



Spatial Relations Trigger Visual Binding of People

Parvaneh Adibpour^{1,2}, Jean-Rémy Hochmann^{1,2}, and Liuba Papeo^{1,2}

Abstract

■ To navigate the social world, humans must represent social entities and the relationships between those entities, starting with spatial relationships. Recent research suggests that two bodies are processed with particularly high efficiency in visual perception, when they are in a spatial positioning that cues interaction, that is, close and face-to-face. Socially relevant spatial relations such as facingness may facilitate visual perception by triggering grouping of bodies into a new integrated percept, which would make the stimuli more visible and easier to process. We used EEG and a frequency-tagging paradigm to measure a neural correlate of grouping (or visual binding), while female and male participants saw images of two bodies face-to-face or back-to-back. The two bodies in a dyad flickered at frequency F1 and F2, respectively, and appeared together at a third frequency Fd (dyad frequency). This stimulation should

elicit a periodic neural response for each body at F1 and F2, and a third response at Fd, which would be larger for face-to-face (vs. back-to-back) bodies, if those stimuli yield additional integrative processing. Results showed that responses at F1 and F2 were higher for upright than for inverted bodies, demonstrating that our paradigm could capture neural activity associated with viewing bodies. Crucially, the response to dyads at Fd was larger for face-to-face (vs. back-to-back) dyads, suggesting integration mediated by grouping. We propose that spatial relations that recur in social interaction (i.e., facingness) promote binding of multiple bodies into a new representation. This mechanism can explain how the visual system contributes to integrating and transforming the representation of disconnected body shapes into structured representations of social events. ■

INTRODUCTION

The visual world is an array of objects, but our understanding of the world goes beyond single object representation, involving, among many others, the processing of how individual entities relate to one another. The analysis of relations is essential to represent a scene or an event.

In social scenes (i.e., scenes involving multiple people), certain relational features, such as spatial proximity and positioning angle between bodies, are reliably correlated with the occurrence of social interaction (Zhou, Han, Liang, Hu, & Kuai, 2019). These features are immediately available to our visual system, affecting the way in which individuals represent objects (Papeo, Goupil, & Soto-Faraco, 2019; Hafri, Trueswell, & Strickland, 2018; Papeo, Stein, & Soto-Faraco, 2017; Glanemann, Zwitserlood, Bölte, & Dobel, 2016). For instance, it has been shown that, presented at perceptual threshold, two bodies are recognized more easily when they are shown face-to-face than when they are back-to-back (Bellot, Abassi, & Papeo, 2021; Papeo et al., 2017). Face-to-face bodies are also attended to with higher priority, relative to other multiple-body configurations: In visual search through a crowd, face-to-face bodies recruit attention more strongly, and are searched for more efficiently, than back-to-back bodies (Vestner, Gray, & Cook, 2020; Papeo et al., 2019; Vestner, Tipper, Hartley, Over, & Rueschemeyer, 2019).

Efficient processing of physically disconnected object shapes in certain spatial configurations could be achieved through perceptual grouping, the binding of parts (here, bodies) into a new unitary representation (Coren & Girgus, 1980). Grouping based on spatial relations can reflect the gestalt laws of low-level vision (Wagemans et al., 2012; Koffka, 1935), but can also be driven by statistically frequent or semantically relevant relations (Kaiser, Quek, Cichy, & Peelen, 2019; Kaiser, Stein, & Peelen, 2014) or, as suggested for face-to-face bodies, socially relevant spatial relations (Papeo, 2020; Quadflieg, Gentile, & Rossion, 2015). In summary, grouping would give rise to a new composite representation, when the spatial arrangement of objects matches an expected, regular, familiar, or meaningful configuration. The processing of multiple objects as a coherent, unitary structure increases efficiency, possibly by reducing the stimulus complexity and/or the competition for selection between objects (Kaiser et al., 2019; McMains & Kastner, 2010; Reddy, Kanwisher, & VanRullen, 2009).

At the neural level, recent fMRI studies have reported that visual areas underlying body perception in the occipito-temporal cortex participate in the representation of multiple-body stimuli, showing increased response and more accurate representation of dyads of face-to-face bodies (vs. nonfacing bodies or single bodies; Bellot et al., 2021; Abassi & Papeo, 2020; Walbrin & Koldewyn, 2019). Activity in these areas is further modulated by the coherence of a multiple-person scene, differentiating between

¹Institut des Sciences Cognitives—Marc Jeannerod, Bron, France, ²Université Claude Bernard Lyon 1, Lyon, France

scenes depicting people that belong to the same (vs. a different) context (e.g., a club or a party; Quadflieg et al., 2015). Increased activity for socially relevant multiple-body configurations has been interpreted as evidence of grouping, following prior research on multiple-object perception. Indeed, research on multiple-object perception has consistently demonstrated an increase in the neural response to pairs of objects seen in familiar (i.e., regular or expected) configurations (e.g., a screen above a keyboard) as compared with the same objects in a spatial relation that does not promote the formation of composite representation (e.g., a screen below a keyboard; Kaiser et al., 2014; Kim & Biederman, 2011; Roberts & Humphreys, 2010; MacEvoy & Epstein, 2009). Increased activity for regular multiple-object configurations has been specifically associated with markers of integrative processing that could mediate grouping of multiple parts or multiple objects in a composite representation that is more than the sum of the parts (Kaiser & Peelen, 2018; Baldassano, Beck, & Fei-Fei, 2017; Kubilius, Baeck, Wagemans, & Op de Beeck, 2015; Baeck, Wagemans, & Op de Beeck, 2013).

Here, we sought to provide an objective measure of grouping of facing body dyads, using a frequency-tagging EEG paradigm to separate the response to the parts of a dyad (single bodies) from the response to the whole (the dyad). In frequency-tagging EEG, stimuli are presented periodically, at a regular frequency. The periodic stimulation entrains a periodic neural activity at the stimulation frequency, which is easily distinguishable in the frequency domain (Norcia, Appelbaum, Ales, Cottareau, & Rossion, 2015; Regan, 1966). By presenting different parts of a multipart stimulus at different stimulation frequencies, it is possible to dissociate the response to each single part from the response to the whole stimulus.

