Seeing relationships:

The specialization for a two-body shape in human visual perception

Abassi Etienne & Papeo Liuba*

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

*Correspondence to: CNRS, Institut des Sciences Cognitives—Marc Jeannerod, 67 Boulevard Pinel, 69675, Bron, France; Phone: +33 437911266; E-mail: liuba.papeo@isc.cnrs.fr

Abstract (max 249)

The human social nature manifests in, among other ways, a particular visual sensitivity to social entities such as faces and bodies. But, the core of a social representation is not as much the social entity as the relation that binds multiple entities together. We asked whether human vision exhibits a special sensitivity to socially relevant spatial relations, beyond the visual sensitivity to likely members of those relations. Some social relations reliably correlate with spatial relations: people are likely to physically interact face-to-face more than back-to-back. Using functional MRI and behavioral measures, we show that visual sensitivity to social stimuli extends to images encompassing two bodies facing toward (vs. away from) each other. In particular, the visual-perception lateral occipital cortex showed an organization with the inferior part encoding of the number of bodies (one vs. two) and the superior part, encoding the spatial relation between bodies (facing vs. nonfacing). Moreover, the functionally localized body-selective cortex responded to facing bodies more strongly than to identical but not facing bodies. Finally, a multivariate pattern analysis showed that representations of single bodies in facing (vs. nonfacing) dyads were sharpened, suggesting that the spatial positioning cuing interaction put pressure on the discrimination of body postures. Finally, the cost of body inversion (upside-down rotation) on body recognition, a behavioral signature of the specialized mechanism for body perception, was larger for facing than for nonfacing dyads. Thus, the human visual perception system registers spatial relations between multiple bodies, with tuning for those that cue social interaction.

Public Significance Statement (120)

The humans' social nature shapes visual perception. Here, we show that human vision is not only attuned to socially relevant entities, such as bodies, but also to socially relevant spatial relations between those entities. The body-selective visual cortex responds more strongly to multiple bodies that appear to interact (i.e., face-to-face), relative to unrelated bodies, and shows more accurate representation of single body postures in interacting scenarios. Moreover, recognition of facing bodies is particularly susceptible to perturbation by upside-down rotation, indicative of a particular visual sensitivity to the canonical appearance of facing bodies. Encoding of relations between multiple bodies, in the same areas for body-shape recognition, suggests that the context in which a body is seen, deeply affects its processing.

Introduction

The term *social vision* has been coined to refer to a number of phenomena in visual perception that mediate the humans' social life (1).

Among those phenomena, there is a tuning of human visual perception, and visual perceptual areas, to certain features or entities that turn out to have high social value. Stimuli such as bodies and faces are individuated earlier than other stimuli in infancy (2). In cluttered environments, they are the most likely to spontaneously recruit attention (3, 4). They are also particularly susceptible to the cost of inversion, the detrimental effect on recognition, of seeing a stimulus rotated upside-down (5, 6). This effect has been linked to an internal representation of the canonical (upright) structure of faces and bodies, which makes recognition particularly efficient, and the disruption of such structure (e.g., through inversion) particularly harmful to recognition (7). Within the visual perception occipitotemporal cortex, a number of brain areas, collectively called person-perception network, exhibit a preference for faces and bodies, indexed by increased neural activity (8). Core aspects of this network are the face-selective cortex in the fusiform gyrus (FFA) (9) and the extrastriate body-selective cortex (EBA) in the lateral occipital area (LOC) (10).

Face and body perception are fundamental to several social tasks, like identity recognition, social categorization and emotion recognition, which has contributed to characterize human visual perception as social (1, 11, 12). However, *social* is primarily the property of a relation that implies two or more entities, and binds them together. Some social relations, such as physical and/or communicative exchanges between people, reliably correlate with spatial relations: for example, physically interacting people are often close and face toward, rather than away from each other. Detecting and recognizing social interactions is as important to human survival as detecting and recognizing bodies, faces and animate entities in general. Third-party interactions must entail rapid discrimination, to activate adaptive behaviors such as defense or assistance/cooperation, and infer group affiliation and conformity (13, 14). We asked whether, besides the increased sensitivity to social entities, human visual perception registers spatial relations between those entities, and shows preference for socially relevant layouts.

Recent reports have shown that interactions between multiple animate agents are recognized just outside the visual perception network, in posterior superior temporal areas (15–17). However, recent behavioral research suggests that seeing bodies in interaction makes body detection and recognition more efficient with respect to seeing multiple, but unrelated bodies (18). This research encourages the thinking that spatial relations between bodies can be registered earlier, within the brain areas that recognize body-shapes.

In the present fMRI and behavioral studies, we examined how the spatial relation between two bodies, which cued interaction (face-to-face, hereafter facing) or not (back-to-back, hereafter nonfacing), affected neural activity in person-selective cortex, and performance in visual recognition. Images of single bodies and dyads of facing or nonfacing bodies were presented during fMRI, and the same stimuli, upright or inverted, during a visual recognition task with backward masking. We hypothesized a stronger visual sensitivity to facing, relative to nonfacing dyads, which would manifest with stronger neural response and more accurate representation in the body-selective cortex, and a larger inversion effect on visual recognition. Neural and behavioral results converged in demonstrating that human visual perception is tuned to configurations where the relative spatial positioning of bodies suggests an ongoing social exchange.

Results

fMRI Experiment. In the main experiment, twenty participants in the fMRI scanner saw images featuring one or two human bodies, in lateral view and in various poses. In images with two bodies, these could face toward or away from each other. Facing and nonfacing dyads were identical except for the relative positioning of bodies. All the bodies presented in dyads were also presented alone. A fixation cross was always present in the center of the screen and, from time to time, changed color (from black to red). Participants were instructed to fixate the cross, detect and report (through button press) the color change. After the main experiment, participants completed a functional localizer task, in which they saw a new set of images featuring bodies (headless bodies or body parts), faces, places (indoor corridors or houses), objects (cars or guitars), or scrambled objects. They had to press a button when a scrambled object was shown.

