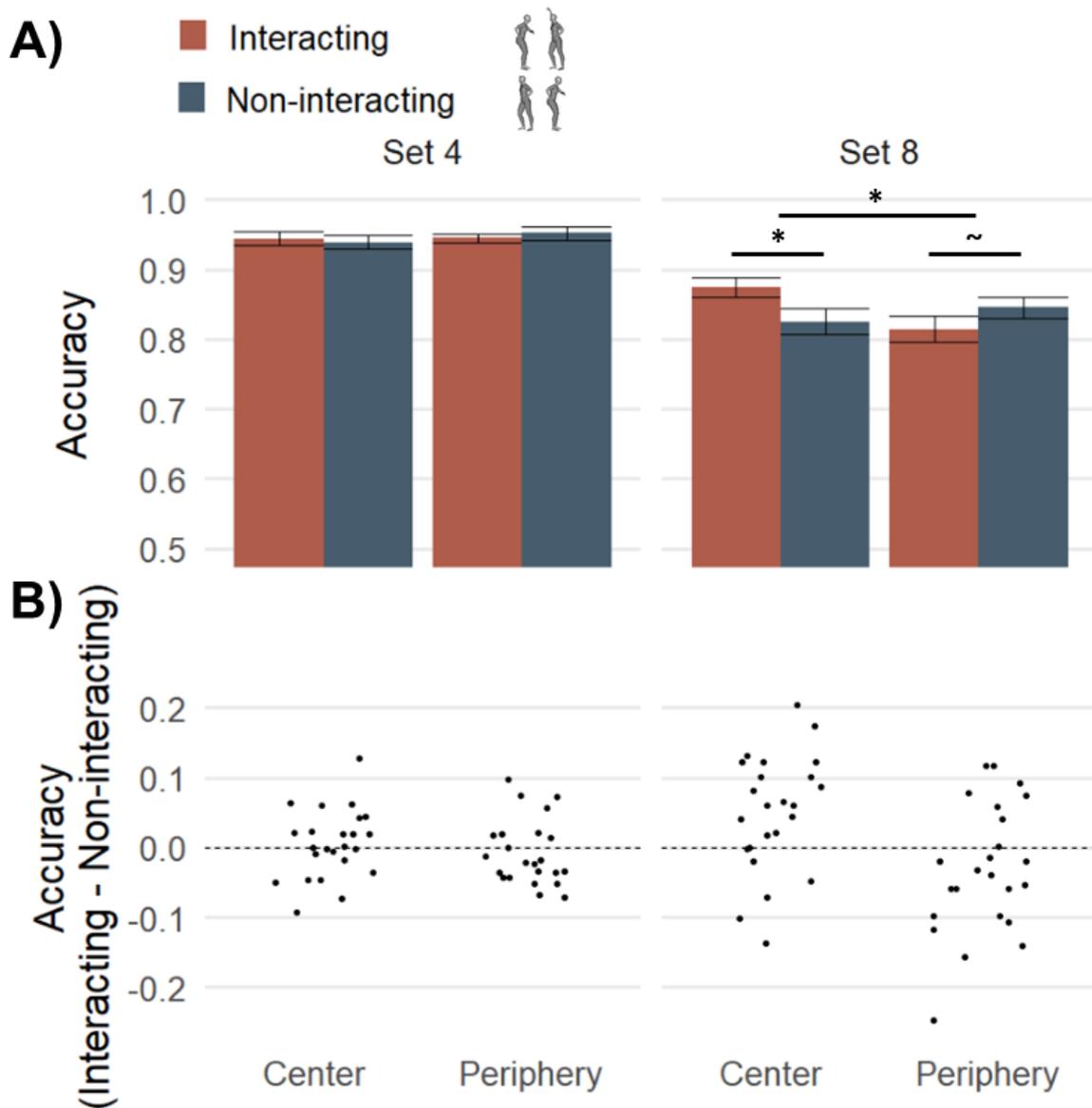
**Accuracy** (Fig. 3). We found a significant effect of set size,  $F(1, 23) = 88.58$ ,  $p < .001$ ,  $\eta_p^2 = 0.95$ , a significant interaction between target configuration and eccentricity,  $F(1, 23) = 8.04$ ,  $p = .009$ ,  $\eta_p^2 = 0.48$ , and a significant interaction between target configuration, target eccentricity and set size,  $F(1, 23) = 11.62$ ,  $p = .002$ ,  $\eta_p^2 = 0.34$ . Other main effects and interactions were not significant (configuration:  $F(1, 23) = 0.28$ ,  $p = .600$ ,  $\eta_p^2 = 0.03$ ; eccentricity  $F(1, 23) = 0.66$ ,  $p = .425$ ,  $\eta_p^2 = 0.07$ ; configuration  $\times$  set size  $F(1, 23) = 0.44$ ,  $p = .513$ ,  $\eta_p^2 = 0.04$ ; eccentricity  $\times$  set size  $F(1, 23) = 3.15$ ,  $p = .089$ ,  $\eta_p^2 = 0.23$ ).

***Effect of eccentricity in the search for interacting vs. non-interacting targets.*** To understand the three-way interaction, we ran two separate ANOVAs for the two different set size conditions, which showed that results of Experiment 1 were fully replicated in Experiment 2. In particular, consistent with Experiment 1, the critical target configuration by eccentricity interaction was significant with set size of eight,  $F(1, 23) = 13.11$ ,  $p = .001$ ,  $\eta_p^2 = 0.36$  (effect of configuration:  $F(1, 23) = 0.42$ ,  $p = .525$ ,  $\eta_p^2 = 0.03$ ; effect of eccentricity  $F(1, 23) = 2.15$ ,  $p = .156$ ,  $\eta_p^2 = 0.12$ ), revealing a search asymmetry in favor of interacting targets with targets at central locations

( $M_{interacting} = 0.87$ ,  $SD = 0.07$ ;  $M_{non-interacting} = 0.83$ ,  $SD = 0.09$ ;  $t(23) = 2.85$ ,  $p = .004$ ,  $d = 0.58$ ), but not with targets at peripheral locations, where, again, an opposite trend was observed: better performance with non-interacting, than with interacting targets ( $M_{interacting} = 0.81$ ,  $SD = 0.09$ ,  $M_{non-interacting} = 0.85$ ,  $SD = 0.07$ ;  $t(23) = -1.63$ ,  $p = .059$ ,  $d = -0.33$ ). Moreover, interacting targets were detected better at the center than at the periphery,  $t(23) = 3.63$ ,  $p = .001$ ,  $d = 0.74$ , but this was not the case for non-interacting targets,  $t(23) = -1.11$ ,  $p = .278$ ,  $d = -0.23$ . The target configuration by eccentricity interaction was not significant with set size of four,  $F(1, 23) = 0.63$ ,  $p = .435$ ,  $\eta_p^2 = 0.03$  (effect of configuration:  $F(1, 23) = 0.02$ ,  $p = .894$ ,  $\eta_p^2 = 0.00$ ; eccentricity:  $F(1, 23) = 0.67$ ,  $p = .421$ ,  $\eta_p^2 = 0.03$ ).