This paradigm has proven effective in capturing neural effects of binding visual features based on gestalt principles (Alp, Kogo, Van Belle, Wagemans, & Rossion, 2016; Aissani, Cottareau, Dumas, Paradis, & Lorenceau, 2011) and binding parts into an object representation (e.g., a face; Boremanse, Norcia, & Rossion, 2013). In particular, using frequency-tagging EEG, Boremanse et al. (2013) could distinguish the responses to the two halves of a face, each presented at a different regular frequency, from the response to the whole face, emerging when the brain integrates information from the two halves, at a third regular frequency. Critically, the response to whole-faces was larger when the halves gave rise to a face, relative to when their arrangement violated the canonical face configuration (i.e., the halves were next to each other but misaligned; Boremanse et al., 2013). Thus, the response to whole-faces was not elicited by the mere occurrence of more facial features, but by the integration of those features in a well-formed face.

Extending the rationale of Boremanse et al. (2013), in this study, we used a frequency-tagging paradigm to obtain objective responses to the parts and whole of a visual

scene, where the constituting parts were two single bodies and the whole was the dyad. Each of the two bodies flickered at a different frequency to elicit a frequency-tagged response at the corresponding frequency (response to parts: F1 and F2). The two bodies appeared together in full view at a third frequency, that we called Fd (i.e., dyad-related frequency). To isolate visual activity associated with body perception, in half of the trials, bodies were presented upside-down, a condition that significantly reduces activity in visual areas for body perception, while providing matched visual stimulation (Brandman & Yovel, 2010, 2016). In different trials, the two bodies faced toward or away from the center. Therefore, when they appeared together at Fd, they faced either toward each other (facing dyad) or away from each other (nonfacing dyad). We reasoned that, if the face-to-face positioning gives rise to a new integrated representation of the two bodies, the response at Fd should be enhanced for facing (vs. nonfacing) dyads. Enhanced response to facing dyads will provide initial evidence for multiple-body integration, triggered by specific-socially relevant-spatial relations. Grouping through integration may begin to account for the neural mechanism that transforms a group of bodies in the representation of a social event such as an interaction.

METHODS

Participants

Twenty healthy adults (13 women; mean age = 23.3, SD = 4.6) with normal or corrected-to-normal vision participated in the study. The sample size was decided based on several EEG studies with a similar paradigm (Kaiser, Häberle, & Cichy, 2020; Alp et al., 2016; Boremanse et al., 2013). The study was approved by the local ethics committee. All participants gave written informed consent and received monetary compensation for participation.

Stimuli

Grayscale images of an identical body in six different postures, seen in profile view, were created with Daz3D (Daz Productions) and the Image Processing Toolbox of MATLAB (The MathWorks). Four facing dyads (Figure 1A) were created by placing two bodies face-to-face. Four nonfacing dyads were created by swapping the position of the two bodies in the facing dyads (Figure 1B). The distance between the center of each body and the center of the screen was fixed (77 pixels); furthermore, the distance between two bodies (i.e., between their inner edges on the horizontal axis) was matched between a facing dyad and its corresponding nonfacing dyad (mean = 38.2 ± 5.4 pixels). Thus, the two bodies were at comparable distance from central fixation across facing and nonfacing dyads.

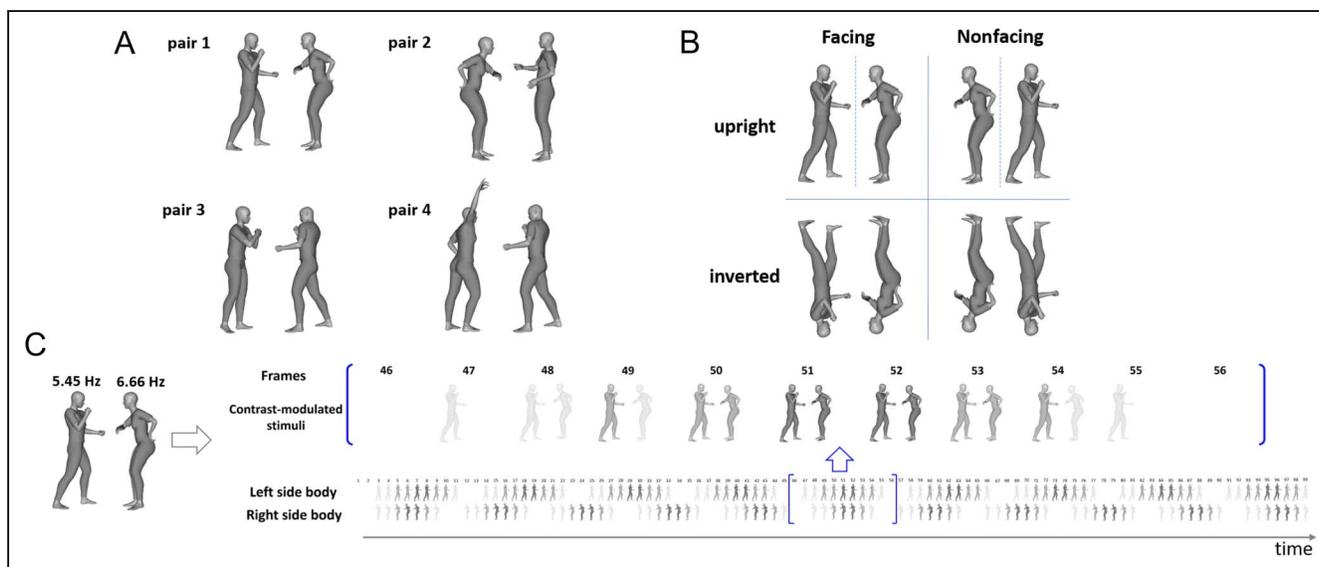


Figure 1. Illustration of stimuli, design, and procedures for stimulus delivery. (A) The four body dyads used in the study. (B) Four variations of a dyad presented in a block. (C) Illustration of the flickering presentation of the stimuli: The two bodies flickered at the frequencies F_1 and F_2 , respectively. The contrast-modulated appearance of body images is shown over a cycle of 99 frames, corresponding to frequency F_d , at which both bodies appeared in full contrast. The blue brackets represent a zoomed-in illustration of the frames where the simultaneous appearance of the two bodies at their full contrast occurs (frames 51 and 52, as marked with the blue arrow).