To identify the brain regions responsive to the number of bodies and to the spatial relation in body dyads, we performed a whole-brain random-effects group analyses with two contrasts: [single body > dyad] and [facing > nonfacing dyads] (Table 1). A large bilateral cluster, centered in the inferior temporal gyrus, responded significantly more strongly to dyads than to single bodies (Fig. S1). This activation spread more posteriorly into the lateral occipital cortex. Stronger response to facing than nonfacing dyads was found in a bilateral cluster, encompassing the posterior middle temporal gyrus and the middle occipital gyrus. In addition, relative to nonfacing dyads, facing dyads elicited stronger bilateral prefrontal activity, centered in the anterior middle frontal gyrus (Fig. 1a).

The group-level contrasts encouraged our hypothesis that the visual perception occipitotemporal cortex is particularly sensitive to scenarios, in which the relative positioning of bodies (face-to-face) cues interaction. To further investigate these findings, and overtake the limits of group analysis in the definition of anatomo-functional correspondences (19), we implemented two further analyses. First, since the whole-brain analysis revealed widespread effects of both number (single body vs. dyad) and spatial relation (facing vs. nonfacing dyads) in the LOC, a conjunction analysis was performed to examine the relationship between the two effects across all the voxels in this territory (20). The conjunction of the statistical maps (dyads > individuals) and (facing >

nonfacing dyads) from the whole-brain analysis, superimposed to the anatomically-defined bilateral LOC, revealed three consecutive regions, organized along the inferior-to-anterior axis, with three different response profiles (Fig. 1b). Voxels in the inferior part of LOC mostly showed an effect of number, with significantly stronger response to dyads than to single bodies. Moving upward, the middle region encompassed voxels than responded more strongly to dyads than single bodies, and to facing dyads than nonfacing dyads. Finally, in the superior region, voxels responded to facing dyads more strongly than to nonfacing dyads, but showed no effect of number. As illustrated in Fig. 1b, the middle and superior regions of LOC, sensitive to the facing vs. nonfacing relation, overlapped with the body-selective EBA, as independently identified with the following analysis.

We studied whether spatial relations between bodies are encoded in the same cortex selective to body perception in LOC, identified in each participant with an independent dataset. Using the data registered during the functional localizer task, we defined two regions of interest (ROIs) for each participant, collapsing across the left- and right-hemisphere voxels: the EBA was defined using the contrast of bodies vs. objects conditions, and the FFA was defined using the contrast of faces vs. objects conditions. In addition, we defined the parahippocampal place area (PPA) with the contrast of places vs. objects conditions, collapsing individual subjects' data across hemispheres. The PPA has been shown to discriminate between object sets based on internal spatial relations, with particularly strong response to sets that form familiar or regular scenes (21, 22). Here, it was included to examine whether spatial relations between bodies are encoded selectively in the body-selective cortex. Finally, the individual early visual cortex (EVC), was defined using a probabilistic map of visual topography (23).

We quantified the response to facing vs. nonfacing dyads for each participant's ROI and observed that the relative positioning of bodies in a dyad affected neural activity, selectively in the EBA and FFA (Fig. 1c). This was confirmed by a two-way ANOVA with relation (facing vs. nonfacing dyads) and ROI (EBA vs. FFA vs. PPA vs. EVC) as repeated-measures factors. The ANOVA revealed a main effect of relation, $F(1,19) = 5.20$, $p = 0.034$, $\eta_p^2 = 0.21$, and of ROI, $F(3,57) = 91.88$, $p < 0.001$, $\eta_p^2 = 0.83$, which were qualified by a significant two-way interaction, $F(3,57) = 4.90$, $p = 0.004$, $\eta_p^2 = 0.21$. This interaction reflected significantly stronger response to facing dyads than to nonfacing dyads in the EBA, $t(19) = 3.09$, $p = 0.006$, and FFA, $t(19) = 3.60$, $p = 0.002$, but not in the PPA, $t(19) = 1.51$, $p = 0.148$, and EVC, $t(19) = 0.53$, $p > 0.250$ (two-tailed *t* tests). Thus, the EBA and the FFA respond more strongly to facing than to nonfacing dyads. This effect was confirmed using ROIs with voxel counts ranging from 50 to 500 (Fig. S2). Like the EVC, the PPA shows no effect of body positioning, suggesting that the EBA/FFA and the PPA participate in two separate streams that detect the relations between objects of different classes, and might link visual perception to action-event and scene representation, respectively. An analysis with the left and right ROIs considered separately for each site, showed no hemispheric difference across all the ROIs, and confirmed the above effects (Text SI and Fig. S1).

Three analytic strategies above (whole-brain analysis, voxel-by-voxel conjunction analysis in LOC, and ROI analysis) converged in showing stronger response to dyads with facing –seemingly interacting– bodies in the EBA. The following analysis sought to shed light on the mechanism underlying this increase of neural activity. What does the EBA do? The EBA is known to encode the specific body shape, or posture, in a percept (11, 24). We asked whether this functionality could just be enhanced in the context of facing dyads. We reasoned that, in a facing dyad, a body (posture) could provide a context to the other body (posture), which could enrich (i.e., by disambiguating) the representation of individual postures, and/or put particular pressure on their discrimination for the purpose of understanding the ongoing interaction. Using multivariate pattern analysis (MVPA), we measured how well single body postures could be discriminated in facing and in nonfacing dyads, in the four ROIs. In each ROI, for each participant, we estimated multivariate patterns of activity for every facing dyad, nonfacing dyad and single body. For each participant, a support vector machine (SVM) classifier was trained to classify the patterns of eight classes corresponding to the eight single bodies, and tested on the classification of patterns corresponding to facing dyads made of the same single bodies. In each test, a pattern corresponding to a facing dyad was classified in one of eight classes of single bodies. The same analysis was repeated using the patterns for nonfacing dyads in the test. For each participant, for each ROI, we obtained two values of classification accuracy, one for the train-and-test on facing dyads and one for the train-and-test on nonfacing dyads. We found that single bodies could be discriminated accurately only from facing dyads, and only using neural patterns extracted from the EBA (Fig. 1e; for the results of the left and right ROIs, separately, see Fig. S1). In particular, in the EBA, classification accuracy was significantly above chance for facing dyads, $t(19) = 4.69$, $p < 0.001$, but not for nonfacing dyads, $t(19) = 0.40$, $p > 0.250$. This result was obtained with different voxel counts ranging from 50 to 500 (Fig. 1f). For the other ROIs, the same analysis yielded no significant effects (FFA: test on facing dyads: $t(19) < 1$, *n.s.*, on nonfacing dyads: $t(19) < 1$, *n.s.*; PPA: test on facing dyads: $t(19) < 1$, *n.s.*, on nonfacing dyads: $t(19) <$ 1, *n.s.*; EVC: test on facing dyads: $t(19) = 1.98$, $p = 0.063$, on nonfacing dyads: $t(19) < 1$, *n.s.*).