*Effect of set size in the search for interacting vs. non-interacting targets.* We investigated the effect of set size in the search for interacting vs. non-interacting targets in two separate ANOVAs with factors target eccentricity (central, peripheral) and set size (set 4, set 8). Results showed a significant interaction for interacting targets ( $F(1, 23) = 14.01$ ,  $p = .001$ ,  $\eta_p^2 = 0.38$ ; effect of eccentricity:  $F(1, 23) = 7.73$ ,  $p = .011$ ,  $\eta_p^2 = 0.37$ ; effect of set size:  $F(1, 23) = 43.07$ ,  $p < .001$ ,  $\eta_p^2 = 0.87$ ), showing that the effect of set size was significantly larger in the search for interacting targets at the periphery,  $t(23) = -7.02$ ,  $p < .001$ ,  $d = -1.43$ , compared to the search for interacting targets at central locations,  $t(23) = -4.44$ ,  $p < .001$ ,  $d = -0.91$ . With non-interacting targets, there was only a main effect of set size,  $F(1, 23) = 93.24$ ,  $p < .001$ ,  $\eta_p^2 = 0.85$ , showing that performance was better with the smaller set size arrays, regardless of whether the target was at the center of the periphery. There was no effect of eccentricity,  $F(1, 23) = 1.78$ ,  $p = .195$ ,  $\eta_p^2 = 0.11$ , or interaction,  $F(1, 23) = 0.16$ ,  $p = .694$ ,  $\eta_p^2 = 0.01$ .

**RTs.** We only found an effect of set size,  $F(1, 23) = 17.40$ ,  $p < .001$ ,  $\eta_p^2 = 0.30$ , showing that performance was faster with set size of four than eight. All other effects and interactions were not significant (configuration:  $F(1, 23) = 0.00$ ,  $p = .974$ ,  $\eta_p^2 = 0.00$ ; eccentricity:  $F(1, 23) = 0.01$ ,  $p = .915$ ,  $\eta_p^2 = 0.00$ ; configuration x eccentricity:  $F(1, 23) = 0.98$ ,  $p = .333$ ,  $\eta_p^2 = 0.02$ ; configuration x set size:  $F(1, 23) = 0.42$ ,  $p = .524$ ,  $\eta_p^2 = 0.01$ ; eccentricity x set size:  $F(1, 23) = 1.15$ ,  $p = .295$ ,  $\eta_p^2 = 0.05$ ; configuration x eccentricity x set size:  $F(1, 23) = 1.27$ ,  $p = .271$ ,  $\eta_p^2 = 0.05$ ).



**Figure 3. Results of Experiment 2. (A) Mean Accuracy as a function of target configuration (interacting/non-interacting), target eccentricity (center/periphery) and set size (four/eight dyads).** Bar plots represent the mean group accuracy with interacting and non-interacting dyads; error bars are standard errors from the mean. \* denote significant effects ( $p < 0.05$ ); ~ denote trends ( $p < .10$ ). **(B) Individual mean accuracy differences between interacting and non-interacting dyads.** Positive values indicate higher accuracy with interacting dyads, negative values higher accuracy with non-interacting dyads, dots indicate individual participants.

**Summary.** Accuracy results with set size of eight fully replicated the results of Experiment 1, showing a search asymmetry in favor of interacting dyads at the center, and a trend for an advantage for non-interacting targets at the periphery. Moreover, the search for interacting targets was less affected by the number of distractors (set size) at the center than at the periphery, in line with the hypothesis of a stronger recruitment of attention by an interacting target in the center *versus* greater reliance on serial search through the distractors with the target at the periphery. The same accuracy pattern was not found with set size of four, where

performance approached ceiling, thus reducing the sensitivity to detect differences across conditions.

RT results only showed a general effect of set size (faster performance with set size of four than eight) but were not sensitive to other variations across conditions. In the current study, we opted for short stimulus duration as we were primarily interested in addressing the search advantage related to spontaneous attentional capture by the target. Different task conditions (*e.g.*, unlimited stimulus presentation times) may help to highlight effects in RTs, which is necessary to evaluate the effects of set size on visual search using search slopes.

## Experiment 3

Experiments 1-2 showed that the search advantage for interacting dyads emerged with targets presented at the attended (central) spatial location. In addition, Experiments 1-2 showed that the pattern tended to reverse towards an advantage for non-interacting dyads, with targets at the periphery, a condition that increases the demand for a search through the array, beyond the attended location. Here, we asked whether this reversal of effects would become statistically more reliable when extending the stimulus duration to allow more time to explore the periphery of the visual field.

### **Methods**

#### ***Participants***

Twenty-four new healthy participants (7 females; age:  $M = 26$  years,  $SD = 4.96$ ) were recruited as paid volunteers. Participants were included if they had a track record of “correct participant” in 90% of the online studies in Testable.org, and provided informed consent.

#### ***Procedures***

Stimuli, familiarization, instructions and procedures were identical to Experiment 2, except for two details. First, stimulus duration was increased by 50%, for a total of 1200 ms per stimulus. Second, the set size was kept constant and corresponded to eight dyads. Each participant saw 400 arrays featuring one interacting target among seven non-interacting distractors (200 trials), or one non-interacting target among seven interacting distractors (200 trials). Targets appeared randomly in central (50% trials) or peripheral locations.

