Category-selective representation of relationships in visual cortex

Liuba Papeo, Etienne Abassi

Category-selective representation of relationships in visual cortex

Etienne Abassi[1] & Liuba Papeo[1]

[1]Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, Centre National de la Recherche Scientifique (CNRS) and Université Claude Bernard Lyon 1, 67 Bd. Pinel – 69675 Bron (France)

Correspondence to:

E.A. etienne.abassi@gmail.com

L.P. liuba.papeo@isc.cnrs.fr

Competing interests: All authors declare that they have no financial or non-financial competing interests.

Abstract

Understanding social interaction requires processing social agents and their relationship. Latest results show that much of this process is visually solved: visual areas can represent multiple people encoding emergent information about their interaction that is not explained by the response to the individuals alone. A neural signature of this process is an increased response in visual areas, to face-to-face (seemingly interacting) people, relative to people presented as unrelated (back-to-back). This effect highlighted a network of visual areas for representing relational information. *How is this network organized?* Using functional MRI, we measured brain activity of healthy female and male humans (*N*=42), in response to images of two faces or two (head-blurred) bodies, facing toward or away from each other. Taking the *facing>non-facing* effect as signature of relation perception, we found that relations between faces and between bodies were coded in distinct areas, mirroring the categorical representation of faces and bodies in visual cortex. Additional analyses suggest the existence of a third network encoding relations between (non-social) objects. Finally, a separate occipitotemporal network showed generalization of relational information across body, face and non-social object dyads (multivariate-pattern classification analysis), revealing shared properties of relations across categories. In sum, beyond single entities, visual cortex encodes the relations that bind multiple entities into relationships; it does so in a category-selective fashion, thus respecting a general organizing principle of representation in high-level vision. Visual areas encoding visual relational information can reveal the processing of emergent properties of social (and non-social) interaction which trigger inferential processes.

Keywords: social perception; social interaction; relational information; social brain; visual perception; fMRI

Significance statement

Understanding social interaction requires representing the actors as well as the relation between them. We show that the earliest, rudimentary representation of a social interaction is formed in visual cortex. Using fMRI on healthy adults, we measured the brain responses to two faces or two (head-blurred) bodies, and found that, beyond representing faces and bodies, the visual cortex represents their relations, distinguishing between seemingly interacting (face-to-face) and non-interacting (back-to-back) faces/bodies. Moreover, we found that information about face and body relations is represented in separate networks, in line with the general organizing principle of categorical representation in visual cortex. The brain network encoding visual relational information may represent emergent properties of interacting people, which underlie the cognitive representation of social interaction.

Introduction

The core meaning of a social interaction is not so much in the individual entities involved in the interaction, as in the relationship holding between them. Much is known about how representation of social agents comes about, starting with the visual representation of faces and bodies. Much less is known about the representation of higher-level relations such as face-to-face social interaction, which bind people together in scenes/events, but are not explained by the information about individuals alone, or exhausted by low-level relational properties of stimuli (proximity, similarity, contours, symmetry, etc.).

Latest research suggests that the representation of social interaction is grounded in visual perception. A neural signature of *social interaction perception* is an increased response in visual areas, to two people face-to-face, relative to the same people presented as independent (Walbrin and Koldewyn, 2019; Abassi and Papeo, 2020; Bellot et al., 2021). This univariate *facing* > *non-facing* effect is intriguing as it denotes not just discrimination between two stimuli, but additional integrative processing emerging from the representation of seemingly related (i.e., facing) individuals (see Adibpour et al., 2021; Goupil et al., 2023).

In those previous studies, the body was treated as a "monolithic" structure with all the parts oriented in the same direction, providing consistent information about an individual's relation with another (facing or not). In a body, however, different effectors can indicate an individual's direction of social attention and interaction (Emery, 2000; Nummenmaa and Calder, 2009): Depending on the context, individuals can rely on gaze, face or rest-of-the-body orientation, independently.

Here, we considered two of such effectors, face/head (hereafter, face) and rest-of-the-body (hereafter, body), and addressed whether spatial relations between faces and relations between bodies are encoded in different aspects of the visual cortex. A categorical organization of relational information would match a general organizing principle in visual cortex, which emerges from its anatomical and functional properties, as extensively studied in object representation (Grill-Spector and Weiner, 2014; Op de Beeck et al., 2019). Such organization could also be behaviorally relevant, as different effectors of interindividual relationships can provide different, even conflicting, information: one can look at another while walking in opposite direction!

Using functional MRI (fMRI), we measured brain activity in healthy human adults, in response to images of two human faces or two (head-blurred) human bodies, facing toward or away from each other. Building on previous results, we used the increased response to facing *versus* non-facing dyads to identify areas for spatial relation perception across the whole brain, and in a set of functionally-defined visual areas, specialized in face- or body-perception (Kanwisher et al., 1997; Downing et al., 2001; Puce et al., 1997; Schwarzlose et al., 2005) and previously implicated in encoding relations between people (Walbrin and Koldewyn, 2019; Abassi and Papeo, 2020, 2022). In a subgroup of subjects, category-selectivity was further investigated measuring brain responses to pairs of non-social objects with anteroposterior morphology (i.e., machines), presented face-to-face or back-to-back. Since we found evidence for a category-selective organization in visual cortex, with segregated representation of relations between faces and between bodies, we performed two additional analyses. First, we tested to which extent relational information for a category (e.g., facing *versus* non-facing faces) overlapped with the visual representation of the corresponding category (e.g., faces *versus* other objects). Second, we tested whether, besides category-selective

areas, there were also visual areas that encoded relations (facing/non-facing) in a stimulus-independent manner, generalizing across dyads of bodies and faces, as well as objects.

In sum, this work examines a division of the visual cortex sensitive to socially relevant spatial relations between people, its internal organization, and its relationship with the face- and body-perception visual networks. In doing so, it lays the groundwork for investigating a new function of vision, and a new visual brain network, which can reveal the earliest perception-based stages towards understanding of social interaction.

Materials and Methods

Participants. A total of 42 subjects took part in the main study (24 females; mean age 24.9 years ± 3.3 *SD*). Among them, a subgroup of 20 was tested with body and face dyads only, and a subgroup of 22 was tested with body and face dyads and, in addition, pairs of non-social objects. All subjects 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 contraindications to fMRI and gave written informed consent before participation. All procedures were approved by the local ethics committee (CPP Sud Est V, CHU de Grenoble).

Stimuli. Using Daz3D (Daz Productions, Salt Lake City) and the Image Processing Toolbox of MATLAB (The MathWorks Inc, Natick, Massachusetts), we created and edited sixteen grayscale renderings of human bodies in profile view, with blurred heads and in various biomechanically possible poses, and sixteen grayscale renderings of human faces in profile view and in various poses and expressions. Additionally, sixteen grayscale renderings of non-social artificial objects (eight chairs, four automated teller machine, and four game machines (i.e., slot machines or arcade game machines) were imported from the Turbosquid database (*www.turbosquid.com*) and edited with the Image Processing Toolbox of MATLAB. All stimuli were matched for luminance and contrast using the SHINE toolbox (Willenbockel et al., 2010). From the sixteen single bodies, eight facing body dyads were created, with each single body appearing only in one pair. Non-facing body dyads were created by swapping the position of the two bodies in each facing dyad. Likewise, eight facing face dyads were created from the sixteen single faces, with each single face appearing only in one pair. Non-facing face dyads were created by swapping the position of the two faces in each facing dyad. Finally, four pairs of machines and four pairs of chairs positioned face-to-face were created from the eight machines and eight chairs, with each single object appearing only in one pair, yielding a total of eight pairs of facing objects. Non-facing objects dyads were created by swapping the position of the two objects in each facing dyad. A mirror version of each dyad was obtained by flipping horizontally each dyad, which yielded a total of 32 body dyads, 32 face dyads and 32 non-social object dyads (for each category, 16 facing and 16 non-facing). The distance between the two terms in a dyad (i.e., between the two closest points) was identical (47 pixels) for all three categories of stimuli (bodies, faces and non-social objects) and for facing and non-facing stimuli.