Visual recognition experiment. We tested whether the neural tuning hypothesized for facing dyads, was associated with increased visual sensitivity to those stimuli, as indexed by the magnitude of inversion effect: the more the visual system is attuned to the canonical (upright) appearance of a stimulus, the higher the cost on recognition, of disrupting that canonical configuration (e.g., through inversion; (25, 26). To test this, we presented the same images of single bodies, facing dyads and nonfacing dyads used for fMRI, randomly interleaved with images of chairs (single chair, or pairs of facing or nonfacing chairs), for a visual recognition task. Stimuli were presented briefly (30 ms), upright or inverted, and then masked. Twenty participants (15 of whom already participated in the fMRI experiment) were instructed to report, as accurately and fast as possible, whether they had seen bodies or chairs, irrespective of number, positioning or orientation. The magnitude of the inversion effect on both visual-recognition accuracy and RTs varied as a function of the stimulus (Fig. 1d). An ANOVA with stimulus (single body vs. facing dyads vs. nonfacing dyads) and orientation (upright vs. inverted) as repeated-measures factors, showed significant effects of stimulus, $F(2,38) = 5.52$, $p = 0.008$, $\eta_p^2 = 0.22$, and orientation, $F(1,19) = 8.52$, $p = 0.009$, $\eta_p^2 = 0.31$, and a significant two-way interaction, $F(2,38) = 3.94$, $p = 0.028$, $\eta_p^2 = 0.17$. This two-way interaction reflected a statistically significant inversion effect for single bodies, $t(19) = 2.75$, $p = 0.013$, and facing dyads, $t(19) = 3.31$, $p = 0.004$, but only marginal for nonfacing dyads, $t(19) = 2.01$, $p = 0.058$. Most importantly, the inversion effect was significantly larger for facing dyads than for nonfacing dyads, $t(19) = 3.68$, $p = 0.002$. Similar results were found in the ANOVA on RTs [effects of stimulus: $F(2,38) = 15.90$, $p < 0.001$, $\eta_p^2 = 0.45$; effect of orientation: $F(1,19) = 24.13$, $p < 0.001$, $\eta_p^2 = 0.56$; interaction, $F(2,38) = 4.22$, $p = 0.22$, $\eta_p^2 = 0.18$]. In the RTs, the inversion effect was significant in all conditions [single bodies: $t(19) = 4.60$, $p < 0.001$; facing dyads, $t(19) = 5.14$, $p < 0.001$; nonfacing dyads, $t(19) = 3.73$, $p = 0.001$]; but, again, it was significantly larger for facing dyads than for nonfacing dyads, $t(19) = 2.72$, $p = 0.001$. No differences in the inversion effect were found with chair-trials; the whole set of accuracy and RTs results was obtained including only the 15 participants who took part in the fMRI experiment (SI Text). Thus, as an index of efficiency in visual perception, the inversion effect offers a behavioral counterpart to the increase of neural activity and the sharpening of neural representation for facing dyads. As such, it contributes to support a special visual sensitivity for two bodies in a spatial positioning that cues social interaction.

Discussion

We investigated whether the relative positioning of two bodies –facing one another as if interacting, or not– affected neural and behavioral signatures of body perception. Our findings demonstrate that the well-established sensitivity of human visual perception to single bodies extends to configurations of multiple bodies, whose relative positioning cue interaction.

Three fMRI data analyses and the behavioral results converged on demonstrating that cortical areas specialized in body perception capture spatial relations between bodies. Particularly in the EBA, stronger response to facing vs. nonfacing dyads was found in the group-level whole-brain contrast, in the voxel-by-voxel conjunction analysis in the LOC, and in the analysis of independently, functionally-defined ROIs. Visual sensitivity to a stimulus, which accounts for neural activity in dedicated brain structures, should also be captured in visual perception behavior (27). Thus, the two-body inversion effect, larger for facing than nonfacing dyads, provides the fourth piece of evidence in favor of a tuning of human visual perception to configurations of multiple, seemingly interacting bodies.

The stronger neural response to facing dyads echoes category-specific effects reported in occipital and temporal regions, for stimuli such as bodies and faces (9, 28), as well as the effect, in the general object-selective LOC, of multiple objects appearing in functionally relevant relations (e.g., a hammer tilted toward a nail) vs. multiple unrelated objects (29–31). But, increase in neural response can cover different underlying processes. What specific process, triggered by interacting bodies, accounts for the increased neural response in the EBA? The results of the MVPA analysis suggest that the natural functionality of EBA is enhanced for facing dyads. The EBA is known to encode visual body shapes or postures, to feed the process toward action understanding (11, 24); here we show that it does so more, or better, when a body appears in the context of another interacting body. Discrimination of single bodies in facing dyads (but not in nonfacing dyads) was accurate, despite the high visual similarity across bodies, and despite the instruction to fixate a cross and detect a color change diverted participants' attention from bodies.

It is possible that, by signaling an ongoing or upcoming interaction, the face-to-face positioning puts pressure on the encoding of body postures, to streamline the processing toward action (and interaction) understanding. The link between the effect in the EBA and the action understanding network might be emphasized by the fact that the stronger response to facing (vs. nonfacing) dyads carries on beyond the EBA, in the superior LOC and the middle/superior temporal gyrus, an important territory for action understanding (32), with patches of selectivity for social interactions (15, 17).