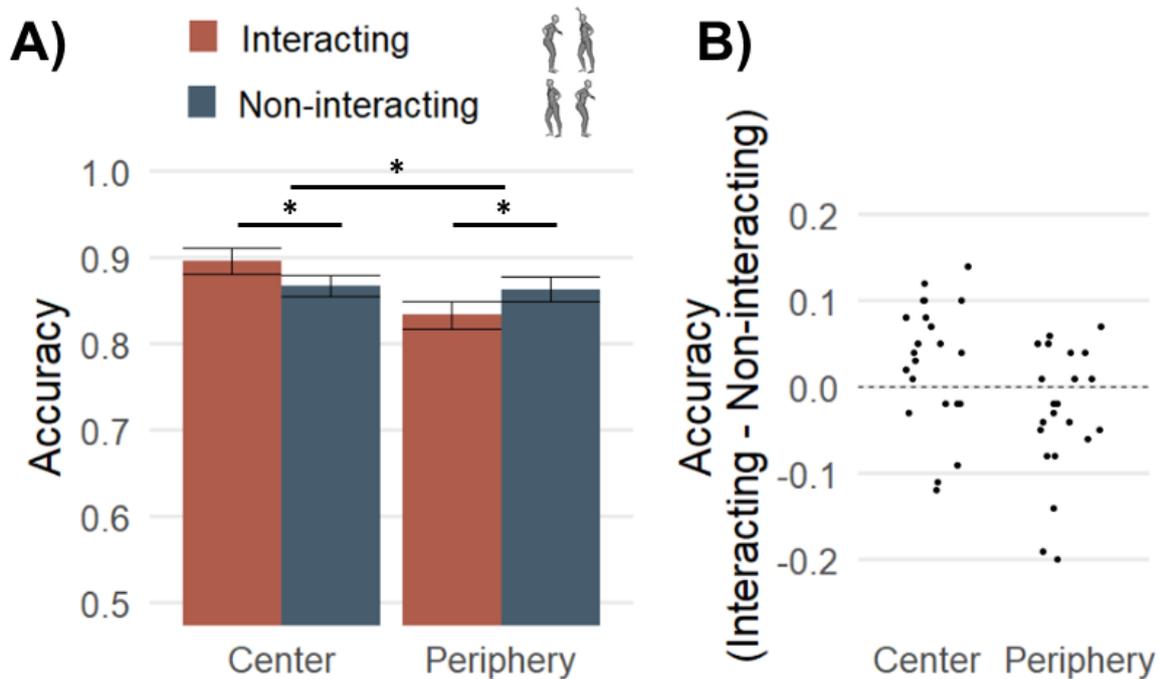
### **Results**

We excluded data from two participants who had a mean accuracy 2.5  $SD$  below the group mean and mean RTs more than 2.5  $SD$  away from the group mean. Mean accuracy rates and RTs of the remaining participants were analyzed in 2 (target configuration: interacting, non-interacting)  $\times$  2 (target eccentricity: central, peripheral) repeated-measures ANOVAs.

**Accuracy** (Fig. 4). We found an interaction between target configuration and eccentricity,  $F(1, 21) = 7.44$ ,  $p = .013$ ,  $\eta_p^2 = 0.26$ , a main effect of eccentricity,  $F(1, 21) = 5.79$ ,  $p = .025$ ,  $\eta_p^2 = 0.32$ , but no effect of configuration,  $F(1, 21) = 0.01$ ,  $p = .938$ ,  $\eta_p^2 = 0.00$ . Pairwise comparisons showed a significant search asymmetry in favor of interacting targets at central locations ( $M_{interacting} = 0.90$ ,  $SD = 0.07$ ;  $M_{non-interacting} = 0.87$ ,  $SD = 0.06$ ;  $t(21) = 1.82$ ,  $p = .041$ ,  $d = 0.39$ ), and the reversed asymmetry at peripheral locations,  $t(21) = -1.87$ ,  $p = .038$ ,  $d = -0.40$ , where participants were more accurate for non-interacting dyads ( $M = 0.86$ ,  $SD = 0.07$ ) than

interacting dyads ( $M = 0.83$ ,  $SD = 0.08$ ). Moreover, interacting targets were found better at the center than at the periphery,  $t(21) = 3.28$ ,  $p = .004$ ,  $d = 0.70$ . This effect was not found for not non-interacting targets,  $t(21) = 0.26$ ,  $p = .797$ ,  $d = 0.06$ .

**RTs.** We found a significant main effect of eccentricity,  $F(1, 21) = 17.08$ ,  $p < .001$ ,  $\eta_p^2 = 0.30$ , as participants were faster with central ( $M = 2001$ ,  $SD = 279$ ) than peripheral targets ( $M = 2061$ ,  $SD = 282$ ). There was no main effect of configuration,  $F(1, 21) = 2.24$ ,  $p = .150$ ,  $\eta_p^2 = 0.38$ , and no interaction,  $F(1, 21) = 0.09$ ,  $p = .769$ ,  $\eta_p^2 = 0.00$ .



**Figure 4. Results of Experiment 3. (A) Mean Accuracy as a function of target configuration (interacting/non-interacting) and target eccentricity (center/periphery).** Bar plots represent the mean group accuracy with interacting and non-interacting dyads; error bars are standard errors from the mean. \* denote significant effects ( $p < 0.05$ ). **(B) Individual mean accuracy differences between interacting and non-interacting dyads.** Positive values indicate higher accuracy with interacting dyads, negative values higher accuracy with non-interacting dyads; dots indicate individual participants.

**Summary.** The longer stimulus presentation implemented in Experiment 3 did not change the search advantage for interacting targets at central locations: participants detected interacting dyads better than non-interacting dyads, when the target appeared at the attended central location. With longer stimulus presentation, however, the advantage for non-interacting targets at the periphery, which was a trend in Experiments 1-2, here became statistically reliable. This result indeed indicates that target presentation at peripheral locations induced serial search through the array and that this search benefited from a meaningful, structured configuration of the distractors. Altogether Experiments 1-3 suggest that the advantage for interacting dyads can be explained by fast and automatic perceptual mechanisms operating on the attended location, such as a perceptual enhancement of parts (*i.e.*, the interacting dyad) of a complex scene. While making the task overall more difficult (see main effect of eccentricity in Experiments 1-3), peripheral target presentation did not just abolish the above asymmetry, but

reversed the effect: participants were more likely to detect non-interacting dyads than interacting dyads, suggesting a change in the underlying mechanism, from immediate attentional capture when a salient target (the interacting dyad) was in central vision, to more efficient serial search through arrays of distractors that could be processed more efficiently (the –grouped– interacting dyads). To assess the statistical significance of this reversal, the mean accuracy of participants from Experiments 1, 2 and 3 was pooled, and entered in a 2 (target configuration: interacting, non-interacting) × 2 (target eccentricity: central, peripheral) repeated-measures ANOVAs. This found a significant interaction between target configuration and eccentricity,  $F(1, 68) = 26.29, p < .001, \eta_p^2 = 0.28$ , a main effect of eccentricity,  $F(1, 68) = 8.10, p = .006, \eta_p^2 = 0.14$ , but no effect of configuration,  $F(1, 68) = 0.38, p = .537, \eta_p^2 = 0.01$ .