Procedures

fMRI data acquisition. Imaging was conducted on a MAGNETOM Prisma 3T scanner (Siemens Healthcare) run by CERMEP at Primage (Bron, France). T2*-weighted functional volumes were acquired using a gradient-echo echo-planar imaging sequence (GRE-EPI; TR/TE = 2000/30 ms, flip angle = 80°, acquisition matrix = 96 x 92, FOV = 210 x 201, 56 transverse slices, slice thickness

= 2.2mm, no gap, multiband acceleration factor = 2 and phase encoding set to anterior/posterior direction). Acquisition of high-resolution T1-weighted anatomical images was performed in the middle of the main experiment and lasted 8 min (MPRAGE; TR/TE/TI = 3000/3.7/1100 ms, flip angle = 8°, acquisition matrix = 320 x 280, FOV = 256 x 224 mm, slice thickness = 0.8mm, 224 sagittal slices, GRAPPA accelerator factor = 2). Acquisition of two field maps was performed at the beginning of the fMRI session.

Main experiment. The experiment consisted of two parts: the main fMRI experiment, and a functional localizer task, which we describe below.

Runs. In the main experiment, 20 of the 42 subjects saw three runs with (facing and non-facing) body dyads and three runs with (facing and non-facing) face dyads. Each run lasted 5.23 min and consisted of two sequences of 16 blocks (8 facing and 8 non-facing), for a total of 32 blocks of 6 s each, and a total duration of the experiment of 31.38 min. Runs with body dyads and runs with face dyads 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. Blocks in the first sequence were presented in random order, and blocks in the second sequence were presented in counterbalanced (i.e., reversed) order relative to the first sequence. Thus, blocks that were presented at the end of the first sequence were shown at the beginning of the second sequence, and *vice versa*. 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. For the remaining 22 subjects, everything was identical to the above, except that three runs of non-social object pairs were presented in addition to three runs of body dyads and three runs of face dyads. In order to limit the duration of the scanning session while increasing the number of stimuli, the duration of a run was reduced to 4.17 min by shortening the block duration (see below), and the experiment lasted 37.53 min.

Blocks. For the 20 subjects who saw body and face dyads only, each block featured five repetitions of the same dyad, randomly alternating between one view and its flipped version. Within a run, the onset time of each block was jittered (range of inter-block interval duration: 2-6 s; total inter-block time for each run: 96 s) to remove the overlap from the estimate of the hemodynamic response (Dale, 1999). Jittering was optimized using the optseq tool of Freesurfer (Fischl, 2012). During each block, a black cross was always present in the center of the screen, while stimuli appeared for 500 ms, separated by an interval of 875 ms. For the remaining 22 subjects who also saw non-social object pairs, in order to limit the duration of the scanning session while increasing the number of stimuli, the duration of each block was reduced to 4 s, by modifying its structure (four repetitions of a stimulus, each shown for 520 ms, separated by an interval of 640 ms).

Task. For both groups, in a subset (37.5.%) of stimulus and fixation blocks, the cross changed color (from black to red). Subjects were instructed to fixate the cross throughout the experiment and detect and report the color change by pressing a button with their right index finger. This task was used to minimize eye movements away from the center and maintain vigilance in the scanner. During fMRI acquisition, stimuli were back projected onto a screen by a liquid crystal projector (frame rate: 60 Hz; screen resolution: 1024x768 pixels, screen size: 40x30 cm). For all the stimuli, the center of the image overlapped with the center of the screen. Subjects, lying down inside the scanner, viewed the stimuli binocularly (~7° of visual angle) through a mirror above their head. Stimulus presentation, response collection and synchronization with the scanner were controlled with the Psychtoolbox (Brainard, 1997) through MATLAB (The MathWorks Inc, Natick, MA).

Functional localizer task. All subjects completed a functional localizer task prior to the main experiment, with stimuli and task adapted from the fLoc package (Stigliani et al., 2015). During this task, subjects saw 180 grayscale photographs of the following five object-categories: 1) body-stimuli (headless bodies in various views and poses, and body parts); 2) faces (adults and children); 3) places (houses and corridors); 4) inanimate objects (various exemplars of cars and musical instruments); and 5) scrambled objects. Stimuli were presented over two runs (5.27 min each). Each run began with a warm-up block (12 s) and ended with a cool-down block (16 s), and included 72 blocks of four seconds each: 12 blocks for each object class with eight images per block (500 ms per image without interruption), randomly interleaved with 12 baseline blocks featuring an empty screen. 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. During blocks containing images, some images were repeated twice, interleaved by a different image (2-back task). Subjects had to press a button when they detected the repetition.

Preprocessing of fMRI data. Functional images were preprocessed and analyzed using MATLAB (MathWorks), in combination with SPM 12 (Friston et al., 2007) and the CoSMoMVPA toolbox (Oosterhof et al., 2016). The first four volumes of each run were discarded, taking into account initial scanner gradient stabilization (Soares et al., 2016). Preprocessing of the remaining volumes involved despiking (SPMUP; *https://github.com/CPernet/spmup/*), slice time correction, geometric distortions correction using field maps, spatial realignment and motion correction using the first volume of each run as reference. 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) space. The DARTEL method (Ashburner, 2007) was used to create a flow field for each subject and an inter-subject template, which was registered in the MNI space and used for normalization of functional images. Final steps included spatial smoothing with a Gaussian kernel of 6 mm FWHM for univariate analyses and 2 mm FWHM for multivariate analyses, and removing low-frequency drifts with a temporal high-pass filter (cutoff 128 s).

Analyses

Whole-brain univariate analysis. Considering the runs with face and body dyads in the main experiment (N=42), the blood-oxygen-level-dependent (BOLD) signal of each voxel in each subject, was estimated in two random-effects general linear model (RFX GLM) analyses, each including two regressors for the experimental conditions (facing and non-facing body dyads, or facing and non-facing face dyads), one regressor for fixation blocks, and six regressors for movement correction parameters as nuisance covariates. Separately for bodies and faces, we ran the RFX GLM contrasts facing > non-facing and non-facing > facing. Data of the two groups (group 1 with face and body stimuli and group 2 with face, body and non-social stimuli) were normalized separately using a min-max normalization (Ahammed et al., 2021) to control for changes in signal intensity due to the different designs, and then analyzed together. With the same methods, we analyzed the runs involving non-social object pairs, presented to a subgroup of 22 participants. For all whole-brain analyses, statistical significance of second-level effects was determined using a voxelwise threshold of $p \leq 0.001$, and correction for multiple comparisons with false discovery rate (FDR) at the cluster level, taking into account all preprocessing steps, including smoothing, as implemented in SPM 12 (Friston et al., 2007).