In another perspective, one seemingly interacting body can provide a meaningful context to the other body, which enriches or sharpens body representations in EBA. The advantage on object representation that results from seeing an object in a meaningful, or in its natural context, has been largely illustrated by visual context effects, such as the word superiority effect (33) and the face superiority effect (34, 35): an object (e.g., a letter or a nose) is identified better, when it is seen in its regular context (i.e., a familiar word or a face, respectively). Recently, neuroimaging methods have allowed demonstrating that a stimulus that meets subjective expectations, or an internal template, is represented more accurately than an unexpected stimulus, in the visual cortex, and that the surrounding context can favor this representational sharpening (36, 37). Consistent with fMRI results, the two-body inversion effect contributes to suggest that the facing dyad meets an internal template, which makes it easier to recognize, and particularly susceptible to spatial perturbation through inversion. In particular, the difference in the inversion effect between facing and nonfacing dyads might be reminiscent of the difference between bodies and scrambled bodies (for which the body-inversion effect does not occur; 6): like a scrambled body, a nonfacing dyads would break the prototypical, regular, and expected configuration of two spatially close bodies, which reduces the cost of inversion.

In sum, our results show that, in the EBA, facing dyads evoke overall stronger response than nonfacing dyads, and sharpen the representation of single bodies. These findings open questions for future research. First, like the EBA, the FFA registers the relation between bodies with a stronger response to facing vs. nonfacing dyads. However, we found no evidence of representational sharpening for individuals in facing vs. nonfacing dyads. Lack of discrimination is not surprising given that all bodies had identical (emotionally neutral) head/faces. Thus, the contribution of face/head vs. body, and the role of the face-selective cortex in the effects of body positioning remain to establish. Moreover, the dissociation of the two effects (increased activation and representational sharpening) raises the possibility that different visual perception regions (e.g., EBA and FFA) implement different operations, and/or that the two effects are signatures of different processes. Second, and related to the last point, we have somehow implied that, in the EBA, the representational sharpening accounts for the increased response to facing dyads. There is an alternative to this interpretation. Behavioral research suggests that relations between bodies (facing vs. nonfacing) are captured early and automatically –i.e., before an elaborate, conscious recognition of actions from body postures (18). Instead, sharpening of single objects' representation has been characterized as a top-down effect on the visual cortex (37). Thus, in a hypothetical architecture, regions such as the EBA (and the FFA) could capture spatial relations between bodies, yielding stronger neural response to facing dyads, and feed, with this information, the system for action understanding. This higher-level system could in turn put pressure on the encoding of single body postures, yielding representational sharpening in the EBA. The activation of temporal and prefrontal regions triggered by facing dyads, might signal the recruitment of higher-level regions for action and/or social processing (15, 17, 38, 39).

Laying the foundation for new research, this initial report of a special visual sensitivity to a two-body shape demonstrates that the relation of a body with another nearby body is captured in the same regions that serve body perception, and affects the way in which bodies are processed. Moreover, it suggests that, beyond the specialization for socially relevant entities, human vision is specialized to processing multi-body configurations, where the relative positioning of bodies signals the unfolding of a social event. This special sensitivity to a socially relevant spatial relation between people decisively contributes to characterize human vision as *social*.

Methods

fMRI experiment

Participants. Twenty healthy adults (12 female; mean age 24.1 years, *SD* = 3.1) took part in the fMRI study as paid volunteers. All had normal or corrected-to-normal vision and reported no history of psychiatric or neurological disorders, or use of psychoactive medications. They were screened for counter indications to fMRI and gave informed consent before participation. The local ethics committee (CPP Sud Est V, CHU de Grenoble) approved the study.

Stimuli. Stimuli were gray scale renderings of one or two human bodies seen from lateral view, in various biomechanically possible poses, created and edited with Daz3D (Daz Productions, Salt Lake City, UT) and the Image Processing Toolbox of MATLAB (The MathWorks, Natick, MA). Eight unique body poses and their flipped version formed the single-body set. Eight facing dyads were created from the eight unique body poses; therefore, each body was used twice, every time paired with a different body. The final facing-dyad set included the eight dyads and their flipped version. Nonfacing dyads were created by swapping the position of the two bodies in each facing dyad (i.e., the body on the left side was moved to the right side and *vice versa*). The centers of the two minimal bounding boxes that contained each figure of a dyad, and the distances between the closest points of the two bodies were matched across facing and nonfacing dyads (distance centers: $t(7) = 0.78$, $p > 0.250$; distance extremities: $t(7) = 0.78$, $p > 0.250$). Thus, facing and nonfacing dyads were identical except for the relative positioning of bodies. The final stimulus set included a total of 48 stimuli (16 single bodies and 32 dyads).

Procedure. Facing and nonfacing dyads were presented over three runs. Each run included 32 blocks of 5.6 s, 16 with facing and 16 with nonfacing dyads, presented in two sequences: the first 16 blocks in random order; the remaining 16 in the counter-balanced order relative to first sequence. Each block featured five repetitions of the same image, randomly alternating between the original view and its flipped version. Thus, in each of the three runs, there were two blocks for each dyad, for a total of six blocks for each dyad across the whole experiment. Three additional runs featured single-body images. Each run included 16 blocks, two for each stimulus, presented in the original view or flipped. A total of six blocks for each single body was shown across the experiment. Runs with dyads and runs with single bodies were presented in pseudorandom order to avoid more than two consecutive runs of the same stimulus group. Each run began with a warm-up block (10 s) and ended with a cool-down block (16 s), during which a central fixation cross was presented. Within a run, the onset time of each block was jittered to remove the overlap from the estimate of the hemodynamic response (40). Jittering was optimized using the optseq tool of Freesurfer (41). Throughout a block, a black cross was always present in the center of the screen, while stimuli appeared for 400 ms, separated by an interval of 900 ms. In a subset (37.5.%) of stimulus and fixation blocks, the cross turned red. Participants were instructed to fixate the cross throughout the experiment, and press the button of a remote with their right index finger, when the cross changed color. This task was used to minimize eye movements and maintain vigilance in the scanner. Stimuli were back-projected onto a screen by a liquid crystal projector (frame rate: 60 Hz; screen resolution: 1024 × 768 pixels, screen size: 40 x 30 cm). Participants viewed the stimuli binocularly (~7° of visual angle) through a mirror above the head coil. Stimulus presentation and response collection were controlled through MATLAB (Psychtoolbox;42).

