

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, 67 Bd. Pinel, 69675, Bron, France.

Running title: Representation of dyadic relations in infants

* **Correspondence to:** N.G., L.P., CNRS, Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, 67 Boulevard Pinel, 69675, Bron, France; Phone: +33 043791 1266; E-mail: nicolas.goupil@isc.cnrs.fr, liuba.papeo@isc.cnrs.fr

Word count: 8061

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

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). 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; 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.

INSERT FIGURE 1 ABOUT HERE

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 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, 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.

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 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 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° , to the Inverted group.

INSERT FIGURE 2 ABOUT HERE

Procedure

Infants sat on their parent's lap at a distance of ~ 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.

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 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.

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 2.

Results

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.

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 ± 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 ± 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 ± 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 ± 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 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 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.

INSERT FIGURE 3 ABOUT HERE

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 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.

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.

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 $(LT_{\text{single}} - LT_{\text{dyad}})/(LT_{\text{single}} + LT_{\text{dyad}})$, 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.

Results

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 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 ($M = -.12$, $SD = .19$ corresponding to a difference of 264 ± 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 ± 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 ± 379 ms; $W = 16$, $P = .001$; $r_{\text{Rosenthal}} = .73$) and the facing condition ($M = -.10$, $SD = .20$ corresponding to a difference of 208 ± 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.

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 1). 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.

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.

Results

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)¹. 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).

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 ± 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 ± 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 ± 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 ± 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 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.

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 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.

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.

Results

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

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 ± 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 ± 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 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*.

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 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 (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 (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 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.

Acknowledgements

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.

References

- Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience*, 40(4), 852-863.
- Aslin, R. N. (2007). What's in a look?. *Developmental science*, 10(1), 48-53.
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, 20(3), 191-208.
- 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*, 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.
- 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>
- Birmingham, E., & Kingstone, A. (2009). Human social attention. *Annals of the New York Academy of Sciences*, 1156, 118-140.
- 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.
- 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.
- 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.
- 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.
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic bulletin & review*, 5(3), 490-495.
- 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.
- 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.
- Graziano, M. S., & Kastner, S. (2011). Human consciousness and its relationship to social neuroscience: a novel hypothesis. *Cognitive neuroscience*, 2(2), 98-113.

- 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.
- Hochmann, J.-R., & Papeo, L. (2014). The invariance problem in infancy: A pupillometry study. *Psychological science*, 25(11), 2038-2046.
- 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.
- 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.
- 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.
- 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).
- 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.
- 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.
- Leslie, A. M. (1982). The perception of causality in infants. *Perception*, 11(2), 173-186.
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3), 265-288.
- Lewkowicz, D. J., & Ghazanfar, A. A. (2012). The development of the uncanny valley in infants. *Developmental psychobiology*, 54(2), 124-132.
- Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences*, 115(38), E8835-E8843.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, 164(1), 177-190.
- 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.
- 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(2), 164.
- Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive psychology*, 61(2), 63-86.
- 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.
- 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.
- Papeo, L. (2020). Twos in human visual perception. *Cortex*. OSFPreprints [10.31219/osf.io/yuxza](https://doi.org/10.31219/osf.io/yuxza)
- 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.
- Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, 28(3), 369-379.
- 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 (2005).
- 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.
- Powell, L. J., & Spelke, E. S. (2018). Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition*, 170, 31-48.
- Powell, L. J., & Spelke, E. S. (2018). Third party preferences for imitators in preverbal infants. *Open Mind*, 2, 61-71.
- Ro, T. Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition*, 15(3), 322-348.
- 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.
- Scholl, B. J. , & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, 4(8), 299-309.
- 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.
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness?. *Psychological science*, 23(2), 196-204.
- Spelke, E. S. (1990). Principles of object perception. *Cognitive science*, 14(1), 29-56.
- Stahl, A. E., & Feigenson, L. (2014). Social knowledge facilitates chunking in infancy. *Child development*, 85(4), 1477-1490.