For every (facing and nonfacing) dyad, we created an inverted version by flipping the image by 180° . This yielded four variations of the same stimulus (i.e., upright facing, upright nonfacing, inverted facing, and inverted nonfacing), corresponding to the four experimental conditions of the study (Figure 1B).

EEG Recordings

Brain activity was recorded using a 128-electrode EEG net (EGI), with a reference on the vertex. Recordings were continuously digitized at a sampling rate of 1000 Hz (net amp 400 system EGI). Participants were seated on a chair, 1 m away from a computer screen where the stimuli were presented. Single bodies on the screen had a mean visual angle of approximately 4.6° in height and 2.1° in width, and their inner edge were placed at 0.3° from the center. Dyads subtended a mean visual angle of 4.6° in height and 4.2° in width.

Paradigm

The experiment consisted of four blocks. Each block included four trials of 70.91 sec (~ 1.18 min) each, in which the same dyad was presented in one of the four experimental conditions (upright facing, upright nonfacing, inverted facing, and inverted nonfacing). Trials were separated by a blank screen. After 20 sec of blank screen, participants were free to start a new trial when they felt ready. Each block lasted about 6–8 min depending on the duration of the blank intertrial intervals. The order of blocks and trials was randomized across participants.

The experiment included a total of 16 trials (4 blocks, each including 4 trials). The total duration did not exceed 30 min.

Each trial consisted of fast periodic visual stimulation, where the two bodies flickered at frequency F_1 of 6.6667 Hz (~ 6.66 Hz) and frequency F_2 of 5.4545 Hz (~ 5.45 Hz), respectively (Figure 1C). Those frequencies were chosen so to prevent the overlap with the alpha band frequency (8–12 Hz) and, thus, maximize the signal-to-noise ratio (SNR) of the evoked responses (Boremanse et al., 2013; Regan, 1989). That choice also took into account the refresh rate of the display (60 Hz), to allow an integer number of frames for each cycle of stimulus presentation ($60 \text{ Hz} / 6.6667 \text{ Hz} \approx 9$ frames, and $60 \text{ Hz} / 5.4545 \text{ Hz} \approx 11$ frames), preventing dropped frames or inaccuracies in the frequency of stimulus presentation. The contrast in the stimuli was sinusoidal-modulated (Figure 1C). Stimuli were presented in cycles of 99 frames ($9 \times 11 = 99$); in most frames, the two bodies were visible at a different contrast; they appeared together in full maximal contrast once during a cycle (at Frames 51–52; see Figure 1C), corresponding to frequency F_d (dyad-related frequency) of 0.6061 Hz ($60 \text{ Hz} / 99 \approx 0.6061 \text{ Hz}$). Two input frequencies, F_1 and F_2 , can give rise to nonlinear interactions at frequencies $nF_1 \pm mF_2$ (with n and m being integers; Zemon & Ratliff, 1984). Therefore, besides the response at F_d , the response at frequencies $nF_1 \pm mF_2$ can reflect the interaction between processing Stimulus 1 at F_1 and Stimulus 2 at F_2 . In our study, nonlinear interaction frequencies corresponded to $F_1 + F_2 = 12.12 \text{ Hz}$ and to $F_1 - F_2 = 1.21 \text{ Hz}$. We note that the latter response overlapped with the second harmonic

of the response at Fd, $F1 - F2 = (n2 - n1) Fd$, where $n1 (= 9)$ and $n2 (= 11)$ are the number of frames corresponding to F1 and F2, respectively.

The oscillation frequency of a body (F1 or F2) and side of the screen (left or right) was kept constant for facing and nonfacing dyads within a block and was counterbalanced across blocks. Throughout a trial, a black cross was present at the center of the screen roughly aligned with the shoulders of the bodies (whether upright or inverted). The cross color changed from black to red for 200 msec, 8 times at random intervals during the trial duration (i.e., 1.18 min). Participants were instructed to press the spacebar on a keyboard with their right hand when they detected the color change. This task was meant to maintain the participants' vigilance and fixation at the center of the screen. Stimulus presentation, communication of triggers with the recording EEG system, and response collection were controlled with Psychtoolbox (Brainard, 1997) through MATLAB.

Preprocessing

EEG recordings were band-pass filtered between 0.1 and 100 Hz using a zero-phase lag filter and were further processed using the EEGLAB (Delorme & Makeig, 2004) and Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011) toolboxes in MATLAB. Recordings were segmented for the duration of the trials, based on the trigger marking the trial onset. In each trial, the first 3.3 sec of the data (i.e., two cycles of stimulus presentation) were removed, taking into account the time needed for the entrainment to become effective on the brain activity. The analysis considered the following 66 sec of the data for each trial, corresponding to 360 cycles for the stimulation frequency of 6.66 Hz, and 440 cycles for the stimulation frequency of 5.45 Hz, and 40 cycles of the dyad-presentation frequency 0.61 Hz. The final 1.6 sec (i.e., one cycle of stimulus presentation) were not analyzed, to take into account the potential imprecision of boundary markers.

Frequency Domain Analysis

In the periodic activity evoked by the periodic presentation of the stimuli, responses to the two individual bodies were frequency-tagged at F1 and F2, whereas responses to dyads were frequency-tagged at Fd. Time-series from each trial were transformed into the frequency domain using discrete Fourier transform, with a high frequency resolution 0.0152 Hz ($1/66 \text{ sec} = 0.0152 \text{ Hz}$), which allows studying the response at the frequencies of interest with high precision. For each resulting frequency spectrum, the SNR was then computed as the amplitude at each frequency divided by the mean amplitude of the 80 neighboring frequencies (40 frequency bins on each side), after excluding the four immediately adjacent bins (Leleu et al., 2020; Boremanse et al., 2013). The SNR

was then averaged across trials of the same experimental condition, separately for each participant. Mean SNR values were analyzed to determine the response to individual bodies and dyads, at the corresponding tagged frequencies.