Functional Localizer task. In addition to the six experimental runs, participants performed three functional localizer runs (4.26 min each), to identify, at the individual level, the body-selective EBA, the face-selective FFA and the place-selective PPA. Stimuli and task were adapted from the fLoc package (43). Participants saw 144 grey-level images of bodies (hands, arms, feet or whole headless bodies), faces, places (indoor corridors or houses), and inanimate objects (cars or guitars). To minimize low-level differences across categories, the view, size, and retinal position of the images varied across trials, and each item was overlaid on a 10.5° phase-scrambled background generated from another image of the set, randomly selected. Each run included 52 blocks of four seconds: 10 blocks for each object-class (bodies, faces, places and objects) with eight images per block (500 ms per image without interruption), randomly interleaved with 12 baseline-blocks featuring an empty screen.

Data acquisition. Imaging was conducted on a MAGNETOM Prisma 3T scanner (Siemens Healthcare). T2*-weighted functional volumes were acquired using a gradient-echo echo-planar imaging sequence (GRE-EPI) (repetition time = 2.2 s, echo time = 30 ms, 40 slices, slice thickness = 3 mm, no gap, field-of-view = 220 mm, matrix size = 74 x 74, acceleration factor of 2 with GRAPPA reconstruction and phase encoding set to anterior/posterior direction). For the main experiment and the functional localizer session, we acquired nine runs for a total of 1275 frames per participant. Acquisition of high-resolution T1-weighted anatomical images was performed after the third functional run and lasted 8 min (MPRAGE; 0.8 mm isotropic voxel size, repetition time = 3 s, echo time = 3.7 ms, TI = 1.1 s, field-of-view = 256 x 224 mm, acceleration factor of 2 with GRAPPA reconstruction).

fMRI Preprocessing. Functional images were preprocessed and analyzed using SPM12 (44) and MATLAB. The first four volumes of each run were discarded. Preprocessing of the remaining volumes involved: spatial realignment and motion correction using the first volume of each run as reference, slice timing correction, removing of low-frequency drifts with a temporal high-pass filter (cutoff 128 s), spatial smoothing with a Gaussian kernel of 8 mm FWHM for univariate analysis, and of 2 mm FWHM for multivariate analysis. Anatomical volumes were co-registered to the mean functional image, segmented into gray matter, white matter and cerebrospinal fluid in native space, and aligned to the probability maps in the Montreal Neurological Institute (MNI) as included in SPM12. The method DARTEL (45) was used to create one flow field for each subject and one inter-subject template, which was registered in the MNI space and used for normalization of functional images.

Whole-brain group analysis. Two whole-brain contrasts were performed to identify the neural effects of number (dyads > single-bodies) and relation (facing > nonfacing dyads). The blood-oxygen-level-dependent (BOLD) signal of each voxel in each participant was estimated in two general linear model

(GLM) analyses, both with two regressors for the experimental conditions (single-bodies and dyads, or facing and nonfacing dyads), one regressor for fixation-trials, and six regressors for the movement correction parameters as nuisance covariates. Statistical significance of second-level (group) analysis was determined using a voxelwise threshold of $p < 0.001$, family-wise error corrected at cluster-level.

Voxel-by-voxel conjunction analysis in LOC. Effects of number (single-bodies vs. dyads) and relation (facing *vs.* nonfacing dyads) were tested, voxel-by-voxel, in the LOC. To this end, the left and right functional maps from the group-level contrasts [dyads > individual bodies] and [facing dyads > nonfacing dyads] were superimposed on the masks of the left and right LOC, respectively. Three functional maps were defined within each mask, which encompassed all the voxels with activation higher than a threshold of $p = 0.001$ in the contrast facing > nonfacing only, in the contrast dyads > individuals only, or in both contrasts, as shown by a conjunction analysis (20).

ROIs definition and activity-based analysis. ROIs were defined by entering the individual data, registered during the functional localizer runs, into a GLM with four regressors for the four object-classes and movement correction parameters as nuisance covariates. Three bilateral masks of the inferior LOC, the Temporal Occipital Fusiform Cortex (TOFC) and the inferior Parahippocampal Cortex (PHC) were created using FSLeyes (46) and the Harvard-Oxford Atlas (47) through FSL (48). Within each mask of each participant, we first selected the voxels with activity above a voxelwise threshold of $p > 0.05$, for the contrasts of interest. The EBA within LOC was identified with the contrast of bodies > objects conditions, the FFA within TOFC was identified with the contrast of faces > objects conditions, and the PPA was localized within PHC with the contrast of places > objects conditions. All the voxels from the one mask and its contralateral homologous that passed the threshold were ranked by activation level (t value). The final ROI included up to 200 (right or left) best voxels (mean number of voxels in EBA and PPA: 200; in FFA: 198.4 ± 7.16 *SD*). Additionally, a bilateral mask of the early visual cortex (EVC) was created using a probabilistic map of visual topography in human cortex (23). After transforming the mask in each subject space, up to 200 (right or left) voxels with the highest probability ranking (mean = 200 voxels, SD = 0) were selected to form the EVC-ROI. From the four ROIs of each participant, the mean neural activity values (mean β-weights minus baseline) for facing and nonfacing dyads were extracted, and analyzed in a 2 relation x 4 ROI repeated-measure ANOVA, run with Statistica (StatSoft Europe, Hamburg). A second ANOVA included, in addition to the above factors, the hemisphere (left vs. right), corresponding to the mean β-weights extracted, separately, from the left and right ROIs (up to 200 voxels for each hemisphere; see SI text, Figure S1).