- Sun, Y., Stein, T., Liu, W., Ding, X., & Nie, S. P. (2017). Biphasic attentional orienting triggered by invisible social signals. *Cognition*, 168, 129-139.
- 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 (2015).
- 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.
- 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.
- Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, 200, 104270.
- Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, 198, 296-302.
- Walbrin, J., Mihai, I., Landsiedel, J., & Koldewyn, K. (2020). Developmental changes in visual responses to social interactions. *Developmental Cognitive Neuroscience*, 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.
- 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.
- 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.

Footnotes

1- 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 ($P_s > .77$).

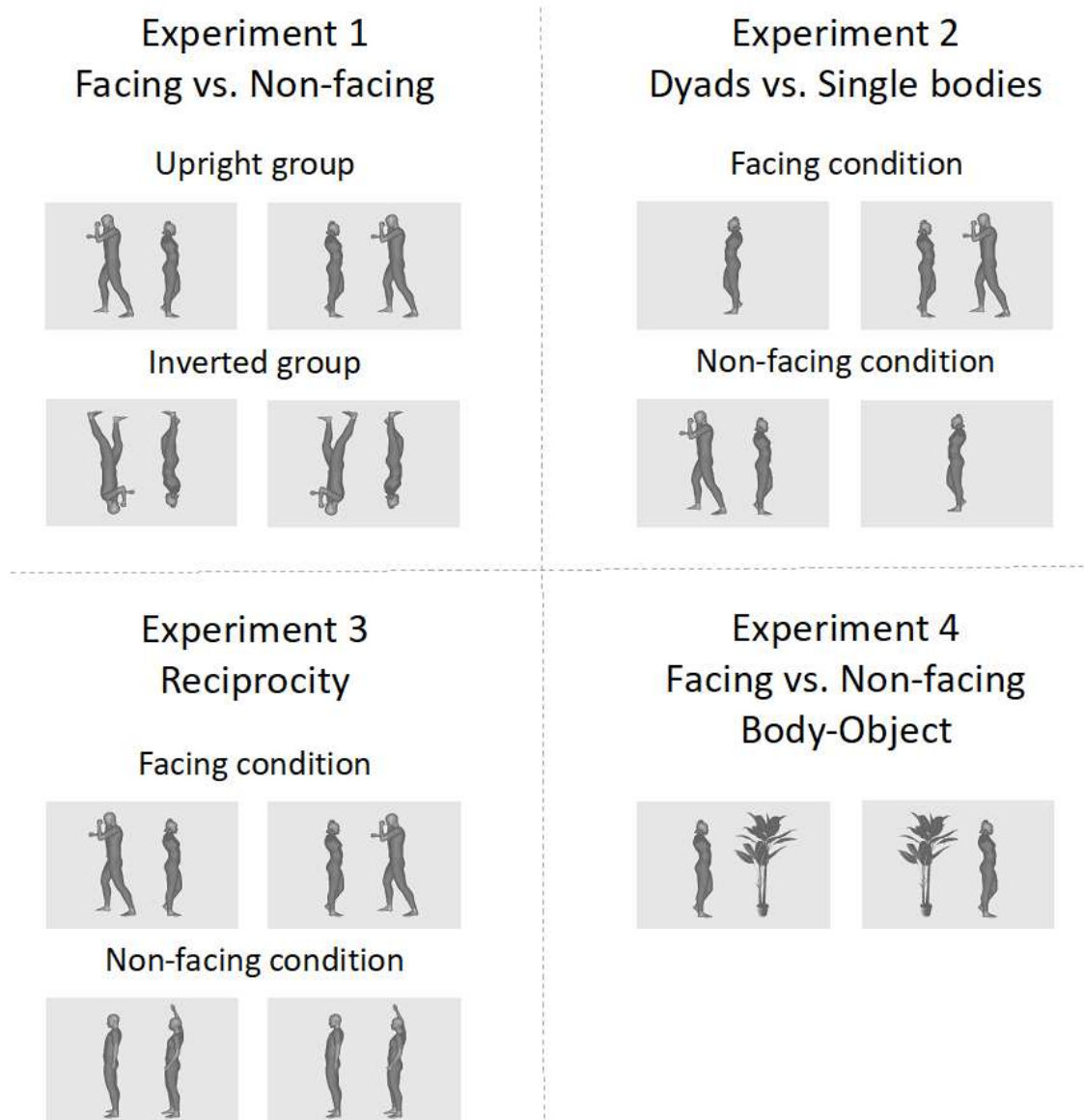


Figure 1 – Examples of the stimuli used in Experiments 1-4.