## Experiment 4

Experiments 1-3 revealed a complex pattern of asymmetries in visual search for interacting or non-interacting dyads: interacting dyads were detected better in central vision and non-interacting dyads were detected better at the periphery. In Experiment 4, we tested whether the pattern found for interacting vs. non-interacting dyads in Experiments 1-3 generalized to object pairs appearing in their regular spatial and functional arrangement (hereafter: interacting pairs; e.g., a computer screen above a keyboard, a mirror above a sink, or a lamp above the table), compared with object pairs arranged in an irregular way (hereafter: non-interacting pairs; e.g., a computer screen below a keyboard, a mirror below a sink, or a lamp below the table).

### **Methods**

#### ***Participants***

A new group of twenty-four healthy participants (19 females; age  $M = 22, SD = 2.83$ ) were recruited and tested in the lab. Participants were paid volunteers with normal or corrected-to-normal vision, no psychiatric history or current medication reported. They were enrolled after providing informed consent.

#### ***Stimuli and procedures***

Search arrays were created from eleven pairs of everyday gray-scale objects used in previous studies (Kaiser et al., 2014). In the arrays, the pairs could appear in their typical vertical configuration for common usage (interacting pairs: e.g., a lamp above a dining table, a mirror above a bathroom sink, or an air vent above a stove; Fig. 1B, left), or vertically swapped (non-interacting pairs: e.g., a lamp below a table, a mirror below a bathroom sink, an air vent below a stove; Fig. 1B, right). For each single objects (e.g., table), two different exemplars were included (table A and table B), resulting in four different exemplars for each pair (table A-lamp A, table B-lamp A, table B-lamp B), and a total of 44 functionally interacting and 44 non-interacting pairs. Arrays were generated following the same procedure described for body dyads, except for two differences. First, search arrays did not display more than one pair from the same type (e.g., only one lamp-table pair per display). Second, the two halves of the arrays contained different pairs (i.e., one half was not the mirror version of the other half, as in Experiments 1-3). This was done to prevent the target to be found because it corresponded to the only asymmetric part of the array. Search arrays with body dyads were modified relative to Experiment 1, to match these features of the object-pair arrays (only one repetition of the same dyad in each array and different dyads in the two halves of the array).

For each participant, 800 unique arrays were generated, 400 with body dyads and 400 with object pairs. Half of the arrays included a target among seven distractors; the other half included a target among nine distractors (Fig. 1D). Targets occurred in a central (50% trials) or at a peripheral location. Arrays with object pairs and arrays with body dyads were presented in two different blocks.

In the object-pair condition, in two separate blocks, participants were instructed to search for the only interacting pair among the non-interacting pairs (Fig. 1D, left), or for the only non-interacting pair among the interacting pairs (Fig. 1D, right). In the body-dyad conditions, in two separate blocks, participants were instructed to search for the only interacting dyad among the non-interacting dyad pairs, or for the only non-interacting dyad among the interacting dyads, as in previous experiments. The order of blocks was alternated across participants: the experiment started with either body-dyads or object-pairs, and within each type-condition, with either interacting or non-interacting targets. Before each block, participants were familiarized with the instructions and stimuli of each sub-block, in the same order, with 8 trials presenting 4 stimuli for each of the 2 experimental conditions. Arrays were displayed on a 17-in. CRT monitor (1024 × 768 pixel resolution, 85-Hz refresh rate) positioned 60 cm from the participant's eyes. Participants sat on a height-adjustable chair with their eyes aligned to the center of the screen. Stimulus presentation and response collection were controlled through the Psychophysics Toolbox extension of MATLAB (Brainard, 1997). The entire experiment lasted ~60 min.

## Results

One participant with a mean accuracy rate 2.5 SD below the group mean was discarded. All other participants had mean RTs within 2.5 SD from the group mean and were included in analysis. Mean accuracy rates and RTs were analyzed in 2 (target configuration: interacting, non-interacting) × 2 (target eccentricity: central, peripheral) × 2 (set size: set 8, set 10) × 2 (target type: bodies, objects) repeated-measures ANOVAs.