ROI-based MVPA. ROIs for this analysis were defined as above, using the data from the functional localizer task preprocessed with a spatial smoothing of 2 mm FWHM, to be consistent with the 2 mm FWHM preprocessing used for MVPA (mean number of voxels in EBA: 198 ± 2.85 *SD*; in FFA:

198.45 ± 2.35; in PPA: 193.3 ± 18.01; in EVC: 188.1 ± 10.94). For each participant, in each ROI, we estimated 48 multivariate β-patterns for facing dyads and 48 for nonfacing dyads (16 patterns for 3 runs), along with 48 β-patterns for single bodies (16 patterns for 3 runs). β-patterns were run-wise normalized to avoid spurious correlations within runs (49). The SVM classifier (LIBSVM, http://www.csie.ntu.edu.tw/~cjlin/libsvm), run through the CosmoMVPA toolbox (50), was trained to discriminate patterns of eight classes corresponding to the eight single bodies, using six samples per class, and then tested on the 48 patterns of facing dyads using a *one-against-one* approach and voting strategy, as implemented in LIBSVM. In 48 testing iterations, each dyad was classified in one of the eight classes of single bodies. Classification accuracy values were averaged across all iterations. Since the test-item included two bodies, classification of each body could be correct in two out of eight cases; therefore, the chance level was 25%. This train-and-test procedure was repeated with nonfacing dyads. Individual accuracy values form the two classifications (train-and-test on facing dyads and train-and-test on nonfacing dyads) were tested against chance with two separate *t* tests ($\alpha = 0.05$, two-tailed).

Visual recognition experiment

Participants. This experiment involved 16 participants of the fMRI-experiment sample, who volunteered to return to the lab, in addition to five new participants (15 female; mean age 24.10 years ± 2.96 *SD*). Twenty was the required sample size to obtain an effect size $\eta_p^2 = 0.453$ ($\beta = 0.80$, alpha = 0.05; G*Power 3.1), comparable to a previous study (18) that used a similar design.

Stimuli. In addition to the 48 images of body-stimuli (16 single bodies, 16 facing dyads and 16 nonfacing dyads) used for fMRI, the experiment included 48 images of single chairs or pairs of chairs (facing or nonfacing). Chair-stimuli were created from eight grey-scale exemplars of chairs and their flipped version, which were combined in 16 pairs of facing, and 16 pairs of nonfacing chairs. All body- and chair-stimuli were inverted upside down, yielding a total of 192 stimuli presented against a grey background. The same number of masking stimuli was created, consisting of high-contrast Mondrian arrays (11° x 10°) of grey-scale circles (diameter 0.4°-1.8°).

Procedure. Participants sat on a height-adjustable chair, 60 cm away from a computer screen, with their eyes aligned to the center of the screen (17-in. CRT monitor, 1024 x 768 pixel resolution, 85-Hz refresh rate). Stimuli on the screen did not exceed 6° of visual angle. Each trial began with a blank display (200 ms), followed by a fixation cross (500 ms), another blank (200 ms), target stimulus (30 ms), mask (250 ms) and, finally, blank until the participant gave a response. The next trial began after a variable interval between 500 and 1000 ms. For each trial, participants provided a response by pressing one of two keys on a keyboard in front of them (“1” with the index finger for “bodies”, or “2” with the middle finger for “chair”; this stimulus-response mapping was counterbalanced across participants). The task included two identical runs, each containing 32 trials for each of the twelve conditions (upright and inverted single, facing and nonfacing bodies and chairs), presented in random order. Each stimulus appeared twice in a run. Every 32 trials participants were invited to take a break. Two blocks of familiarization preceded the actual experiment. In the first, four stimuli per condition were shown for 250 ms, so that the participant could see the stimuli clearly. In the second, eight stimuli per condition were shown for 30 ms, like in the actual experiment. The instructions for the familiarization blocks were identical to those of the actual experiment. The experiment lasted 40 min. Stimulus presentation and response collection (accuracy and RTs) were controlled with the Psychtoolbox and MATLAB.

Behavioral data analyses. Data from one participant were discarded because the average RTs were more than 2.5 *SD* above the group mean. Mean accuracy (mean proportion of correct responses) and RTs for the remaining 20 participants were used for the analysis of the body inversion effect, in a 3 stimulus x 2 orientation repeated-measures ANOVA. Comparisons between critical conditions were performed with pairwise *t* tests ($\alpha = 0.05$, two-tail). The same analysis was repeated on RTs, after removing trials in which the participant’s response was inaccurate or 2 *SD* away from the individual mean (9.6% of the total number of trials).

Acknowledgements

The authors thank Jean-Remy Hochmann and Moritz Wurm for comments on an earlier version of this manuscript. This work was supported by a European Research Council Starting Grant awarded to L.P. (Project: THEMPO, Grant Agreement 758473).

References

1. Adams RB, Ambady N, Nakayama K, Shimojo S eds. (2010) *The Science of Social Vision* (Oxford University Press) doi:10.1093/acprof:oso/9780195333176.001.0001.

2. Bonatti LL, Frot E, Mehler J (2005) What Face Inversion Does to Infants' Counting Abilities. *Psychol Sci* 16(7):506–510.

3. Downing PE, Bray D, Rogers J, Childs C (2004) Bodies capture attention when nothing is expected. *Cognition* 93(1):B27–B38.

4. New J, Cosmides L, Tooby J (2007) Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc Natl Acad Sci* 104(42):16598–16603.

5. Yin (1969) Looking At Upside-Down Faces 1. *J Exp Psychol* 81(1):141–145.

6. Reed CL, Stone VE, Bozova S, Tanaka J (2003) The body-inversion effect. *Psychol Sci* 14(4):302–308.

7. Maurer D, Le Grand R, Mondloch CJ (2002) The many faces of configural processing. *Trends Cogn Sci* 6(6):255–260.

8. Gobbini MI, Haxby J V. (2007) Neural systems for recognition of familiar faces. *Neuropsychologia* 45(1):32–41.

9. Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17(11):4302–11.

10. Downing PE, et al. (2001) A cortical area selective for visual processing of the human body. *Science* 293(5539):2470–3.

11. Downing PE, Peelen M V. (2011) The role of occipitotemporal body-selective regions in person perception. *Cogn Neurosci* 2(3–4):186–203.

12. Stolier RM, Freeman JB (2016) *The Neuroscience of Social Vision* (Elsevier Inc.) doi:10.1016/B978-0-12-800935-2.00007-5.