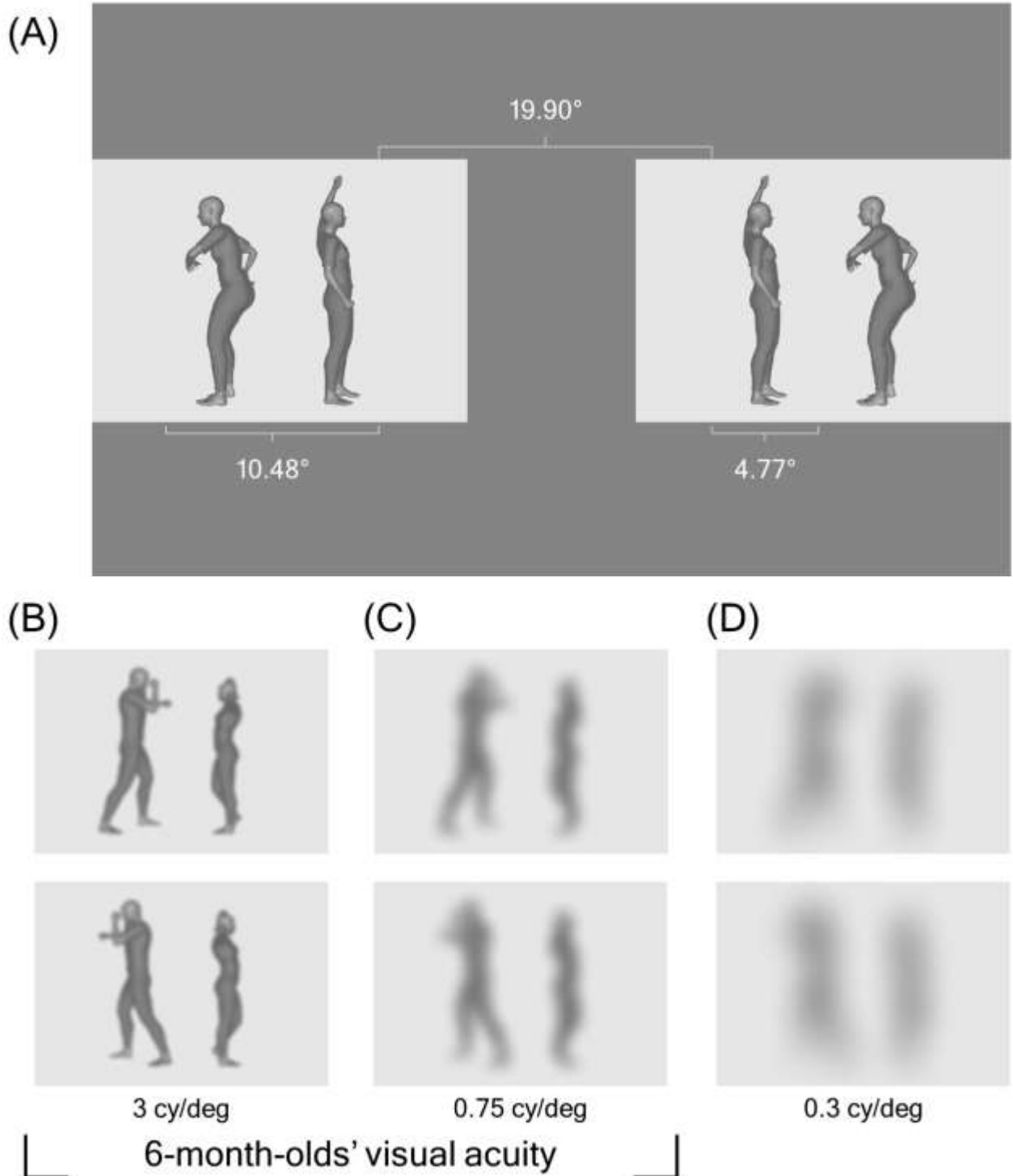


Figure 2 – Illustration of the stimulation with (A) display layout and (B-D) stimulus visibility at 6 months. (A) Image outlay and dimensions. (B-D) The same exemplar stimulus was filtered by 3 low-pass Gaussian filter. (B) Stimulus low-pass filtered at the highest spatial frequencies of the 6-months-old 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).