Statistical Analyses and Results

Grand Average Identification of Frequency-Tagged Responses

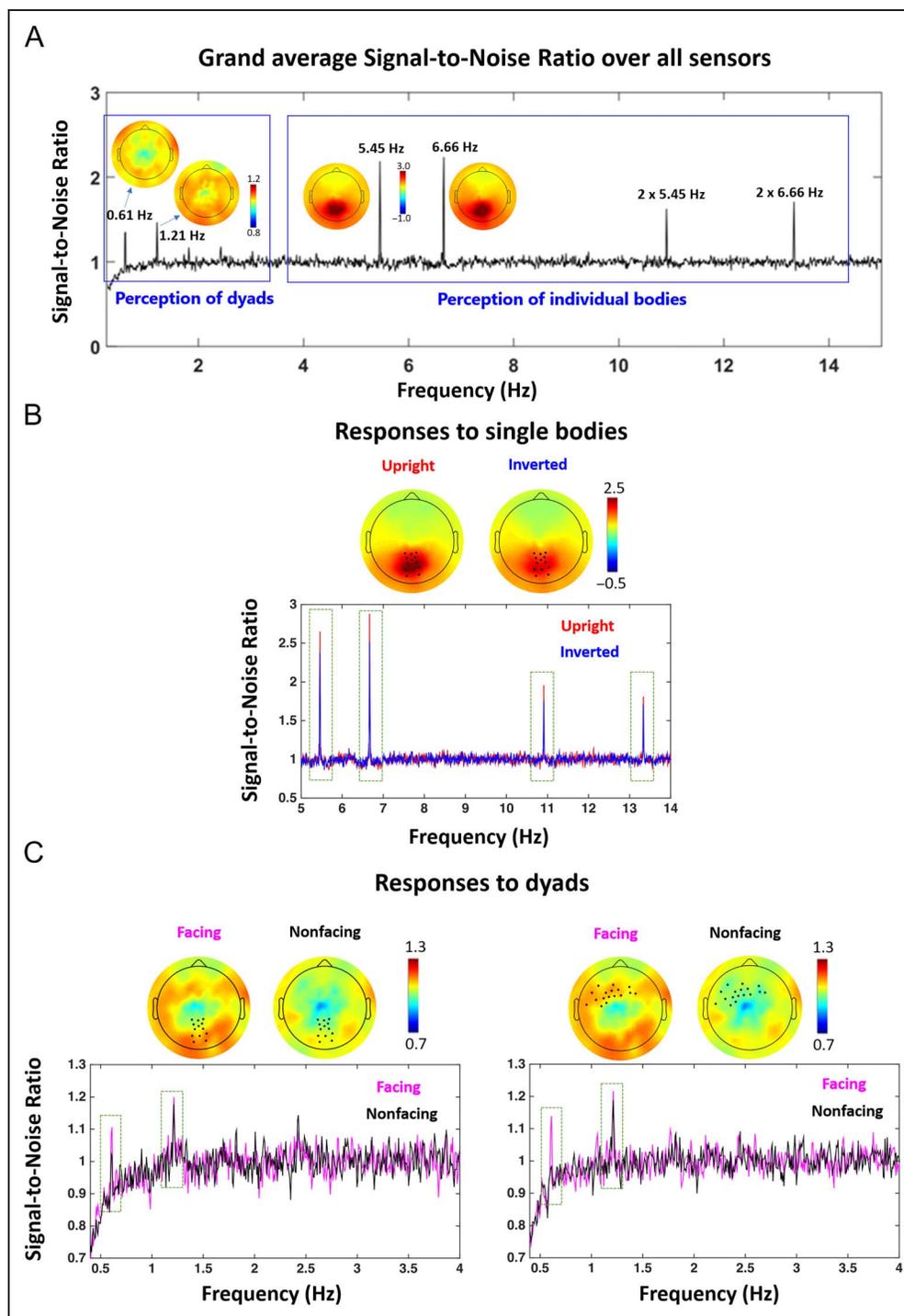
We first evaluated how many participants showed above-noise-level responses at the predefined frequencies, irrespective of experimental conditions and spatial distribution of the responses, to ascertain that our stimulation was effective in evoking frequency-tagged responses. To this aim, we averaged the SNR spectrum over all EEG sensors and trials for each participant (Figure 2A). Then, we compared the SNR at the frequencies F1 and F2 (6.66 Hz and 5.45 Hz), Fd (0.61 Hz) and their harmonics, and the second-order intermodulation responses at $nF1 \pm mF2$, against the noise level (80 neighboring frequency bins excluding the immediately adjacent four bins), with a one-tailed *t* test. This analysis showed significant responses at the main frequencies of interest F1, F2, and Fd, as well as at multiple harmonics of those frequencies ($2 \times F1/F2/Fd$, $3 \times F1/F2/Fd$, $4 \times F1/F2/Fd$). Note that the even harmonics of the response at Fd also corresponded to harmonics of the intermodulation response F1–F2 (e.g., $2Fd = F1 - F2$). Other second-order intermodulation responses (i.e., $F1 + F2$) were not significantly above noise level in this whole-sensor analysis.

For both single-body and dyad-related responses, we constrained the analyses to the first two harmonics, because these responses were the most robust in 18 out of 20 participants and allowed keeping the same number of harmonics for both responses (statistical analysis of all harmonic responses is reported in Table 1). Thus, to study the response to single bodies, we averaged the SNR at 6.66 Hz and $2 \times 6.66 \text{ Hz}$ (Body 1) and at 5.45 Hz and $2 \times 5.45 \text{ Hz}$ (Body 2); to study the response to dyads, we averaged the SNR at 0.61 Hz and 1.21 Hz ($2 \times 0.606 \text{ Hz}$).

Responses at F1, F2, and their harmonics were predominantly distributed over posterior electrodes with medial distribution. Responses at Fd and $2 \times Fd (= F1 - F2)$ were more broadly distributed over the scalp, encompassing posterior and anterior areas (Figure 2).