13. Quadflieg S, Koldewyn K (2017) The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Ann N Y Acad Sci* 1396:166–182.

14. Powell LJ, Spelke ES (2018) Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition* 170:31–48.

15. Isik L, Koldewyn K, Beeler D, Kanwisher N (2017) Perceiving social interactions in the posterior superior temporal sulcus. *Proc Natl Acad Sci*:201714471.

16. Masson HL, Plas S Van De, Daniels N, Beeck H Op De (2018) NeuroImage The multidimensional representational space of observed socio-affective touch experiences. 175(July 2017):297–314.

17. Walbrin J, Downing P, Koldewyn K (2018) Neural responses to visually observed social interactions. *Neuropsychologia* 112:31–39.

18. Papeo L, Stein T, Soto-Faraco S (2017) The Two-Body Inversion Effect. *Psychol Sci* 28(3):369–379.

19. Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: A defense of functional localizers. *Neuroimage* 30(4):1088–1096.

20. Friston KJ, Penny WD, Glaser DE (2005) Conjunction revisited. *Neuroimage* 25(3):661–667.

21. Aminoff E, Gronau N, Bar M (2007) The Parahippocampal Cortex Mediates Spatial and Nonspatial Associations. *Cereb Cortex* 17(7):1493–1503.

22. Kaiser D, Stein T, Peelen M V. (2014) Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc Natl Acad Sci* 111(30):11217–11222.

23. Wang L, Mruczek REB, Arcaro MJ, Kastner S (2015) Probabilistic Maps of Visual Topography in Human Cortex. *Cereb Cortex* 25(10):3911–3931.

24. Downing PE, Peelen M V., Wiggett AJ, Tew BD (2006) The role of the extrastriate body area in action perception. *Soc Neurosci* 1(1):52–62.

25. Diamond R, Carey S (1986) Why Faces Are and Are Not Special : An Effect of Expertise. 115(2):107–117.

26. Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* 3(2):191–197.

27. DiCarlo JJ, Zoccolan D, Rust NC (2012) How Does the Brain Solve Visual Object Recognition? *Neuron* 73(3):415–434.

28. Haxby J V., et al. (2001) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science (80-)* 293(5539):2425–2430.

29. Baeck A, Wagemans J, Op de Beeck HP (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.

30. Kim JG, Biederman I (2011) Where do objects become scenes? *Cereb Cortex* 21(8):1738–1746.

31. Roberts KL, Humphreys GW (2010) Action relationships concatenate representations of separate objects in the ventral visual system. *Neuroimage* 52(4):1541–1548.

32. Lingnau A, Downing PE (2015) The lateral occipitotemporal cortex in action. *Trends Cogn Sci* 19(5):268–277.

33. Reicher GM (1969) Perceptual recognition as a function of meaningfulness of stimulus material. *J Exp Psychol* 81(2):275–280.

34. Homa D, Haver B, Schwartz T (1976) Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Mem Cognit* 4(2):176–185.

35. van Santen JPH, Jonides J (1978) A replication of the face-superiority effect. *Bull Psychon Soc* 12(5):378–380.

36. Brandman T, Peelen XM V (2017) Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. 37(32):7700–7710.

37. Kok P, Jehee JFM, Lange FP De (2012) Report Less Is More : Expectation Sharpens Representations in the Primary Visual Cortex. 265–270.

38. Farrow TFD, et al. (2011) Higher or lower? The functional anatomy of perceived allocentric social hierarchies. *Neuroimage* 57(4):1552–1560.

39. Kujala M V., Kujala J, Carlson S, Hari R (2012) Dog Experts' Brains Distinguish Socially Relevant Body Postures Similarly in Dogs and Humans. *PLoS One* 7(6):e39145.

40. Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8(2–3):109–114.

41. Fischl B (2012) FreeSurfer. *Neuroimage* 62(2):774–781.

42. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433–436.

43. Stigliani A, Weiner KS, Grill-Spector K (2015) Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *J Neurosci* 35(36):12412–12424.

44. Friston KJ (Karl J., Ashburner J, Kiebel S, Nichols T, Penny WD (2007) *Statistical parametric mapping : the analysis of funtional brain images* (Elsevier/Academic Press).

45. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38(1):95–113.

46. McCarthy P (2018) FSLeyes. doi:10.5281/ZENODO.1887737.

47. Desikan RS, et al. (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31(3):968–980.

48. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. *Neuroimage* 62(2):782–790.

49. Lee S, Kable JW (2018) Simple but robust improvement in multivoxel pattern classification. *PLoS One* 13(11):1–15.

50. Oosterhof NN, Connolly AC, Haxby J V. (2016) CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Front Neuroinform* 10(July):1–27.