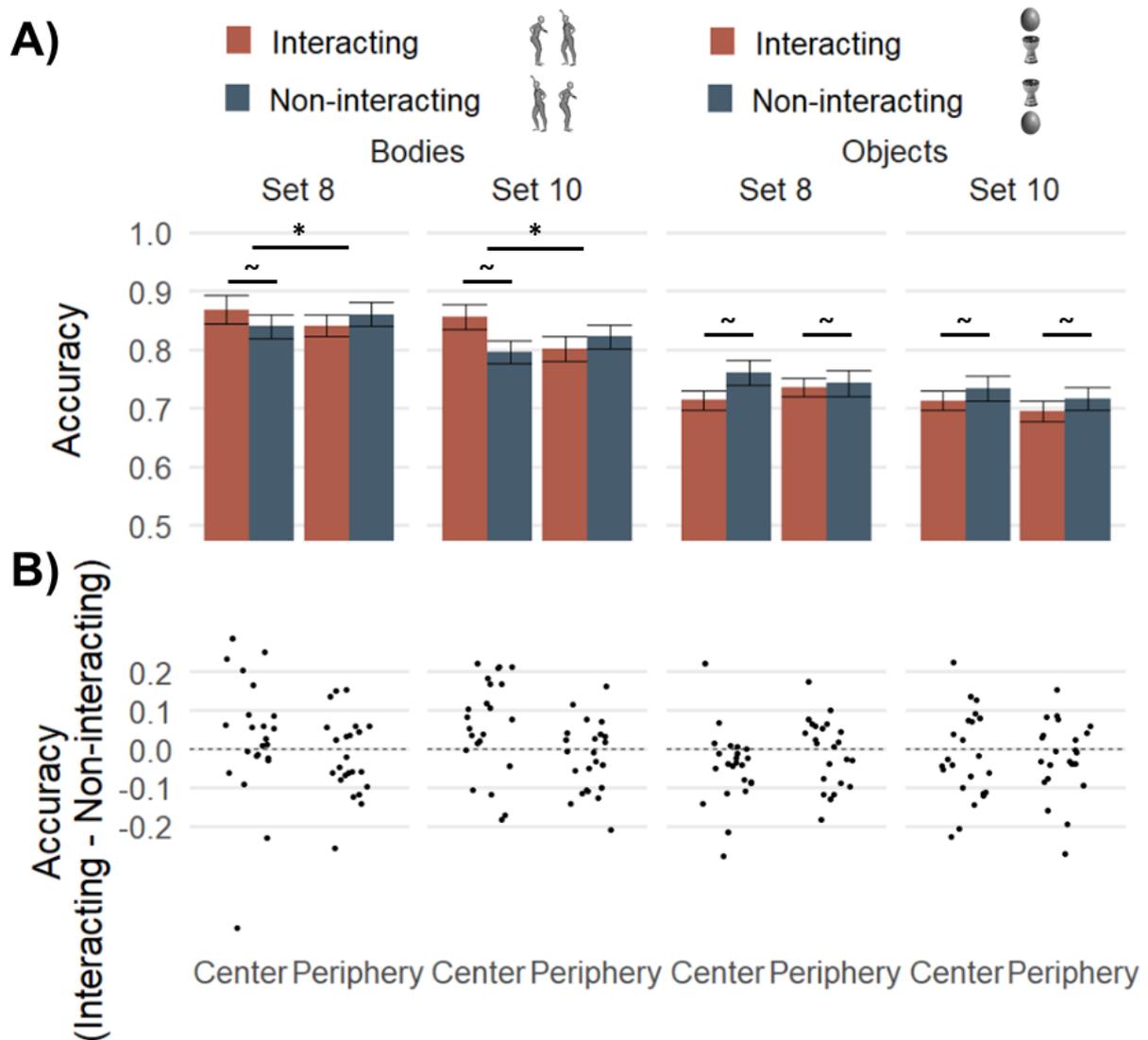
**Accuracy** (Fig. 5). We found an effect of set size,  $F(1, 22) = 27.12$ ,  $p < .001$ ,  $\eta_p^2 = 0.48$ , as participants were more accurate with smaller set size arrays ( $M_{eight} = 0.80$ ,  $SD = 0.06$ ;  $M_{ten} = 0.77$ ,  $SD = 0.07$ ), an effect of target type,  $F(1, 22) = 114.132$ ,  $p < .001$ ,  $\eta_p^2 = 0.93$ , as participants were more accurate with bodies ( $M = 0.84$ ,  $SD = 0.07$ ) than objects ( $M = 0.73$ ,  $SD = 0.07$ ), and a significant interaction between target configuration, eccentricity and target type,  $F(1, 22) = 5.71$ ,  $p = .026$ ,  $\eta_p^2 = 0.33$ . Other effects and interactions were not significant (configuration:  $F(1, 22) = 0.55$ ,  $p = .467$ ,  $\eta_p^2 = 0.04$ ; eccentricity:  $F(1, 22) = 0.90$ ,  $p = .353$ ,  $\eta_p^2 = 0.07$ ; configuration × eccentricity:  $F(1, 22) = 1.65$ ,  $p = .213$ ,  $\eta_p^2 = 0.13$ ; configuration × set size:  $F(1, 22) = 0.69$ ,  $p = .414$ ,  $\eta_p^2 = 0.03$ ; configuration × type:  $F(1, 22) = 2.04$ ,  $p = .167$ ,  $\eta_p^2 = 0.27$ ; eccentricity × set size:  $F(1, 22) = 1.81$ ,  $p = .193$ ,  $\eta_p^2 = 0.06$ ; eccentricity × type:  $F(1, 22) = 0.00$ ,  $p > .999$ ,  $\eta_p^2 = 0.00$ ; set size × type:  $F(1, 22) = 0.75$ ,  $p = .396$ ,  $\eta_p^2 = 0.02$ ; configuration × eccentricity × set size:  $F(1, 22) = 3.41$ ,  $p = .078$ ,  $\eta_p^2 = 0.09$ ; configuration × set size × type:  $F(1, 22) = 0.13$ ,  $p = .717$ ,  $\eta_p^2 = 0.01$ ; eccentricity × set size × type:  $F(1, 22) = 0.16$ ,  $p = .695$ ,  $\eta_p^2 = 0.01$ ; configuration × eccentricity × set size × type:  $F(1, 22) = 0.02$ ,  $p = .900$ ,  $\eta_p^2 = 0.00$ ).

To understand the interaction, we computed two 2 (target configuration: interacting, non-interacting) × 2 (target eccentricity: central, peripheral) repeated-measures ANOVAs for body- and object-trials, separately. For bodies, we found a significant interaction between target configuration and eccentricity,  $F(1, 22) = 5.52$ ,  $p = .028$ ,  $\eta_p^2 = 0.20$ , with no main effects of configuration,  $F(1, 22) = 0.50$ ,  $p = .488$ ,  $\eta_p^2 = 0.03$ , or eccentricity,  $F(1, 22) = 0.47$ ,  $p = .499$ ,  $\eta_p^2 = 0.02$ . Like in Experiments 1-3, interacting targets were detected better at the center than

at the periphery,  $t(22) = 2.21, p = .038, d = 0.46$ , while this was not the case for non-interacting targets,  $t(22) = -1.32, p = .200, d = -0.28$ . Moreover, consistent with Experiments 1-3, there was a trend for a search asymmetry in favor of interacting dyads, with targets in central locations ( $M_{interacting} = 0.86, SD = 0.10; M_{non-interacting} = 0.82, SD = 0.08; t(22) = 1.68, p = .054, d = 0.35$ ), but not with targets at the periphery ( $M_{interacting} = 0.82, SD = 0.08; M_{non-interacting} = 0.84, SD = 0.09; t(22) = -1.27, p = .108, d = -0.27$ ). Numerically, however, the search for the target at the periphery matched the effect in Experiments 1: better search for non-interacting than interacting targets. A weaker effect here could reflect design differences between the previous experiments and the current one, which involved modified search arrays and body-trials interspersed with object-trials in the same session.