Definition of regions of interest (ROIs). We used data from the functional localizer task to define, for each subject, the following visual areas: extrastriate body area (EBA), fusiform body area (FBA), fusiform face area (FFA), occipital face area (OFA), the object-selective lateral occipital cortex (obj-LOC), and the place-specific parahippocampal place area (PPA). The first four ROIs are selective to body (EBA and FBA; Downing et al., 2001; Taylor et al., 2007) or face perception (FFA and OFA; Kanwisher et al., 1997; Pitcher et al., 2011), and have been also implicated in processing multiple-person scenarios (Abassi and Papeo, 2020, 2022), thus appearing to be plausible candidates in processing relations between faces/bodies. The other two ROIs (obj-LOC and PPA) were included as control areas, as they are involved in generally processing objects but with no selectivity for faces and bodies. To define those ROIs, individual data were entered into a General Linear Model with five regressors for the five object-class conditions (bodies, faces, places, objects and scrambled objects), one regressor for baseline blocks, and six regressors for movement correction parameters as nuisance covariates. Four bilateral masks of the middle occipito-temporal cortex (MOTC), the inferior occipital cortex (IOC), the occipito-temporal fusiform cortex (OTFC) and the inferior parahippocampal cortex (PHC) were created using FSLeyes (McCarthy, 2018) and the Harvard-Oxford Atlas (Desikan et al., 2006) through FSL (Jenkinson et al., 2012). For each subject, within each mask, we selected the voxels with significant activity (threshold: $p = 0.05$) for the contrasts of interest: EBA in the MOTC with the contrast bodies>[objects+faces+places], FBA in the OTFC with the contrast bodies>[objects+faces+places], OFA in the IOC with the contrast faces>[objects+bodies+places], FFA in the OTFC with the contrast faces>[objects+bodies+places], obj-LOC in the MOTC with constrast objects>[bodies+faces+places] and PPA in the PHC with the contrast places>[objects+faces+bodies]. For each ROI, all the voxels within the bilateral mask that passed the threshold were ranked by activation level (*t* values). The final ROI included up to 100 best voxels across the right and left ROI.

ROIs analyses. From the six ROIs of each subject in the main experiment, we extracted the mean neural activity values (mean β-weights minus baseline) for facing and non-facing bodies and facing and non-facing faces and analyzed them in a 2 category (Body, Face) x 2 configuration (Facing, Non-facing) x 6 ROI (EBA, FBA, OFA, FFA, obj-LOC and PPA) repeated-measures ANOVA. We then used *t*-tests (two-tail) for pairwise comparisons addressing the difference between the target effect of relation (facing *vs.* non-facing) for bodies *versus* faces, in each ROI. In a secondary analysis controlling for the selectivity of the effects of faces and bodies, we extracted from the above ROIs the β values (mean β-weights minus baseline) associated with facing and non-facing non-social objects, and used pairwise *t* tests to measure the effect of non-social objects configuration in each ROI. Cohen's *d* was used as a measure of effect size with exact confidence intervals calculated with alpha = 0.05%.

Multivariate searchlight cross-decoding. We used searchlight MVPA analysis (Kriegeskorte et al., 2006) to identify, across the whole brain, areas that encoded relational (facing *vs.* non-facing) information, generalizing across body and face dyads. Separately for each subject, in each sphere of 3-voxel radius centered in each voxel across the brain, we trained a support vector machine classifier (LIBSVM; Chang and Lin, 2011) to discriminate between the β-patterns responding to facing *vs.* non-facing body dyads (48 patterns by condition), and tested the classification accuracy using the β-patterns for facing *vs.* non-facing face dyads (48 patterns by condition), and *vice versa* (train on faces, test on bodies). For each subject, we obtained one map, averaging the two accuracy maps (one for training on bodies and test on faces, the other for training on faces and test on bodies). Individuals' accuracy maps were tested, at the group level, with a one-sample *t*-test against chance

(50 %). Statistical significance of group-level map was determined using a voxel-wise threshold of $p \leq 0.001$ corrected for multiple comparisons using FDR at the cluster level.

Spatial relations between brain networks. We addressed the anatomical relation between the effect of configuration (facing > non-facing) and the activity for face and body perception. In particular, we identified the peak of each effect to test whether the processing of social objects (faces and bodies) and of their relations could recruit dissociable neuronal populations. We tested separately the relation between representation of bodies and their relations, and representation of faces and their relations.

We asked whether for each subject the voxels with the strongest facing>non-facing effect overlapped with the voxels with highest selectivity for single bodies or faces. For each subject, from each anatomical region (MOTC, IOC and OTFC) with significant clusters for the contrast facing>non-facing body dyads, facing>non-facing face dyads, bodies>[objects+faces+places], or faces>[objects+bodies+places], we extracted the best 100 voxels (based on *t* values) using group-constrained, subject-specific localization (Saxe et al., 2006; Fedorenko et al., 2010). More precisely, for effects related to bodies, we created two bilateral masks based on group-level data, one corresponding to the most significant cluster for the contrast facing>non-facing body dyads and one corresponding to the most significant cluster for bodies>[objects+faces+places] in the MOTC; then, for each subject, from each mask, we extracted the best 100 voxels for each contrast, based on *t* values. For effects related to faces, we created bilateral masks based on group-level data, one corresponding to the most significant cluster for the contrast facing>non-facing face dyads and one corresponding to the most significant cluster for the contrast faces>[objects+bodies+places] in the IOC and OTFC; then, for each subject, from each mask, we extracted the best 100 voxels for each contrast, based on *t* values. Finally, for each subject, within each mask, we also computed the number of voxels overlapping for the two contrasts (processing of bodies and their relations and processing of faces and their relations).

While the analysis of individual subjects is more informative to characterize functionally distinct regions of the cortex (Fedorenko et al., 2010; Saxe, Brett & Kanwisher, 2006), we considered group level effects for illustration purposes. On group-level maps, we plotted the voxels with the highest selectivity for the facing>non-facing body dyads, for facing>non-facing face dyads, for single bodies and for single faces. The peak for facing>non-facing body dyads was defined considering the 100 most significant voxels in the group level maps (based on *t* values), within the MOTC (data from main experiment); the peak for the body-selective response was defined considering the 100 most significant voxels in the group level map (based on *t* values) within the MOTC for the contrast bodies>[objects+faces+places] (data from function localizer task). We finally computed the number of overlapping voxels between the two effects. For the relation between processing faces and processing relations between faces, we used the same method as above, but in two different anatomical regions of the visual cortex, the IOC and the OTFC, and using the contrasts facing>non-facing face dyads (data from main experiment) and faces>[objects+bodies+places] (data from the functional localizer task). Finally, the above analysis were repeated to investigate the spatial relationship between voxels with the highest selectivity for the effect facing>non-facing for non-social object dyads, and voxels with the highest selectivity for single non-social objects, within the MOTC, using the contrasts facing>non-facing object dyads (data from main experiment) and objects>[scrambled objects] (data from the functional localizer task).

Results

Higher-level relations are represented in a category-selective fashion in visual cortex. Building on previous research, (Isik et al., 2017; Walbrin and Koldewyn, 2019; Abassi and Papeo, 2020, 2022; Bellot et al., 2021), we used the increased response to facing *versus* non-facing dyads to identify areas sensitive to socially relevant relational information. Going beyond previous research, we tested whether relations between two different effectors of interindividual relationship (faces and bodies) were processed in separate areas of the visual cortex.

For bodies (Fig. 1a; Table 1), the contrast facing > non-facing dyads showed effects in a bilateral cluster peaking in the anterior MOTC and overlapping with the (group-level) body-selective EBA, and in a left cluster in the posterior MOTC (Fig. 1c). The contrast non-facing > facing yielded a cluster in the left cuneus. For faces (Fig. 1b; Table 1), the contrast facing > non-facing dyads showed an effect in a bilateral cluster peaking in the fusiform gyrus and overlapping with the (group-level) face-selective FFA, as well as in a cluster peaking in the posterior MOTC, extending into the IOC and overlapping with the (group-level) face-selective OFA (Fig. 1c). The contrast non-facing > facing faces revealed two clusters peaking in the right lingual gyrus and in the left cuneus.