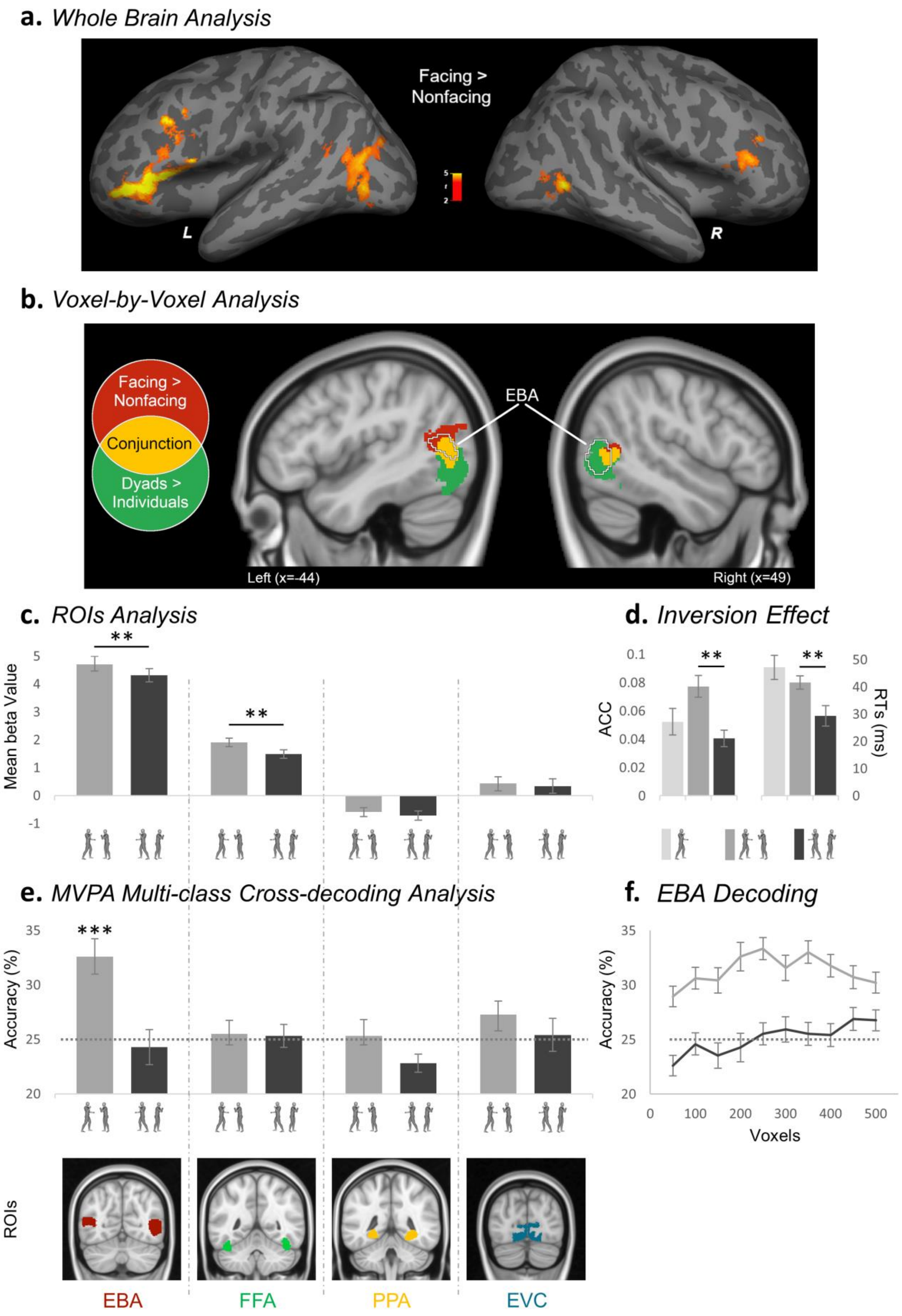
Table 1. Location and significance of clusters showing stronger response to dyads relative to single bodies, and to facing dyads relative to nonfacing dyads. The results are cluster-wise corrected using FWE. Peak coordinates are in MNI space.

Contrasts	**Hemispheres**	**Locations**	**Peak coordinates**			***t***	**Cluster-Wise FWE**	**Cluster size**
			x	**y**	**z**			
Dyads > Individuals	Right	Inferior Temporal Gyrus	47	-71	-3	7.82	<0.001	3232
	Left	Middle Temporal Gyrus	-51	-74	8	6.21	<0.001	2348
Facing > Nonfacing	Right	Middle Frontal Gyrus	-41	44	-5	7.59	<0.001	3307
		Middle Occipital Gyrus	-41	-72	2	4.75	<0.001	1335
	Left	Middle Frontal Gyrus	45	33	24	4.87	0.004	805
		Middle Temporal Gyrus	48	-59	6	5.11	0.011	652

Figure Caption

Fig. 1. Increased activity for facing vs. nonfacing dyads in the EBA and FFA, and representational sharpening in the EBA**.** **(a)** Left (L) and right (R) group random-effect map (N = 20) for the contrast facing dyads > nonfacing dyads. The color bar indicates *t* values. **(b)** Conjunction of the group random-effect maps of the contrasts facing dyads > nonfacing dyads (effect of relation) and dyads > single bodies (effect of number), in the bilateral LOC. Highlighted in red are the voxels showing significant effect of relation only; highlighted in yellow are the voxels showing the effect of both relation and number (conjunction); highlighted in green are the voxels showing the effect of number only. The EBA location corresponds to the group random-effect contrast bodies > objects using the data of the functional localizer task. (**c**) Average beta values (± within-subjects normalized *SEM*) across participants, in each individually defined ROI (EBA, FFA, PPA and EVC), in response to facing and nonfacing dyads. $^{**}p \leq 0.01$. **(d)** Inversion effect (proportion of correct responses with upright minus inverted trials ± within-subjects normalized *SEM*) for accuracy (ACC) and response time (RTs) as a function of the stimulus: single body, facing dyad and nonfacing dyad. $^{**}p \leq 0.01$. **(e)** Classification accuracies (± within-subjects normalized *SEM*) for multi-class cross-decoding of patterns for single bodies in facing dyads and nonfacing dyads, in each individually defined ROI (EBA, FFA and PPA and EVC). Horizontal grey bar represents the chance level (25%). Asterisks indicate accuracy of classification significantly above chance. *** *p < .001*. For illustration purposes, ROIs correspond to the results of the group-level random-effect contrasts of bodies > objects for the EBA, faces > objects for the FFA, and places > objects for the PPA and the probabilistic map of EVC. **(f)** Classification accuracies for multi-class cross-decoding in the EBA-ROI using voxels counts from 50 to 500 voxels.

Figure 1


a. Whole Brain Analysis
Facing > Nonfacing
5
2
L
R
b. Voxel-by-Voxel Analysis
Facing > Nonfacing
Conjunction
Dyads > Individuals
EBA
Left (x=-44)
Right (x=49)
c. ROIs Analysis
**
**
Mean beta Value
5
4
3
2
1
0
-1
d. Inversion Effect
**
**
ACC
0.1
0.08
0.06
0.04
0.02
0
RTs (ms)
50
40
30
20
10
0
e. MVPA Multi-class Cross-decoding Analysis
***
Accuracy (%)
35
30
25
20
f. EBA Decoding
Accuracy (%)
35
30
25
20
0
100
200
300
400
500
Voxels
ROIs
EBA
FFA
PPA
EVC