For object pairs, the ANOVA only showed a trend for the effect of configuration,  $F(1, 22) = 3.49, p = .075, \eta_p^2 = 0.18$ : regardless the location, non-interacting targets tended to be detected more successfully than interacting targets. There was no effect of eccentricity,  $F(1, 22) = 0.88, p = .358, \eta_p^2 = 0.03$ , or interaction,  $F(1, 22) = 0.75, p = .395, \eta_p^2 = 0.03$ .

**RTs.** RTs showed effects of eccentricity,  $F(1, 22) = 11.89, p = .002, \eta_p^2 = 0.62$ , target type,  $F(1, 22) = 35.52, p < .001, \eta_p^2 = 0.99$ , and set size  $F(1, 22) = 43.32, p < .001, \eta_p^2 = 0.77$ . The interaction between set size and target type was significant,  $F(1, 22) = 6.26, p = .020, \eta_p^2 = 0.34$ , reflecting a larger effect of set size (faster responses with smaller than larger sets) for body-trials,  $t(22) = -6.45, p < .001, d = -1.35$ , than objects  $t(22) = -2.73, p = .012, d = -0.57$  (8 pairs:  $M = 1501, SD = 219$ ; 10 pairs:  $M = 1535, SD = 218$ ). Finally, there was a significant interaction between target configuration and eccentricity,  $F(1, 22) = 4.63, p = .043, \eta_p^2 = 0.21$ . This interaction showed that the effect of eccentricity (faster responses at central than peripheral locations) was larger for interacting bodies/objects,  $t(22) = -3.39, p = .003, d = -0.71$ , than for non-interacting bodies/objects  $t(22) = -2.27, p = .033, d = -0.47$ . Other effects and interactions were not significant (configuration:  $F(1, 22) = 0.86, p = .364, \eta_p^2 = 0.39$ ; configuration x set size:  $F(1, 22) = 0.54, p = .469, \eta_p^2 = 0.04$ ; configuration x type:  $F(1, 22) = 3.35, p = .081, \eta_p^2 = 0.58$ ; eccentricity x set size:  $F(1, 22) = 0.02, p = .889, \eta_p^2 = 0.00$ ; eccentricity x type:  $F(1, 22) = 0.55, p = .467, \eta_p^2 = 0.06$ ; configuration x eccentricity x set size:  $F(1, 22) = 1.20, p = .286, \eta_p^2 = 0.11$ ; configuration x eccentricity x type:  $F(1, 22) = 1.47, p = .238, \eta_p^2 = 0.10$ ; configuration x set size x type:  $F(1, 22) = 0.47, p = .502, \eta_p^2 = 0.04$ ; eccentricity x set size x type:  $F(1, 22) = 0.06, p = .809, \eta_p^2 = 0.00$ ; configuration x eccentricity x set size x type:  $F(1, 22) = 1.08, p = .311, \eta_p^2 = 0.05$ ).



**Figure 5. Results of Experiment 4. (A) Mean Accuracy as a function of target type, target configuration (interacting/non-interacting), target eccentricity (center/periphery) and set size.** Bar plots represent the mean group accuracy with interacting dyads (light brown) and non-interacting dyads (dark blue); error bars are standard errors from the mean. \* denote significant effects ( $p < 0.05$ ). ~ denote trends ( $p < .10$ ). **(B) Individual mean accuracy differences between interacting and non-interacting dyads.** Positive values indicate higher accuracy with interacting dyads, negative values indicate higher accuracy with non-interacting dyads, dots indicate individual participants.

**Summary.** Experiment 4 showed different patterns of performance in visual search through body dyads and object pairs. Participants were generally more accurate and responded faster with bodies than objects and in smaller than larger sets. Set size had a stronger influence on search time for bodies than objects but did not interact with the target configuration. More importantly, consistent with the previous experiments, the search advantage (in accuracy rates) for interacting dyads was only found when targets were in central locations. With targets at peripheral locations, performance was qualitatively better in the search for non-interacting (vs. interacting) targets. A similar effect was found with object pairs regardless of the target

eccentricity: whether at the center or at the periphery, non-interacting object pairs tended to be detected more frequently than interacting pairs. Thus, the multifaceted effect of social interaction does not extend to non-social relations. Object sets in meaningful (e.g., functionally relevant) spatial relations are generally grouped and thus easier to reject, but they do not possess the same ability to capture attention when shown at the center.

## General discussion

The current study addressed whether and how spatially defined higher-level relations (e.g., a social interaction implied by *facingness*) affect visual search. In particular, we investigated *a*) how people searched for interacting (face-to-face) bodies in a crowd of non-interacting (back-to-back) dyads, and *vice versa*; *b*) how the spatial location of the target in the visual field (*i.e.*, its eccentricity) affected search; and *c*) whether the pattern of search performance with body dyads generalized to other pairs instantiating another type of high-level relation (*i.e.*, objects in functional configurations for common usage).

Results reveal a complex pattern of search asymmetries. Interacting dyads among non-interacting dyads were detected better than non-interacting dyads among interacting dyads. This advantage for interacting dyads was only found when the target appeared in a central location, that is, in a location that was in the participant's focus of attention. When the target appeared at more peripheral locations, non-interacting dyads were detected better than interacting dyads. Thus, peripheral target location did not just abolish the asymmetry in favor of interacting dyads but reversed the effect. The latter effect generalized to the search for object pairs, independently of target location: whether presented in central or peripheral locations, non-interacting objects were detected better than interacting objects.

The advantage for interacting body dyads in central vision replicates the visual search asymmetry reported in Papeo *et al.* (2019), who showed that, with fast stimulus presentation, interacting targets among non-interacting distractors were detected better than non-interacting targets among interacting distractors. Adding to those findings, the current results specify the boundaries of the asymmetry, showing that attentional capture by interacting dyads only happens with target presentation at the attended central location.