Contrast	Category	Hemisphere	Peak Location	Peak coordinates			Peak z	Cluster-Wise FDR-corr	Cluster Size
				x	y	z			
Facing > Non-Facing	Bodies	Left	Middle Occipital Gyrus	-44	-72	2	4.81	< 0.001	560
			Middle Occipital Gyrus	-26	-100	-6	4.23	0.018	227
		Right	Middle Temporal Gyrus	52	-70	10	3.96	0.04	170
	Faces	Left	Middle Occipital Gyrus	-28	-100	2	6.65	< 0.001	795
			Fusiform Gyrus	-44	-48	-22	4.41	0.011	117
		Right	Middle Occipital Gyrus	28	-94	4	6.31	< 0.001	1254
			Fusiform Gyrus	44	-44	-22	4.53	0.004	163
	Machines /Chairs	Left	Middle Occipital Gyrus	-28	-89	4	4.34	0.044	195
Non-Facing > Facing	Bodies	Left	Cuneus	-8	-110	12	5.30	0.022	137
	Faces	Left	Cuneus	-12	-94	2	4.81	< 0.001	839
		Right	Lingual Gyrus	18	-90	-16	5.21	< 0.001	919

Table 1. Activations for the whole-brain contrasts. Location and significance of clusters showing stronger responses to facing dyads relative to non-facing dyads and non-facing dyads relative to facing dyads, separately for bodies, faces and non-social objects (machines/chairs). Peak coordinates are in MNI space.

Body- and face-perception ROIs represent relational information in a category-selective fashion. Category-selective organization of relational information in visual cortex was confirmed by the ROI analyses. The individual β values extracted for each condition, from each ROI, were entered in a 2 category (Body, Face) x 2 configuration (Facing, Non-facing) x 6 ROI (EBA, FBA, OFA, FFA, obj-LOC, PPA) repeated-measures ANOVA. Results showed main effects of category, $F(1,41) = 23.60$, $p < 0.001$, $\eta_p^2 = 0.37$, configuration, $F(1,41) = 20.70$, $p < 0.001$, $\eta_p^2 = 0.34$, and ROI, $F(5,205) = 3.45$, $p = 0.005$, $\eta_p^2 = 0.08$, along with significant interactions between category and configuration, $F(1,41) = 6.25$, $p = 0.017$, $\eta_p^2 = 0.13$, category and ROI, $F(5,205) = 65.48$, $p < 0.001$, $\eta_p^2 = 0.61$, and configuration and ROI, $F(5,205) = 11.67$, $p < 0.001$, $\eta_p^2 = 0.22$. All effects were qualified by a significant three-way interaction, $F(5,205) = 16.55$, $p< 0.001$, $\eta_p^2 = 0.29$.

To explain this interaction, pairwise comparisons showed higher activity for facing than non-facing bodies in the EBA and FBA, but not in other ROIs (EBA: $t(41) = 4.22$, $p < 0.001$, $d = 0.65$, $CI = [0.02;0.07]$; FBA: $t(41) = 2.20$, $p = 0.033$, $d = 0.34$, $CI = [0.01;0.03]$; OFA: $t(41) = 0.63$, $p > 0.250$, $d = 0.10$, $CI = [-0.02;0.03]$; FFA: $t(41) = 0.62$, $p > 0.250$, $d = 0.10$, $CI = [-0.02;0.03]$; obj-LOC:

$t(41) = 0.59$, $p > 0.250$, $d = 0.09$, $CI = [-0.02;0.03]$; PPA: $t(41) = 0.38$, $p > 0.250$, $d = 0.06$, $CI = [0.03;0.07]$), and higher activity for facing than non-facing faces in the EBA, FBA, OFA, FFA, and obj-LOC, but not in the PPA (EBA: $t(41) = 3.04$, $p = 0.004$, $d = 0.47$, $CI = [0.01;0.03]$; FBA: $t(41) = 4.40$, $p < 0.001$, $d = 0.68$, $CI = [0.02;0.07]$; OFA: $t(41) = 9.46$, $p < 0.001$, $d = 1.46$, $CI = [0.06;0.10]$; FFA: $t(41) = 7.38$, $p < 0.001$, $d = 1.14$, $CI = [0.07;0.12]$; obj-LOC: $t(41) = 4.93$, $p < 0.001$, $d = 0.76$, $CI = [0.03;0.07]$; PPA: $t(41) = 1.04$, $p > 0.250$, $d = 0.16$, $CI = [-0.04;0.01]$). Crucially, we found that the *facing > non-facing* effect (Fig. 2) was higher for bodies than for faces in the EBA, $t(41) = 2.29$, $p = 0.027$, $d = 0.35$, $CI = [0.00;0.05]$, and for faces than for bodies in the OFA, $t(41) = 5.52$, $p < 0.001$, $d = 0.85$, $CI = [0.05;0.10]$, and FFA, $t(41) = 5.71$, $p < 0.001$, $d = 0.88$, $CI = [0.06;0.12]$, as well as in the obj-LOC, $t(41) = 3.19$, $p = 0.003$, $d = 0.49$, $CI = [0.06;0.12]$. The difference was not significant in the FBA, $t(41) = 1.37$, $p = 0.178$, $d = 0.21$, $CI = [-0.04;0.01]$, and PPA, $t(41) = 0.84$, $p > 0.250$, $d = 0.13$, $CI = [-0.03;0.07]$.

Contribution of low-level visual properties to the facing>non-facing effect. In the following analysis, we sought to shed light on the features that contribute to the facing/non-facing distinction in high-level visual areas. In particular, we quantified to what extent the facing/non-facing distinction relied on low-level differences between the stimuli. To this end, we used a Gabor-jet model (Lades et al., 1993) to measure the V1-like visual properties of each stimulus-image (eight facing and eight non-facing dyads of bodies and faces), represented as a vector (Foster et al., 2022; Yue et al., 2012; Haghighat et al., 2015). Then, separately for each category (bodies and faces), we used Pearson correlations (1–Pearson's *r*) between Gabor-based vectors to create representational dissimilarity matrices (RDM) reflecting low-level differences (i.e., dissimilarities) between stimuli. Upon visual inspection, those Gabor-based RDMs revealed a clear distinction between facing and non-facing pairs for both bodies (fig. 3a) and faces (fig. 3b), suggesting that low-level image information aids facing/non-facing discrimination. To quantify this effect, we tested whether the facing/non-facing distinction in the EBA, OFA, FBA and FFA remained after removing low-level visual information. Separately for bodies and faces, we created a synthetic RDM where cells had value of 0 (for two stimuli of the same type, e.g., facing) or 1 (for two stimuli of a different type). For each subject, for each ROI, we extracted the pattern of activity for each stimulus (eight facing and eight non-facing dyads) and created an RDM using Pearson correlations (1–Pearson's *r*), to measure stimulus dissimilarities in terms of neural patterns. Finally, we computed the correlation (Pearson correlation) between each activity-based RDM (from EBA, OFA, FBA and FFA) and the above synthetic RDM, regressing out similarity based on low-level information (i.e., the Gabor-model RDM was used as regressor). For each ROI, for each subject, we obtained a correlation coefficient *r*. Normalized (fisher-transformed) correlation coefficients, for each ROI separately, were tested against chance with a one-tail *t*-test. Results showed no significant correlation between the activity-based and synthetic RDM after removing low-level visual information, in all the tested ROIs except the EBA (EBA: $t(41) = 4.00$, $p < 0.001$, $d = 0.62$, $CI = [0.03;0.10]$; FBA: $t(41) = 0.15$, $p > 0.250$, $d = 0.02$, $CI = [-0.03;0.03]$; OFA, $t(41) = 1.24$, $p = 0.111$, $d = 0.19$, $CI = [-0.01;0.04]$; FFA: $t(41) = 0.03$, $p > 0.250$, $d = 0.00$, $CI = [-0.03;0.03]$). This means that in higher-level ROIs the facing/non-facing distinction was built –almost exclusively– on the visual properties of the stimuli, in line with the view that information in the physical structure of the stimuli is already relevant to the representation of social interaction (see also McMahon et al., 2023). The fact that the EBA encodes the facing/non-facing distinction beyond the low-level visual properties of the stimuli, suggests functional differences between the ROIs, with EBA serving a distinctive role in the early transformative processes of related bodies (see Abassi & Papeo; 2022; Gandolfo et al., 2023;

Walbrin & Koldewin, 2019). In sum, this analysis showed that the facing/non-facing distinction in higher-level visual areas primarily leverages low-level visual information. What remains to be understood is the precise computation that, based on such distinction, yields a stronger response to facing (vs. non-facing) dyads. This computation, in our framework, would encode emergent properties of related/interacting entities.