The intermodulation response at $F1 + F2$ was not significant in the first (whole-sensor) analysis. Because this response has often been described in similar frequency-tagging studies, we ran a secondary, less conservative analysis, restricting our search space to the 20 posterior sensors extending from O1/O2 to T5/T6. We tested the SNR value at $F1 + F2$ against the noise level for each sensor using a one-tailed *t* test. This analysis revealed a response at $F1 + F2$ significantly above the noise level,

Figure 2. Frequency-tagged responses and their modulation depending on the experimental conditions. (A) Grand average SNR over the group of participants, averaged over all EEG sensors. Peaks are visible at the stimulation frequencies corresponding to single bodies and dyads. Adjacent to the peak, the spatial distribution of responses is shown on a topographical map. (B) Response to upright and inverted single bodies in a posterior cluster (body-related cluster) as highlighted on the topographical maps. Average SNR value within this cluster at frequencies F1 and F2 are highlighted in red for upright bodies and in blue for inverted bodies. (C) Response to facing and nonfacing dyads in the posterior body-related cluster of sensors (left), and in the anterior cluster pointed out by the analysis across all EEG sensors (right). These clusters are highlighted on the topographical maps, and the average SNR values within each of them is plotted below the corresponding topographies, showing higher activations for facing dyads (in magenta) than for nonfacing dyads (in black).



on three adjacent electrodes, $t(19) = 3.17$; $p = .050$, correcting for 20 multiple comparisons.

Definition of Body-related Response

Previous fMRI studies have reported that the encoding of spatial relations between multiple bodies begins in visual areas for body perception (Abassi & Papeo, 2020; Walbrin & Koldewyn, 2019). To follow up on those studies, we singled out visual activity for viewing bodies, using the

contrast upright versus inverted bodies at F1, F2, $2 \times F1$ and $2 \times F2$. Indeed, inverted bodies typically reduce or change the response to bodies in occipital visual areas (Brandman & Yovel, 2010, 2016; Minnebusch, Suchan, & Daum, 2009; Stekelenburg & de Gelder, 2004).

The comparison between SNR values for upright versus inverted bodies was carried out for each sensor separately with a one-tail t test. Neighboring sensors showing an effect with a t value larger than a threshold corresponding to $p < .05$, $t(19) > 1.73$, were clustered and tested for

Table 1. Significance of the Responses at Stimulation Frequencies of Interest and Their Multiple Harmonics as Assessed on the Grand Average SNR of All EEG Sensors

	<i>F1</i>	$2 \times F1$	$3 \times F1$	$4 \times F1$	$5 \times F1$	$6 \times F1$	$7 \times F1$	$8 \times F1$	$9 \times F1$
<i>Frequency</i>	6.66 Hz	13.33 Hz	20.00 Hz	26.66 Hz	33.33 Hz	40.00 Hz	46.66 Hz	Above 50 Hz	Above 50 Hz
Participants (out of 20) with significant responses	20	20	19	20	13	18	14	–	–
Statistics	$t = 8.7,$ $p = .001$	$t = 8.1,$ $p = .001$	$t = 5.3,$ $p = .001$	$t = 7.3,$ $p = .001$	$t = 2.3,$ $p = .110$	$t = 5.6,$ $p = .001$	$t = 2.5,$ $p = .071$	–	–
	<i>F2</i>	$2 \times F2$	$3 \times F2$	$4 \times F2$	$5 \times F2$	$6 \times F2$	$7 \times F2$	$8 \times F2$	$9 \times F2$
<i>Frequency</i>	5.45 Hz	10.90 Hz	16.36 Hz	21.81 Hz	27.27 Hz	32.72 Hz	38.18 Hz	43.63 Hz	49.09 Hz
Participants (out of 20) with significant responses	18	19	20	18	19	17	15	15	13
Statistics	$t = 6.2,$ $p = .001$	$t = 6.9,$ $p = .001$	$t = 6.3,$ $p = .001$	$t = 4.0,$ $p = .003$	$t = 4.5,$ $p = .001$	$t = 4.5,$ $p = .001$	$t = 3.7,$ $p = .007$	$t = 2.9,$ $p = .004$	$t = 2.6,$ $p = .007$
	<i>Fd</i>	$2 \times Fd$	$3 \times Fd$	$4 \times Fd$	$5 \times Fd$	$6 \times Fd$	$7 \times Fd$	$8 \times Fd$	$9 \times Fd$
<i>Frequency</i>	0.606 Hz	1.21 Hz	1.81 Hz	2.42 Hz	3.03 Hz	3.63 Hz	4.24 Hz	4.84 Hz	Above 5 Hz
Participants (out of 20) with significant responses	20	19	14	15	16	14	13	13	–
Statistics	$t = 7.5,$ $p = .001$	$t = 6.4,$ $p = .001$	$t = 3.6,$ $p = .007$	$t = 3.2,$ $p = .018$	$t = 2.5,$ $p = .083$	$t = 2.0,$ $p = .229$	$t = 2.2,$ $p = .163$	$t = 1.2,$ $p = .934$	–

Significant responses at the frequencies of interest correspond to SNR values higher than noise-level SNR. For responses to individual bodies, significance was tested for frequencies < 50 Hz; for responses to dyads, significance was tested for frequencies < 5 Hz. *p* Values are Bonferroni corrected for the number of harmonics tested at frequency.

significance using nonparametric cluster-mass permutation test (Maris & Oostenveld, 2007), with 5000 random permutations of the condition labels on the original data. For each shuffled data set, a t test was carried out for each sensor separately. Neighboring sensors yielding above-threshold t values were clustered together. The significance probability of the original clusters was computed as the number of times the shuffled data produced clusters with higher summed t values than the real data. This analysis revealed a posterior cluster of electrodes with significantly higher response to upright than inverted bodies ($p = .03$; Figure 2B).

Effect of Positioning and Orientation of Body Dyads

In the first analysis, we targeted neural correlates of body perception and tested the hypothesis that body positioning affects the visual processing of bodies, as suggested by recent fMRI studies (Bellot et al., 2021; Abassi & Papeo, 2020; Walbrin & Koldewyn, 2019). In the cluster of electrodes that showed higher activity for upright versus inverted single bodies (see above Definition of Body-related Response section), we studied the effect of positioning (facing or nonfacing) and/or orientation (upright or inverted) on the response to dyads. To this end, the mean SNR values for Fd and 2Fd (= F1 – F2) over the predefined cluster were entered into a repeated-measures ANOVA, with factors Positioning and Orientation. This analysis revealed a main effect of Positioning, $F(1, 19) = 5.5, p = .030$, with higher activity for facing than for nonfacing dyads (Figure 2C), but no effect of Orientation, $F(1, 19) = 2.2, p = .152$, or interaction, $F(1, 19) < 1, ns$.