This result shows that *facingness*, a reliable cue of social interaction, makes two people particularly salient at the attended spatial location (Papeo *et al.*, 2017; Papeo *et al.*, 2019; see also Vestner *et al.*, 2019; Skripkauskaitė *et al.*, 2022). This effect may be promoted by the perceptual enhancement of interacting dyads suggested by neuroimaging findings. In effect, fMRI research using classification analyses has shown that facing bodies have a stronger representation in visual cortex, relative to the same bodies presented in non-facing configurations (Abassi & Papeo, 2020; Bellot *et al.*, 2021). This neural effect could be linked to behavioral effects showing that body postures and movements seen in an interacting context are discriminated better (Bellot *et al.*, 2021; Neri *et al.*, 2006) and remembered better (Ding *et al.*, 2017; Paparella & Papeo, 2022; Vestner *et al.*, 2019) than the same stimuli presented as unrelated or in isolation. In sum, interacting dyads capture attention because their relative positioning makes them more salient, and therefore more visible, relative to unrelated bodies.

*Why isn't that the case for facing dyads at peripheral locations?* Interacting and non-interacting bodies are more difficult to discriminate at the periphery of the search arrays due to the effects of eccentricity (*i.e.*, decreased spatial resolution with increasing degrees of eccentricity; Carrasco & Frieder, 1997; De Valois & De Valois, 1988) and crowding (Bouma, 1970). Visual

crowding is the phenomenon in which an object in peripheral vision can be recognized in isolation but not when flanked by other objects, especially when target and flankers are visually similar, as it is the case for our body dyads (Whitney & Levi, 2011). Thus, when the target is not in central vision, interacting and non-interacting dyads cannot be readily discriminated, and the participant needs to start a search through the array. As our results show, this search is easier when the majority of distractors are interacting dyads and more difficult when they are non-interacting dyads. This effect is predicted by possible differences in the way interacting and non-interacting distractors are visually represented. In particular, the facilitatory effect of related distractors in visual search adds up to other behavioral and neural effects suggesting that the spatial positioning of objects according to meaningful relations (e.g., real-world regularities) triggers grouping of multiple objects into a perceptual unit (Kaiser et al., 2019). Grouping in visual search would effectively reduce the number of distractors to check and reject, thus increasing search efficiency (Kaiser et al., 2014). In parallel, and in line with this, growing literature shows that *facingness*, as a cue of interaction, triggers grouping of two people into a unitary perceptual structure (Abassi & Papeo, 2022; Adibpour et al., 2021; Papeo & Abassi, 2019; Vestner et al., 2019).

Overall, our results show that spatial relations that imply interaction between people or objects affect visual search. Visual search benefits from visual relational information by treating related items as single attentional units, thus reducing the number of items to process in an array. This mechanism is common to objects and people, or social and non-social relations. At the same time, interacting dyads capture attention when close to the initial attentional focus, while objects do not. This suggests that there are shared but also different mechanisms for processing of social and non-social visual relational information. Differences can reflect visual and semantic dissimilarities between object categories. For example, here, bodies were all visually very similar (in fact, the very same body was shown in different postures across all dyads); whereas objects belonged to various, visually different, categories. Moreover, while relations between bodies unfolded along the horizontal axis, object pairs were organized vertically. Body dyads and object pairs might also entail different levels of representation. That is, all facing dyads might trigger a general representation of interaction, while each interacting object pair might specify a relation, different from the others (lamp above table could be “dining area”, monitor above keyboard could be “computer” and so on), which would make object pairs easier to separate than body dyads. Difference between body dyads and object pairs may also reflect the use that we, as humans, make of relational information, and/or its relevance in the ecological setting. Detecting and recognizing social interactions in the wild is crucial for survival, and as a primary source of information for social learning. Stimuli that are reliably associated with social interaction (e.g., face-to-face body positioning) might have been prioritized in attention/perception throughout ontogeny and phylogeny (see New et al., 2007, for a perspective on evolutionary constraints on visual attention). In contrast, spotting a group of regularly positioned objects is not typically associated with such behavioral rewards. In navigating the environment, the visual system could use a strategy, where the regular, meaningful, and familiar configurations of objects are efficiently discarded to focus on the important –socially and biologically relevant– information, such as conspecifics, other animals, and their interactions.

In this spirit, the current results support the idea of a special status for interacting bodies in visual attention (see Papeo, 2020). However, it remains possible that other categories of stimuli or relations benefit from a similar advantage. The work of Vestner et al. (2020; 2021; 2022) suggests that an attentional capture similar to the one observed for interacting bodies could be observed for pairs of facing objects whose orientation (leftward or rightward) provides a

strong directional cue (e.g., cars and cameras). A direct comparison with our results however is limited by significant differences in the paradigms. Vestner *et al.* presented arrays of four stimuli around the central fixation (alike our “central” condition) for unlimited time, and did not test for visual search asymmetries: facing and non-facing pairs did not swap role in two conditions, as in classic search asymmetry paradigms (condition A: facing is the target, non-facing is the distractor; condition B: non-facing is the target, facing is the distractor), but were only used as targets among other distractors. Future studies should further extend the present comparison of socially interacting bodies and functionally interacting objects to other classes of stimuli and relations to clarify the types of stimuli/relations that are processed analogously to visual illustrations of social interaction (i.e., facing dyads).

To summarize, here we elicited two apparently conflicting effects reported in the literature in a single design, showing that interacting dyads capture attention more strongly than non-interacting dyads, but are also easier to reject as distractors. The new results reported here show how the same stimuli (interacting dyads) can give rise to different effects in visual search (attentional capture or more efficient search) based on target eccentricity (central/peripheral), determined by how much search thought the distractors is required. Meaningful relations between individual items structure the array in a way that reduces the number of units to process, facilitating the search for a target. This mechanism applies to social dyads as well as to non-social object pairs. The presentation of the target at the attended central location breaks this effect, but only when the target is a social dyad and only when the two members of the dyad are face-to-face, as if interacting. More generally, we showed how new units of perception emerge from higher-level relations such as social interaction or functional association, which are not explained by low-level visual properties of the stimuli (e.g., proximity, similarity, contours, continuity), and yet bind visual entities together, impacting the way in which individuals parse the visual world.