An occipitotemporal network encodes spatial relations across face and body dyads. We used searchlight MVPA analyses (Kriegeskorte et al., 2006) to identify areas, across the whole brain, which encoded relational (facing *vs.* non-facing) information, generalizing across body and face dyads. Searchlight MVPA cross-decoding across face and body dyads showed effects in the bilateral medial and lateral occipitotemporal cortex, and in the right inferior occipital gyrus. The effect in the lateral occipitotemporal cortex fell in between the EBA, OFA and obj-LOC, and was partly segregated from areas showing the *facing>non-facing* effect in the MOTC (Fig. 4; Table 2).

Cross-decoding Facing vs. Non-facing	Hemisphere	Peak Location	Peak coordinates			Peak z	Cluster-Wise FDR-corr	Cluster Size
			x	y	z			
Bodies ↔ Faces	Left	Lingual Gyrus	-18	-94	-16	6.16	< 0.001	786
		Middle Occipital Gyrus	-48	-88	4	6.69	< 0.001	434
	Right	Lingual Gyrus	22	-88	-14	5.37	< 0.001	567
		Middle Occipital Gyrus	52	-80	0	4.74	< 0.001	233
		Inferior Occipital Gyrus	44	-70	-16	4.00	< 0.001	29
Persons ↔ Objects	Left	Middle Occipital Gyrus	-48	-84	8	4.31	< 0.001	33
	Right	Lingual Gyrus	16	-92	-6	5.98	< 0.001	1608

Table 2. Searchlight MVPA analyses. Location and significance of clusters showing significant cross-decoding of spatial relations (facing *vs.* non-facing) across dyads of bodies and faces and across dyads of persons and non-social objects. Peak coordinates are in MNI space.

Distinct neuronal populations represent bodies, faces, and their relations. Above we showed that the encoding of socially relevant relations between bodies were computed in a network that involved key body-perception structures (notably in the EBA). Next, we addressed the relationship between the two processes, by asking whether, in each subject, the 100 voxels that showed the highest selectivity to bodies were also those who showed the strongest *facing>non-facing* effect. The analyses, constrained to the bilateral MOTC, revealed largely distinct peaks for the two processes, with no overlap in the right hemisphere and a modest overlap (14.74 voxels ± 11.73 SD) in the left hemisphere (Fig. 6). While the analysis of individual subjects is more informative to characterize functionally distinct regions of the cortex (Saxe et al., 2006; Fedorenko et al., 2010), we considered group level effects for illustration purposes (Fig. 5a). Consistent with the individual-level analysis, the effect at group-level showed two largely distinct peaks for the two processes, with no overlap in the right hemisphere and a modest overlap of 23 voxels in the left hemisphere (Fig. 5a).

The same subject-by-subject analysis considering the face-selective activity and the *facing>non-facing* effect for face dyads in the IOC (Fig. 7), showed largely segregated peaks with minimal overlap in the left hemisphere (1.64 voxels ±1.79), and modest overlap in the right hemisphere (13.81 voxels ±12.53). Likewise, the two effects showed only modest overlap in the fusiform gyrus (left: 18.21 ±14.41; right: 17.52 ±14.61). Consistent with the effects at the individual level, group-level peaks were largely segregated (Fig. 5b) with an overlap of 4 voxels in the left IOC, 16 voxels in the right IOC, 34 voxels in the left fusiform gyrus, and 39 voxels in the right fusiform gyrus.

The visual processing of non-social object pairs. A subgroup of subjects ($N = 22$) saw pairs of non-social objects with clear antero-posterior morphology, facing toward or away from each other, in addition to face and body dyads. We used these data to further investigate category-selective effects

in relation perception; particularly, to study whether the effects described above for faces and bodies could generalize to other visual object categories. The whole-brain contrast facing > non-facing objects yielded a cluster in the left posterior MOTC, overlapping with the obj-LOC, but distinct from areas encoding relations between bodies, and partly segregated from areas encoding relations between faces (Fig. 8a; Table 1). The contrast non-facing > facing objects showed no effects. Furthermore, we found that, like for faces and bodies, the peak (100 best voxels at the group-level) of the *facing>non-facing* effect for object dyads in the left posterior MOTC was spatially segregated from the (group-level) peak of object-selective activity found with the contrast intact > scrambled objects in functional localizer task (Fig. 8c). Note that, for objects, we only performed this analysis at the group-level, as, in our approach, subject-specific peaks were defined within group-constrained activations, which did not overlap at all between the effect of facing > non-facing objects (main experiment) and the effect of intact > scrambled objects (functional localizer task).

The effect of relations between objects was further investigated in each of the category-selective ROIs based on the functional localizer task. Supporting a category-specific representation of relational information, there was no effect of object relation (facing > non-facing) in body- or face-specific ROIs (EBA: $t(21) = 1.21$, $p = 0.119$, $d = 0.26$, $CI = [-0.02;0.06]$; FBA: $t(21) = 0.28$, $p > 0.250$, $d = -0.06$, $CI = [-0.03;0.02]$; OFA: $t(21) = 0.50$, $p > 0.250$, $d = -0.11$, $CI = [-0.05;0.03]$; FFA: $t(21) = 0.85$, $p = 0.204$, $d = -0.18$, $CI = [-0.04;0.02]$), or in PPA ($t(21) = 0.78$, $p = 0.222$, $d = 0.17$, $CI = [-0.02;0.04]$), but a trend in the obj-LOC ($t(21) = 1.41$, $p = 0.086$, $d = 0.30$, $CI = [-0.01;0.05]$) (Fig. 8b).

In summary, while the focus of this study was on social relationships and the analysis on non-social objects lacked statistical power relative to the analysis for body and face dyads, the exploration of the effects associated with non-social objects provided converging evidence for category-selective effects of relational information. In particular, we found that, in none of the face- or body-processing areas, non-social object pairs evoked a significant *facing>non-facing* effect. Moreover, the whole-brain effect, as well as the trend for an effect of relation in the obj-LOC, suggested that the network of visual areas sensitive to relational information may extend beyond the areas identified with body and face dyads, including other meaningful, category-selective subdivisions (see also Kaiser et al., 2014a, 2019; Roberts and Humphreys, 2010; Kim and Biederman, 2011; Baeck et al., 2013; Kubilius et al., 2015).