In a second analysis without a priori hypothesis about the spatial distribution of the responses, we tested the effects of bodies positioning, orientation, and their interaction on the SNR values, at each sensor, using repeated-measure ANOVAs. Neighboring sensors showing a main effect or an interaction with an F value higher than the threshold corresponding to $p < .05, F(1, 19) > 4.3$, were considered as clusters. Clusters were tested for significance using nonparametric cluster-mass tests, with 5000 random permutations of the condition labels on the original data. For each shuffled data set, an ANOVA was carried out, separately for each sensor. Neighboring sensors yielding above-threshold F values were clustered together. The significance of the original cluster was computed as the number of times the shuffled data set produced clusters with higher summed F values than the real data. This analysis highlighted an anterior cluster, showing higher activity for facing than for nonfacing dyads (cluster-mass permutation test: $p = .03$; Figure 2C). We found no significant effect of Orientation and no interaction between the two factors (all $ps > .05$).

Finally, we tested whether the average response at F1 + F2 identified in the three posterior sensors was affected by positioning or orientation of bodies in the dyads. A repeated-measures ANOVA showed no effect of Positioning

($F < 0.1, p = .980$), Orientation ($F = 4.0, p = .060$), or interaction ($F < 0.1, p = .875$).

In summary, our analyses highlighted activations distributed over posterior and anterior sensors, suggesting an integrated representation of the two bodies in a face-to-face (relative to back-to-back) configuration.

Effect of Single Body Direction

In this analysis, we assessed whether the difference between facing and nonfacing dyads observed in posterior and anterior clusters was because of the processing of the dyad as a whole (i.e., to the processing of the relation between two bodies), or could be rather accounted for by a difference in the visual treatment of each single body in the facing versus nonfacing condition. We reasoned that, in the latter case, the identified clusters should show different responses not only to dyads (i.e., at dyad-related frequencies Fd and 2Fd) but also to single bodies in the facing versus nonfacing condition (i.e., at the single-body-related frequencies F1 and F2).

To this end, we tested the effect of Positioning and Orientation, separately for each cluster, considering the mean SNR values measured at F1 and F2. In the posterior cluster, this analysis confirmed the effect of Orientation, $F(1, 19) = 18.5, p < .001$ (see above Definition of Body-related Response section), with no effect of Positioning, $F(1, 19) = 0.2, p = .643$, or interaction, $F(1, 19) = 0.2, p = .675$. No effect reached significance in the anterior cluster (Orientation: $F(1, 19) = 0.5, p = .498$; Positioning, $F(1, 19) = 0.7, p = .417$; interaction: $F(1, 19) = 3.2, p = .089$). This analysis showed that the above effect of positioning of bodies in dyads could not be explained by the visuospatial features of single bodies such as the direction relative to central fixation (inward vs. outward).

DISCUSSION

Using a frequency-tagging EEG paradigm, we recorded frequency-separable responses to body dyads (at Fd and F1 – F2) and to each single body that formed the dyad (at F1 and F2). We first identified neural response associated with perception of single bodies at F1 and F2, using the contrast of upright versus inverted bodies (Stekelenburg & de Gelder, 2004). This analysis revealed a posterior cluster, compatible with the fMRI activity evoked in the occipito-temporal cortex, by viewing canonical upright (vs. inverted) bodies (Brandman & Yovel, 2010, 2016). Within that cluster, we tested the hypothesis that the neural correlates involved in body perception are sensitive to the relative positioning of multiple bodies in a visual scene. In line with previous studies (Bellot et al., 2021; Abassi & Papeo, 2020; Walbrin & Koldewyn, 2019), we found an effect of positioning, with larger response to facing than to nonfacing dyads at Fd and F1 – F2. Extended to the whole scalp (i.e., all the sensors), our analysis

revealed another cluster with anterior distribution, showing larger response to facing than to nonfacing dyads.

Unlike the responses evoked at the dyad-related frequencies, the responses evoked by single bodies did not vary depending on whether a body was part of a facing or a nonfacing dyad. Thus, the effect of positioning at Fd and F1 – F2 captured the response to the whole dyadic stimulus, rather than a change in the response to a single body seen in a facing versus a nonfacing dyadic context. In other words, the effect reported here appears to reflect encoding of relative positioning (i.e., the positioning of a body relative to another), rather than the absolute body positioning (i.e., body orientation directed toward or away from the center).

Although the neural response to single bodies, at F1 and F2, was reduced when bodies were presented upside-down (see also Brandman & Yovel, 2010, 2016; Minnebusch et al., 2009; Stekelenburg & de Gelder, 2004), we did not see any effect of inversion on the neural response to dyads (i.e., at Fd 0.61 Hz and F1 – F2). Although multiple bodies have also been shown to be susceptible to the effect of inversion (Papeo & Abassi, 2019; Papeo et al., 2017), it is possible that such effect is visible at high stimulation frequencies (i.e., the frequency of single-body presentation) but not at slower stimulation frequencies (i.e., the frequency of dyad presentation). Moreover, a difference between upright and inverted dyads might not occur in the frequency domain, but in the time domain (responses might have different shape in the time course). Future studies should address this question using, for example, event-related designs.

Irrespective of the orientation, the response to dyads was effectively tagged at two frequencies, possibly reflecting different processes: Fd (0.61 Hz), corresponding to the periodic simultaneous appearance of two bodies in full contrast, and 1.21 Hz, corresponding to the second harmonic of Fd ($2 \times 0.606 \text{ Hz} = 1.21 \text{ Hz}$), and the nonlinear interaction between the two input frequencies (F1 – F2 = 1.21 Hz; Zemon & Ratliff, 1984). As one can appreciate from Figure 2A, the response at 0.61 Hz was comparable in strength with the response at 1.21 Hz. Because harmonics typically have lower amplitudes than the main response, it is less likely that the response at 1.21 Hz only reflected the harmonic of the response at Fd. This raises the possibility that the response at 1.21 Hz resulted from both the second harmonic of the response to the periodic visual stimulation (i.e., the periodicity of dyad presentation at Fd), and the nonlinear interaction of the response to the two stimulation frequencies (F1 – F2; Alp et al., 2016; Boremanse et al., 2013). Future studies should be designed to distinguish between the visual responses to the whole scene, (i.e., at Fd), and the nonlinear interactions between the parts that give rise to the whole scene, as marked in the emergent intermodulation responses at $nF1 \pm mF2$.