## Acknowledgements

This work was funded by a European Research Council Starting Grant awarded to L.P. (Project: THEMPO, Grant Agreement 758473), by a European Research Council Starting Grant awarded to D.K. (Project: PEP, Grant Agreement 101076057), by a Deutsche Forschungsgemeinschaft (DFG) grant awarded to D.K. (SFB/TRR135 – INST162/567-1), and “The Adaptive Mind”, funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. The authors are grateful to Mathilde Kremer for help in data collection.

## References

- Abassi, E., & Papeo, L. (2020). The Representation of Two-Body Shapes in the Human Visual Cortex. *Journal of Neuroscience*, 40(4), 852-863. <https://doi.org/10.1523/JNEUROSCI.1378-19.2019>
- Abassi, E., & Papeo, L. (2022). Behavioral and neural markers of visual configural processing in social scene perception. *NeuroImage*, 260, 119506. <https://doi.org/10.1016/j.neuroimage.2022.119506>
- Adibpour, P., Hochmann, J.-R., & Papeo, L. (2021). Spatial Relations Trigger Visual Binding of People. *Journal of Cognitive Neuroscience*, 33(7), 1343-1353. [https://doi.org/10.1162/jocn\\_a\\_01724](https://doi.org/10.1162/jocn_a_01724)
- Anderson, S. F., & Kelley, K. (2020). Bias-and Uncertainty-Corrected Sample Size. R package version 1.2. 1. <https://CRAN.R-project.org/package=BUCSS>
- Bellot, E., Abassi, E., & Papeo, L. (2021). Moving Toward versus Away from Another : How Body Motion Direction Changes the Representation of Bodies and Actions in the Visual Cortex. *Cerebral Cortex*, 31(5), 2670-2685. <https://doi.org/10.1093/cercor/bhaa382>
- Bouma, H. (1970). Interaction Effects in Parafoveal Letter Recognition. *Nature*, 226(5241), 177-178. <https://doi.org/10.1038/226177a0>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436. <https://doi.org/10.1163/156856897X00357>
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, 57(8), 1241-1261. <https://doi.org/10.3758/BF03208380>
- Carrasco, M., & Frieder, K. S. (1997). Cortical Magnification Neutralizes the Eccentricity Effect in Visual Search. *Vision Research*, 37(1), 63-82. [https://doi.org/10.1016/S0042-6989\(96\)00102-2](https://doi.org/10.1016/S0042-6989(96)00102-2)
- De Valois, R. L., & De Valois, K. K. (1980). Spatial vision. *Annual Review of Psychology*, 31, 309-341. <https://doi.org/10.1146/annurev.ps.31.020180.001521>
- Ding, X., Gao, Z., & Shen, M. (2017). Two Equals One: Two Human Actions During Social Interaction Are Grouped as One Unit in Working Memory. *Psychological Science*, 28(9), 1311–1320. <https://doi.org/10.1177/0956797617707318>
- Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences*, 111(30), 11217-11222. <https://doi.org/10.1073/pnas.1400559111>
- Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin & Review*, 22(6), 1784-1790. <https://doi.org/10.3758/s13423-015-0833-4>

- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object Vision in a Structured World. *Trends in Cognitive Sciences*, 23(8), 672-685. <https://doi.org/10.1016/j.tics.2019.04.013>
- Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments. R package version, 4(2). <https://CRAN.R-project.org/package=ez>
- Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, 16(1), 1–10. <https://doi.org/10.1080/13506280701381691>
- Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, 9(9), 1186-1192. <https://doi.org/10.1038/nn1759>
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, 104(42), 16598-16603. <https://doi.org/10.1073/pnas.0703913104>
- Paparella, I., & Papeo, L. (2022). Chunking by social relationship in working memory. *Visual Cognition*, 30(5), 354-370. <https://doi.org/10.1080/13506285.2022.2064950>
- Papeo, L. (2020). Twos in human visual perception. *Cortex*, 132, 473-478. <https://doi.org/10.1016/j.cortex.2020.06.005>
- Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human–human interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 45(7), 877-888. <https://doi.org/10.1037/xhp0000646>
- Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual Search for People Among People. *Psychological Science*, 0956797619867295. <https://doi.org/10.1177/0956797619867295>
- Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychological Science*, 28(3), 369-379. <https://doi.org/10.1177/0956797616685769>
- Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching : How the human brain makes sense of other people’s encounters. *Annals of the New York Academy of Sciences*, 1396(1), 166-182. <https://doi.org/10.1111/nyas.13331>
- Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). Chapter 13 - More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. In B. L. Parkin (Éd.), *Progress in Brain Research* (Vol. 253, p. 243-262). Elsevier. <https://doi.org/10.1016/bs.pbr.2020.06.005>
- Skripkauskaitė, S., Mihai, I., & Koldewyn, K. (2022). Attentional bias towards social interactions during viewing of naturalistic scenes. *Quarterly Journal of Experimental Psychology*, 17470218221140879.

- Stein, T., Kaiser, D., & Peelen, M. V. (2015). Interobject grouping facilitates visual awareness. *Journal of Vision*, 15, 10.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 459-478. <https://doi.org/10.1037/0096-1523.16.3.459>
- Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3), 285-310. <https://doi.org/10.1037/0096-3445.114.3.285>
- Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, 200, 104270. <https://doi.org/10.1016/j.cognition.2020.104270>
- Vestner, T., Gray, K. L. H., & Cook, R. (2021). Visual search for facing and non-facing people: The effect of actor inversion. *Cognition*, 208, 104550. <https://doi.org/10.1016/j.cognition.2020.104550>
- Vestner, T., Over, H., Gray, K. L. H., & Cook, R. (2022). Objects that direct visuospatial attention produce the search advantage for facing dyads. *Journal of Experimental Psychology: General*, 151(1), 161-171. <https://doi.org/10.1037/xge0001067>
- Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S.-A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000545>
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15(4), 160-168. <https://doi.org/10.1016/j.tics.2011.02.005>
- Wickham, H. (2016). Data analysis. In *ggplot2* (pp. 189-201). Springer, Cham.
- Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Attention, Perception, & Psychophysics*, 63(3), 381-389.
- Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In *Integrated Models of Cognitive Systems* (p. 99-119). Oxford University Press, USA.
- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202-238. <https://doi.org/10.3758/BF03200774>