Finally, since we found a network of areas that encoded relational information across faces and bodies, we carried out a second searchlight MVPA analysis to explore to what extent those areas also encoded relations between non-social objects. To do so, we used the data from the 22 subjects who saw faces, bodies and non-social objects. For each subject, in each sphere of 3-voxel radius centered in each voxel across the brain, we trained a support vector machine classifier (LIBSVM) to discriminate between facing and non-facing *people* and tested the classifier on the facing/non-facing discrimination using the β-patterns associated with (facing and non-facing) object pairs, and *vice versa* (training on objects, test on people). As we had 96 patterns for the people condition (48 body-dyad patterns and 48 face-dyad patterns), and 48 patterns in the *objects* condition, we used half of the body patterns (24) and half of the face patterns (24) for the people condition in a first classification, and the remaining halves, in a second classification. For each classification, we averaged the two accuracy maps obtained with training on people and test on objects, and training on objects and test on people. We then averaged the two maps from the two classification analyses. Individuals' accuracy maps were tested at the group level with a one-sample *t*-test against chance

(50 %). The statistical significance of the group-level map was determined using a voxelwise threshold of $p \leq 0.001$ corrected for multiple comparisons using FDR at the cluster level. The result showed effects in the left medial occipitotemporal cortex and in the bilateral primary visual cortex (fig. 8d; table 2). As shown in Fig. 7d, the effects of cross-decoding of people and objects overlapped with the effects of the cross-decoding of bodies and faces in the more posterior (occipital) aspects. More anterior (occipitotemporal) effects found with the cross-decoding of bodies and faces did not appear in cross-decoding of people and objects. These results suggest that neuronal populations in the occipitotemporal cortex are selectively recruited for processing relations between people or person orientation (see also Foster et al., 2021).

Discussion

Understanding how relational information is computed, where in the brain, and in what format, is a challenge for cognitive and computational neuroscience (Green and Hummel, 2004; Hochmann, 2022; Malik and Isik, 2022) Here we addressed the role of visual cortex in encoding visual relational information *en route* to the representation of social interaction.

Building on previous work, we targeted the *facing>non-facing* effect, under the hypothesis that this effect captures the representation of emergent properties of seemingly related (i.e., facing) bodies (Abassi and Papeo; 2022; Bellot et al., 2021). Results showed that a set of visual areas responded more strongly to facing dyads than to non-facing dyads, in a way that respected a central organizing principle in high-level vision: object category. In particular, with whole-brain analysis, we found that the *facing>non-facing* effect for face and body dyads recruited segregated networks, mirroring the categorical object representation in visual cortex. The effect for body dyads encompassed the bilateral middle occipitotemporal cortex overlapping with the EBA; the effect for face dyads encompassed the bilateral fusiform gyrus and middle occipital gyrus, overlapping with the FFA and OFA. Consistent with the results of whole-brain analyses, ROI analyses showed a double dissociation pattern, with stronger *facing>non-facing* effect for body dyads than face dyads in the body-processing EBA, and stronger *facing>non-facing* effect for face dyads than for body dyads in the face-processing FFA and OFA. Supporting category-selectivity, in none of the face- or body-processing areas, non-social object pairs evoked a significant *facing>non-facing* effect. Additional analyses considering Gabor-filtered image properties, showed that the *facing>non-facing* effect in EBA, FBA and OFA is built starting from a facing/non-facing discrimination based on low-level visual properties of the stimuli. Finally, we found that the peaks of *facing>non-facing* effects were spatially distinct from the activity peaks for face/body (and object) perception. This anatomical shift might explain why relational effects for a category encroached on the territory of another category (i.e., why the *facing>non-facing* face effect was found in face-selective, but also in body- and object-selective ROIs). At the same time, this finding provides important indication for future research, showing a possible limitation of targeting classic object-perception areas (EBA, FFA, obj-LOC, etc.) to study multiple-object perception.

Category-specific and category-general relational effects

Representation of bodies and faces is dissociated in many aspects in visual cortex: distinct areas represent body *versus* face parts, configurations or motion (Atkinson et al., 2012; Kanwisher et al., 1997; Downing et al., 2001; Pitcher et al., 2009; Pitcher et al., 2019). Category selectivity, which is especially marked in the case of faces and bodies, is the result of anatomical and functional features

of neurons in high-level visual areas. Our results showed that, in exploiting the same substrate, the representation of relational information in visual cortex preserves one key function of high-level vision: categorization (Grill-Spector and Weiner, 2014; cf. Konkle and Alvarez, 2022).

Since faces and bodies are independent effectors of interpersonal relationship (Emery, 2000), segregated processing of face- and body-relational information can be behaviorally relevant. At the same time, an architecture that processes interpersonal relationships also needs to combine category-specific information into a whole-body representation. Again, such organization would mirror the organization of face- and body-related information in visual cortex, where face- and body-selective areas coexist with areas that integrate category-selective signals into whole-person representation (Bernstein et al., 2014; Kaiser et al., 2014b; Hu et al., 2020). We found that an occipitotemporal network showed generalization of facing/non-facing dyadic information across body dyads and face dyads. This network, in its anterior part, overlaps with a network reported by Foster et al. (2022), showing shared neural code for face and body orientation (i.e., front, leftward, rightward). These results could come closer to the definition of visual areas that encode spatial relations between *things* with respect to the observer's viewpoint. An open question remains whether these effects reflect abstraction from category-specific signals or low-level visuo-spatial features/processing that are common to facing/non-facing things and are then read by other (higher-order) areas for abstraction and integration (see Hochmann & Papeo, 2022).

Another open question remains what kinds of object-categories and relations can trigger the effects found for face and body dyads. We found the *facing>non-facing* effect for non-social objects, in a cluster adjacent to the obj-LOC, and a trend for the same effect in the functionally localized obj-LOC. Moreover, the occipitotemporal network that classified facing/non-facing relations across faces and bodies, in its posterior part, also did so with non-social object pairs. These observations lend credit to the hypothesis that there are other category-selective divisions of the visual cortex for processing object-object or human-object relations (Baeck et al., 2013; Kaiser et al., 2014a, 2019; Kim and Biederman, 2011; Roberts and Humphreys, 2010; Wurm & Caramazza, 2019), together with common mechanisms for generalization of relational structures across social and non-social interaction events (Karakose-Akbiyik et al., 2023).

Visual versus non-visual effects of spatial relations

By implying interaction, facing people could recruit more attention and/or deeper inferential processing. If so, the *facing>non-facing* effect in visual cortex could reflect top-down signals enhancing the perceptual response to facing dyads, rather than inherently different visual representation of facing *vs.* non-facing dyads. We consider a number of facts that favor the visual origin of the effect. First, we found effects for the contrast facing>non-facing bodies/faces, but also for the contrasts non-facing>facing bodies/faces and facing>non-facing objects (Table 1), which already rule out some kind of general bias for facing people. Second, we designed task and stimuli to single out visuo-spatial processes while minimizing the recruitment of inferential processes and attention towards the stimuli. In particular, attention was balanced across conditions using an orthogonal task (color-change detection on central fixation); accordingly, we found no univariate or multivariate effects in areas that could be specifically related to attentional processing (Posner and Dehaene, 1994; Thiebaut de Schotten et al., 2011). Moreover, while preserving prototypical features of social interaction (*facingness* and spatial proximity), our facing dyads did not give rise to any familiar, meaningful interactions. Accordingly, univariate and multivariate analyses revealed

selective effects in visual cortex, but no activity in higher order temporoparietal areas, associated with the processing of dynamic representations of meaningful social interactions (Centelles et al., 2011; Isik et al., 2017; Walbrin et al., 2018; see also McMahon et al., 2023).

In summary, there is no indication in our results that effects in visual cortex are driven by processes occurring upstream. Our results are rather in line with studies where paradigms tackling visual perception performance, were used to demonstrate more efficient processing of facing (*vs.* non-facing) dyads (Papeo et al., 2017; Papeo, 2020), associated with differences in visual cortex activations (Abassi & Papeo, 2022).