The response to dyads at other intermodulation frequencies such as F1 + F2 was clearly weaker than the

response at Fd/F1 – F2 in our study, as it was only observed when restricting the analysis to posterior sensors, in three adjacent sensors. The response at F1 + F2, as weak as it might be, further supports the involvement of integrative processes that arise from the nonlinear interaction of the main (single body) responses.

Frequency-tagged responses at the dyad-related frequencies Fd and F1 – F2 marked the neural representation of the two bodies together. Note that the two bodies were present on the screen in most frames of a trial, although at a varying contrast level. That is, participants experienced the presence of two flickering bodies throughout the trial and were not aware that the bodies were only occasionally (i.e., at Fd) shown together in full view. This feature of the paradigm implies that the response to dyads at Fd captured a process that was spontaneous—if not automatic—given certain input characteristics, and independent of an explicit goal or task (Fodor, 1983); most likely a perceptual process. On this background, the larger response to dyads of facing (vs. nonfacing) bodies suggests the recruitment of additional processing, distributed over posterior and anterior areas, beyond the processing of multiple single bodies that were identical in facing and nonfacing dyads. Although previous studies have primarily focused on the effect of spatial relations between bodies in posterior (visual) areas, increased frontal activity for facing dyads has also been observed (Bellot et al., 2021; Abassi & Papeo, 2020). We note that, here, posterior activity was higher for facing dyads, but was also present for nonfacing dyads; instead anterior activity was observed for facing dyads, but was virtually absent for nonfacing dyads at Fd. Higher posterior (i.e., visual cortex) activity for facing dyads has been interpreted as an effect of the visual enhancement of multiple-body configurations that cue interaction (Papeo, 2020). Less clear is the role of frontal activity in this processing. The selectivity of anterior activity observed here for facing dyads opens to the hypothesis that it reflects a binding process triggered when two bodies are combined into a unitary representation.

We interpret the facing versus nonfacing dyad effect at Fd and F1 – F2 in the spirit of previous studies, where frequency-tagging paradigms were used to mark the integrative processing that mediates grouping of multiple parts into a whole, such as geometrical elements into a new object representation (e.g., the Kanizsa triangle; Alp et al., 2016; Gundlach & Müller, 2013; Aissani et al., 2011), or face-halves in a face (Boremanse et al., 2013). Those studies demonstrated that spatial relations between parts are key to trigger the integration captured by the enhanced response to a canonical versus noncanonical configuration. Thus, for example, two face-halves evoked increased response at the integration frequency when aligned but not when misaligned (Boremanse et al., 2013). By analogy, in the current study, two bodies increased the response at the alleged integration frequency, when they appeared simultaneously and facing

toward—but not away from—each other. Adding to the extant literature, our results show that grouping can account not only for object formation through binding of disconnected, but spatially organized, parts (Alp et al., 2016; Boremanse et al., 2013; Gundlach & Müller, 2013; Aissani et al., 2011) but also for the formation of scenes (or events) by binding multiple disconnected objects (here, bodies) together. The binding mechanism that would underlie the increased response to facing dyads may account for behavioral and neural effects showing attentional/perceptual advantages for sets of objects in spatial configuration that cues interaction (e.g., face-to-face bodies; see Papeo, 2020), a common function or usage (e.g., a pen over a notebook; see Kubilius et al., 2015), or a coherent scene (e.g., a lamp above the table; see Kaiser et al., 2019). It remains unknown whether a single domain-general mechanism accounts for binding of objects of different categories and according to different types of relations (e.g., bodily/social interaction, statistical regularity, semantic relatedness). In other words, whether the effect described here for facing body dyads is specific to bodies, or it applies to other object sets that can form coherent scenes (e.g., a lamp above a table) or functional groups (e.g., a pen above a notebook), remains an outstanding question for future research.

In conclusion, by contrasting the response to facing versus nonfacing dyads involving the same individual bodies, we were able to disentangle objective neural responses to parts (individual body) and to the whole scene (the dyad). The response to facing (vs. nonfacing) dyads described here echoes neural effects that have been consistently associated with perceptual grouping (also called visual binding), and suggests that the face-to-face body positioning can trigger a new representation that is more than the sum of the constituent parts. The representational content of the composite representation emerging from two facing bodies remains a fascinating question for further research. By extending the frequency-tagging paradigm to a new class of stimuli (i.e., multiple-body stimuli), our results contribute to characterizing grouping as a general mechanism that could account not only for the representation of objects from parts but also for the representation of scenes from objects or events from social entities.

Reprint requests should be sent to Parvaneh Adibpour, CNRS, Institut des Sciences Cognitives - Marc Jeannerod, 67 Boulevard, Pinel, 69675, Bron, France, or via e-mail: parvaneh.adibpour@isc.cnrs.fr.

Author Contributions

Parvaneh Adibpour: Conceptualization; Data curation; Formal analysis; Writing—Original draft; Writing—Review & editing. Jean-Rémy Hochmann: Conceptualization; Funding acquisition; Supervision; Writing—Original draft; Writing—Review & editing. Liuba Papeo: Conceptualization; Funding

acquisition; Supervision; Writing—Original draft; Writing—Review & editing.

Funding Information

This work was supported by ANR grant awarded to J.-R. H., Agence Nationale de la Recherche (<https://dx.doi.org/10.13039/501100001665>), Project: TACTIC, grant number: ANR-16-CE28-0006, and H2020 European Research Council Starting (<https://dx.doi.org/10.13039/100010663>), awarded to L. P., Project: THEMPO, grant number: 758473.