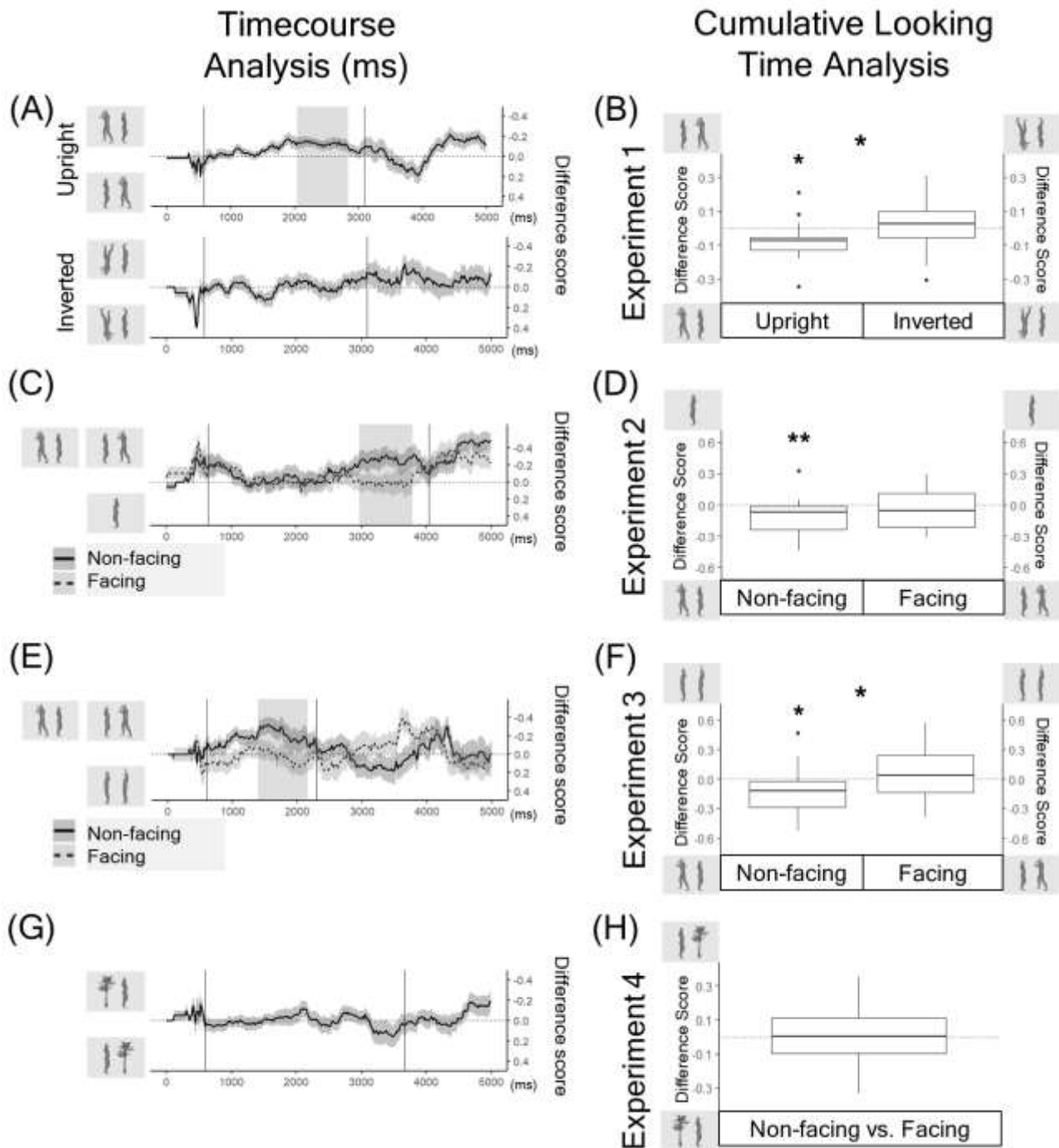


Figure 3 – Results of the time course analyses (A,C,E,G) and of the cumulative looking time analyses inside the Informative Time Window (B,D,F,H) for Experiments 1-4. (A,C,E,G) Each curve represents the time course of the difference scores, measuring the tendency to look towards one or the other type 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). * $P < .05$; ** $P < .01$.

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)

Table 2 – Summary of the 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.

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

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 ($p = .005$) ^a
Inverted		n.s.	n.s.
Upright vs. Inverted		Upright > Inverted (2176-2652, $p = .09$)	Upright > Inverted ($p = .02$) ^b
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 ($p = .003$) ^a
Facing vs. Non-facing		Non-facing > Facing (2907-3791, $p = .03$)	Non-facing > Facing ($p = .09$) ^a
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 ($p = .009$) ^a
Facing vs. Non-facing		Non-facing > Facing (1394-2159, $p = .01$)	Non-facing > Facing ($p = .005$) ^a
Experiment 4 (<i>Facing vs. Non-facing</i>)			