This discussion is not intended to suggest that the processing of relations between people (or objects) begins and ends in the visual cortex. We rather propose that visual areas compute the earliest stages of social-interaction processing, encoding cues of relationship (e.g., *facingness*), which are precursors to, but are not yet a representation of, social interaction (see also Hochmann & Papeo, 2021; McMahon et al., 2023). Spatial positioning (facing/non-facing) is one of such cues but others (e.g., distance, synchrony, coordination) await study. Moreover, by using rich more meaningful and dynamic depictions of social interactions, it should be possible to trigger the whole social-interaction-processing network, revealing where information is channeled beyond visual cortex. Candidate targets of relational information encoded in visual areas are temporo-parietal areas, which appear to be preferentially driven by dynamic stimuli (Bellot et al., 2021; Isik et al., 2017; Pitcher and Ungerleider, 2021; Landsiedel et al., 2022). Another possible target to explore is the attention brain network. As discussed above, we found no evidence that the facing>non-facing effect arises outside of the visual cortex. However, there may be an important role for attention in a process that conceivably has a lot to do with individuals' attention orientation, gaze following and the like. fMRI combined with other methods will reveal feed-forward and feedback dynamics within and between networks that contribute to the understanding of social interaction.

Conclusions

This study shows that, beyond face/body recognition, the visual cortex encodes information that is relevant to construct the representation of meaningful relationships (e.g., a social interaction) between people. It does so in a category-selective fashion, thus respecting a general organizing principle of representation in high-level vision. A separate network of visual areas shows to generalize relational information (facing/non-facing) between body and face dyads, as well as between people and objects. Overall, our results open to the possibility that segregated networks in visual cortex represent objects and their visuo-spatial relations. Understanding the organization and functioning of these previously uncharted divisions of the visual cortex can unravel visual processes that, beyond single face/body recognition, compute emergent properties of complex, multi-person (or multi-object) scenarios, implementing the early stages of compositionality toward the representation of events.

Acknowledgments

This work was supported by a European Research Council Starting Grant to L.P. (Grant number: THEMPO-758473).

References

Abassi E, Papeo L (2020) The representation of two-body shapes in the human visual cortex. Journal of Neuroscience 40:852–863.

Abassi E, Papeo L (2022) Behavioral and neural markers of visual configural processing in social scene perception. Neuroimage:119506.

Adibpour P, Hochmann J-R, Papeo L (2021) Spatial Relations Trigger Visual Binding of People. J Cogn Neurosci 33:1–11.

Ahammed MS, Niu S, Ahmed MR, Dong J, Gao X, Chen Y (2021) DarkASDNet: Classification of ASD on Functional MRI Using Deep Neural Network. Front Neuroinform 15:20.

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

Atkinson AP, Vuong QC, Smithson HE (2012) Modulation of the face- and body-selective visual regions by the motion and emotion of point-light face and body stimuli. Neuroimage 59:1700–1712.

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.

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

Bernstein M, Oron J, Sadeh B, Yovel G (2014) An Integrated Face–Body Representation in the Fusiform Gyrus but Not the Lateral Occipital Cortex. J Cogn Neurosci 26:2469–2478.

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

Chang C, Lin C-J (2011) LIBSVM : A Library for Support Vector Machines. ACM Trans Intell Syst Technol 2:1–27.

Coull, J. T. (2004). fMRI studies of temporal attention: allocating attention within, or towards, time. Cognitive Brain Research, 21(2), 216-226.

Dale AM (1999) Optimal Experimental Design for Event-Related fMRI. 114:109–114.

de Beeck, H. P. O., Pillet, I., & Ritchie, J. B. (2019). Factors determining where category-selective areas emerge in visual cortex. Trends in cognitive sciences, 23(9), 784-797.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980.

Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. Science 293:2470–2473.

Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24:581–604.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175-191.

Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. J Neurophysiol 104:1177–1194.

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

Foster C, Zhao M, Bolkart T, Black MJ, Bartels A, Bülthoff I (2022) The neural coding of face and body orientation in occipitotemporal cortex. Neuroimage 246:118783.

Frankland SM, Greene JD (2015) An architecture for encoding sentence meaning in left mid-superior temporal cortex. Proc Natl Acad Sci U S A 112:11732–11737.

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

Gandolfo, M., Abassi, E., Balgova, E., Downing, P. E., Papeo, L., & Koldewyn, K. (2023). Converging evidence that left extrastriate body area supports visual sensitivity to social interactions. bioRxiv, 2023-05.

Goupil, N., Hochmann, J. R., & Papeo, L. (2023). Intermodulation responses show integration of interacting bodies in a new whole. Cortex, 165, 129-140.

Green C, Hummel JE, Introduction I (2004) Relational Perception and Cognition: Psychol Learn Motiv 44:201–226.

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci 15:536–548.

Hafri A, Firestone C (2021) The Perception of Relations. Trends Cogn Sci:1–18.

Haghighat, M., Zonouz, S., & Abdel-Mottaleb, M. (2015). CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. Expert Systems with Applications, 42(21), 7905-7916.

Hochmann, J. R., & Papeo, L. (2021). How can it be both abstract and perceptual? Comment on Hafri, A., & Firestone, C.(2021), The perception of relations, Trends in Cognitive Sciences.

Hochmann J-R (2022) Representations of Abstract Relations in Infancy. Open Mind:1–20.

Hu Y, Baragchizadeh A, O'Toole AJ (2020) Integrating faces and bodies: Psychological and neural perspectives on whole person perception. Neurosci Biobehav Rev 112:472–486.

Isik L, Koldewyn K, Beeler D, Kanwisher N (2017) Perceiving social interactions in the posterior superior temporal sulcus. Proceedings of the National Academy of Sciences:201714471.

Isik L, Mynick A, Pantazis D, Kanwisher N (2020) The speed of human social interaction perception. Neuroimage 215:116844.

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

Kaiser D, Quek GL, Cichy RM, Peelen M v. (2019) Object Vision in a Structured World. Trends Cogn Sci 23:672–685.

Kaiser D, Stein T, Peelen M v. (2014a) Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. Proceedings of the National Academy of Sciences 111:11217–11222.

Kaiser D, Strnad L, Seidl KN, Kastner S, Peelen M v. (2014b) Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Karakose-Akbiyik, S., Caramazza, A., & Wurm, M. F. (2023). A shared neural code for the physics of actions and object events. Nature Communications, 14(1), 3316.

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

Konkle T, Alvarez GA (2022) A self-supervised domain-general learning framework for human ventral stream representation. Nature Communications 2022 13:1 13:1–12.

Kralik JD, Hauser MD (2002) A nonhuman primate's perception of object relations: experiments on cottontop tamarins, Saguinus oedipus. Anim Behav 63:419–435.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proceedings of the National Academy of Sciences 103:3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in systems neuroscience, 4.

Kubilius J, Baeck A, Wagemans J, op de Beeck HP (2015) Brain-decoding fMRI reveals how wholes relate to the sum of parts. Cortex 72:5–14.

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., Von Der Malsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on computers, 42(3), 300-311.

Landsiedel J, Daughters K, Downing PE, Koldewyn K (2022) The role of motion in the neural representation of social interactions in the posterior temporal cortex. Neuroimage 262:119533.

Malik M, Isik L (2022) Relational Visual Information explains Human Social Inference: A Graph Neural Network model for Social Interaction Recognition.

Markett, S., Reuter, M., Montag, C., Voigt, G., Lachmann, B., Rudorf, S., ... & Weber, B. (2014). Assessing the function of the fronto-parietal attention network: insights from resting-state fMRI and the attentional network test. Human brain mapping, 35(4), 1700-1709.

Masson, H. L., & Isik, L. (2021). Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. NeuroImage, 245, 118741.

McCarthy P (2018) FSLeyes. https://doi.org/10.5281/zenodo.1887737

McMahon, E., Bonner, M. F., & Isik, L. (2023). Hierarchical organization of social action features along the lateral visual pathway. PsyArxiv.