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were $M(\text{an})/M = .408$, $W(\text{oman})/M = .335$, $M/W = .108$, and $W/W = .149$, the comparable proportions for the articles that these authorship teams cited were $M/M = .579$, $W/M = .243$, $M/W = .102$, and $W/W = .076$ (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

- Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience*, 40, 852–863. <https://doi.org/10.1523/JNEUROSCI.1378-19.2019>, PubMed: 31801812
- 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>, PubMed: 21782159
- 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>, PubMed: 26867088
- Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: The weighted average as a general rule. *Neuroimage*, 70, 37–47. <https://doi.org/10.1016/j.neuroimage.2012.12.023>, PubMed: 23266747
- Baldassano, C., Beck, D. M., & Fei-Fei, L. (2017). Human-object interactions are more than the sum of their parts. *Cerebral Cortex*, 27, 2276–2288. <https://doi.org/10.1093/cercor/bhw077>
- 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, 2670–2685. <https://doi.org/10.1093/cercor/bhaa382>
- 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, 6. <https://doi.org/10.1167/13.11.6>, PubMed: 24023273
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. <https://doi.org/10.1163/156856897X00357>, PubMed: 9176952
- Brandman, T., & Yovel, G. (2010). The body inversion effect is mediated by face-selective, not body-selective, mechanisms. *Journal of Neuroscience*, 30, 10534–10540. <https://doi.org/10.1523/JNEUROSCI.0911-10.2010>, PubMed: 20685996
- Brandman, T., & Yovel, G. (2016). Bodies are represented as wholes rather than their sum of parts in the occipital-temporal cortex. *Cerebral Cortex*, 26, 530–543. <https://doi.org/10.1093/cercor/bhu205>, PubMed: 25217470
- Coren, S., & Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The gestalt illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 404–412. <https://doi.org/10.1037/0096-1523.6.3.404>, PubMed: 6447756
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>, PubMed: 15102499
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/4737.001.0001>
- Glanemann, R., Zwitserlood, P., Bölte, J., & Dobel, C. (2016). Rapid apprehension of the coherence of action scenes. *Psychonomic Bulletin & Review*, 23, 1566–1575. <https://doi.org/10.3758/s13423-016-1004-y>
- Gundlach, C., & Müller, M. M. (2013). Perception of illusory contours forms intermodulation responses of steady state visual evoked potentials as a neural signature of spatial integration. *Biological Psychology*, 94, 55–60. <https://doi.org/10.1016/j.biopsycho.2013.04.014>, PubMed: 23665197
- 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>, PubMed: 29459238
- Kaiser, D., Häberle, G., & Cichy, R. M. (2020). Cortical sensitivity to natural scene structure. *Human Brain Mapping*, 41, 1286–1295. <https://doi.org/10.1002/hbm.24875>, PubMed: 31758632
- Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *Neuroimage*, 169, 334–341. <https://doi.org/10.1016/j.neuroimage.2017.12.065>, PubMed: 29277645
- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, 23, 672–685. <https://doi.org/10.1016/j.tics.2019.04.013>, PubMed: 31147151
- 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, U.S.A.*, 111, 11217–11222. <https://doi.org/10.1073/pnas.1400559111>, PubMed: 25024190
- Kim, J. G., & Biederman, I. (2011). Where do objects become scenes? *Cerebral Cortex*, 21, 1738–1746. <https://doi.org/10.1093/cercor/bhq240>, PubMed: 21148087
- Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace and Company.
- Kubilius, J., Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2015). Brain-decoding fMRI reveals how wholes relate to the sum of parts. *Cortex*, 72, 5–14. <https://doi.org/10.1016/j.cortex.2015.01.020>, PubMed: 25771992
- Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., et al. (2020). Maternal odor shapes rapid face categorization in the infant brain. *Developmental Science*, 23, e12877. <https://doi.org/10.1111/desc.12877>, PubMed: 31175678
- MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, 19, 943–947. <https://doi.org/10.1016/j.cub.2009.04.020>, PubMed: 19446454
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>, PubMed: 17517438
- McMains, S. A., & Kastner, S. (2010). Defining the units of competition: Influences of perceptual organization on competitive interactions in human visual cortex. *Journal of Cognitive Neuroscience*, 22, 2417–2426. <https://doi.org/10.1162/jocn.2009.21391>, PubMed: 19925189
- Minnebusch, D. A., Suchan, B., & Daum, I. (2009). Losing your head: Behavioral and electrophysiological effects of body inversion. *Journal of Cognitive Neuroscience*, 21, 865–874. <https://doi.org/10.1162/jocn.2009.21074>, PubMed: 18702581
- 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, 4. <https://doi.org/10.1167/15.6.4>, PubMed: 26024451
- Papeo, L. (2020). Twos in human visual perception. *Cortex*, 132, 473–478. <https://doi.org/10.1016/j.cortex.2020.06.005>, PubMed: 32698947
- Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual search for people among people. *Psychological Science*, 30, 1483–1496. <https://doi.org/10.1177/0956797619867295>, PubMed: 31532709
- Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, 28, 369–379. <https://doi.org/10.1177/0956797616685769>, PubMed: 28140764
- Quaddlieg, 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>, PubMed: 25697049
- Reddy, L., Kanwisher, N. G., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences, U.S.A.*, 106, 21447–21452. <https://doi.org/10.1073/pnas.0907330106>, PubMed: 19955434
- Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, 20, 238–248. [https://doi.org/10.1016/0013-4694\(66\)90088-5](https://doi.org/10.1016/0013-4694(66)90088-5), PubMed: 4160391
- Regan, D. (1989). *Methods for recording steady-state evoked potentials. Human brain electrophysiology. Evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.
- Roberts, K. L., & Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *Neuroimage*, 52, 1541–1548. <https://doi.org/10.1016/j.neuroimage.2010.05.044>, PubMed: 20580845
- Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: An ERP study on the body-inversion effect. *NeuroReport*, 15, 777–780. <https://doi.org/10.1097/00001756-200404090-00007>, PubMed: 15073513
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011, 879716. <https://doi.org/10.1155/2011/879716>, PubMed: 21584256
- 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>, PubMed: 32220782

- 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*, 1251–1268. <https://doi.org/10.1037/xge0000545>, PubMed: 30652892
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., et al. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*, 1172–1217. <https://doi.org/10.1037/a0029333>, PubMed: 22845751
- 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>, PubMed: 31100434
- Zemon, V., & Ratliff, F. (1984). Intermodulation components of the visual evoked potential: Responses to lateral and superimposed stimuli. *Biological Cybernetics*, *50*, 401–408. <https://doi.org/10.1007/BF00335197>, PubMed: 6487677
- 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*, 847–855. <https://doi.org/10.1038/s41562-019-0618-2>, PubMed: 31182793