body-plant dyads	595-3672	n.s.	n.s.

Note: ^a Within group comparisons were assessed through Wilcoxon signed-rank tests. ^b Between group comparison was assessed through a Wilcoxon rank-sum test. n.s. = $P > .10$.

Table S1. Summary of sample sizes in the literature.

Study	Sample size ^a	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
Muentner & 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: ^a a maximal and minimal sample sizes of groups in repeated-measures designs, or groups assigned to only one condition in between-subjects designs; ^b range of samples sizes across different experiments of a study; ^c sample sizes are only approximate given the information in the article.

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

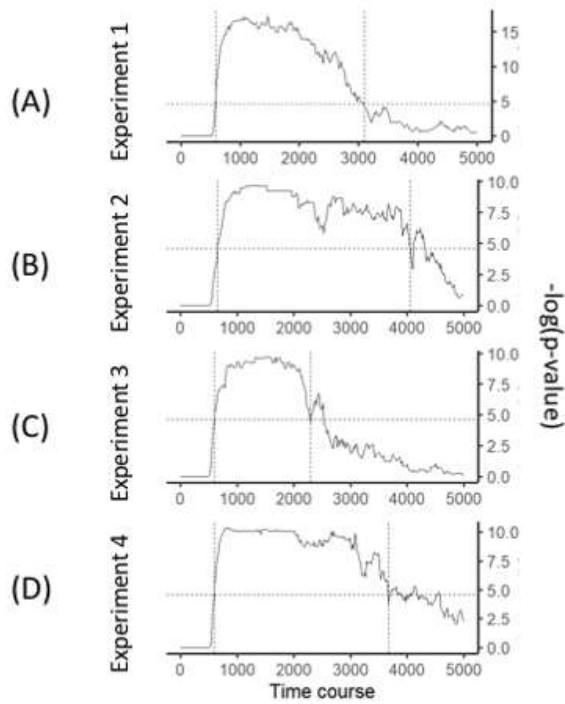


Figure S1 – Identification of the Informative Time Window (ITW) for Experiments 1-4. Each curve represents the time course of $-\log(p\text{-value})$, where p-value is the result of a Wilcoxon signed rank test comparing the proportion of trials looking at one of the picture to .50. The horizontal dotted 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.