Nummenmaa L, Calder AJ (2009) Neural mechanisms of social attention. Trends Cogn Sci 13:135–143.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision, 42, 145-175.

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

Papeo L (2020) Twos in human visual perception. Cortex.

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

Pitcher D, Charles L, Devlin JT, Walsh V, Duchaine B (2009) Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. Current Biology 19:319–324.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage, 56(4), 2356-2363.

Pitcher D, Ianni G, Ungerleider LG (2019) A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. Scientific Reports 2019 9:1 9:1–9.

Pitcher D, Ungerleider LG (2021) Evidence for a Third Visual Pathway Specialized for Social Perception. Trends Cogn Sci 25:100–110.

Pitcher D, Walsh V, Duchaine B (2011) The role of the occipital face area in the cortical face perception network. Exp Brain Res 209:481–493.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. Trends in neurosciences, 17(2), 75-79.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. Journal of neuroscience, 16(16), 5205-5215.

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

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

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. Journal of Neuroscience, 25(47), 11055-11059.

Soares JM, Magalhães R, Moreira PS, Sousa A, Ganz E, Sampaio A, Alves V, Marques P, Sousa N (2016) A Hitchhiker's Guide to Functional Magnetic Resonance Imaging. Front Neurosci 10:515.

Stigliani A, Weiner KS, Grill-Spector K (2015) Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. Journal of Neuroscience 35:12412–12424.

Stuss DT, Benson DF (1987) The frontal lobes and control of cognition and memory. The Frontal Lobes Revisited:141–158.

Taubert J, Ritchie JB, Ungerleider LG, Baker CI (2021) One object, two networks? Assessing the relationship between the face and body-selective regions in the primate visual system. Brain Struct Funct.

Taylor JC, Wiggett AJ, Downing PE (2007) Functional MRI Analysis of Body and Body Part Representations in the Extrastriate and Fusiform Body Areas. J Neurophysiol 98:1626–1633.

Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S., Simmons, A., Vergani, F., Murphy, D. G., & Catani, M. (2011). A lateralized brain network for visuo-spatial attention. Nature Precedings, 1-1.

Voelkl, B. (2019). Multiple testing: correcting for alpha error inflation with false discovery rate (FDR) or family-wise error rate?. Animal behaviour.

Walbrin J, Koldewyn K (2019) Dyadic interaction processing in the posterior temporal cortex Jon Walbrin. Neuroimage 198:1–26.

Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: The SHINE toolbox. Behavior Research Methods 2010 42:3 42:671–684.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron, 30(3), 829-841.

Wurm MF, Caramazza A (2022) Two 'what' pathways for action and object recognition. Trends Cogn Sci 26:103–116.

Wurm, M. F., & Caramazza, A. (2019). Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract representations of sociality. Neuroimage, 202, 116153.

Wurm MF, Caramazza A, Lingnau A (2017) Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. J Neurosci 37:562–575.

Yue, X., Biederman, I., Mangini, M. C., von der Malsburg, C., & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. Vision research, 55, 41-46.

FIGURES

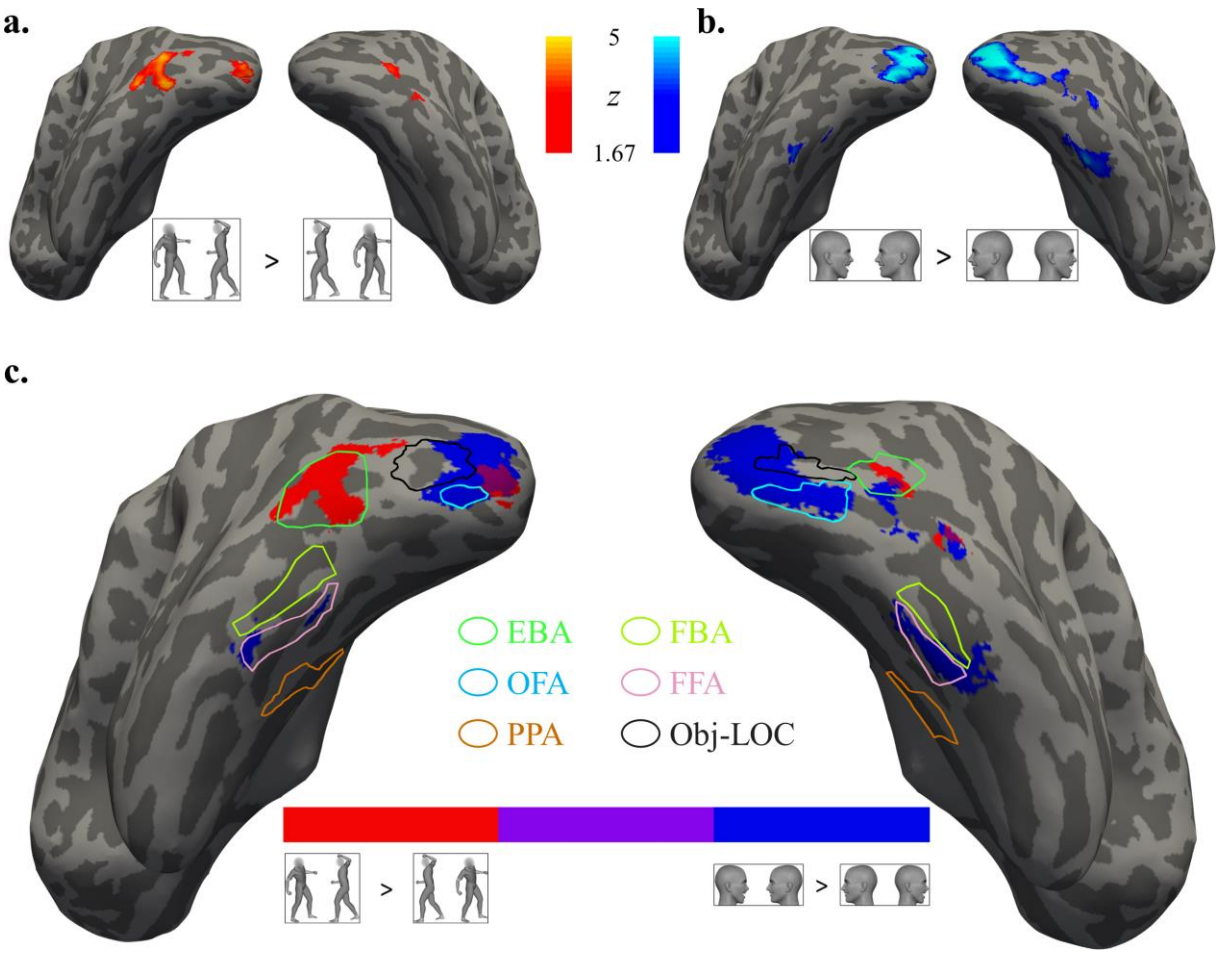

Figure 1. Whole brain univariate analyses for body and face dyads. a. Left and right group random-effect maps (N = 42) for the contrast facing > non-facing body dyads. **b.** Left and right group random-effect maps (N = 42) for the contrast facing > non-facing face dyads. **c.** Overlap (purple) of the clusters for the contrasts facing > non-facing body (red) and face (blue) dyads. Statistical maps are corrected for multiple comparisons using FDR at the cluster level. For illustration purposes, ROIs correspond to the group-level random-effect contrasts (N = 42) of bodies > [objects+faces+places] for the EBA and FBA, faces > [objects+faces+places] for the OFA and FFA, objects > scrambled objects for the obj-LOC and places > [bodies+faces+objects] for the PPA. However, in the actual analyses, we used anatomically-constrained activations in individual subjects, rather than group-level activations.

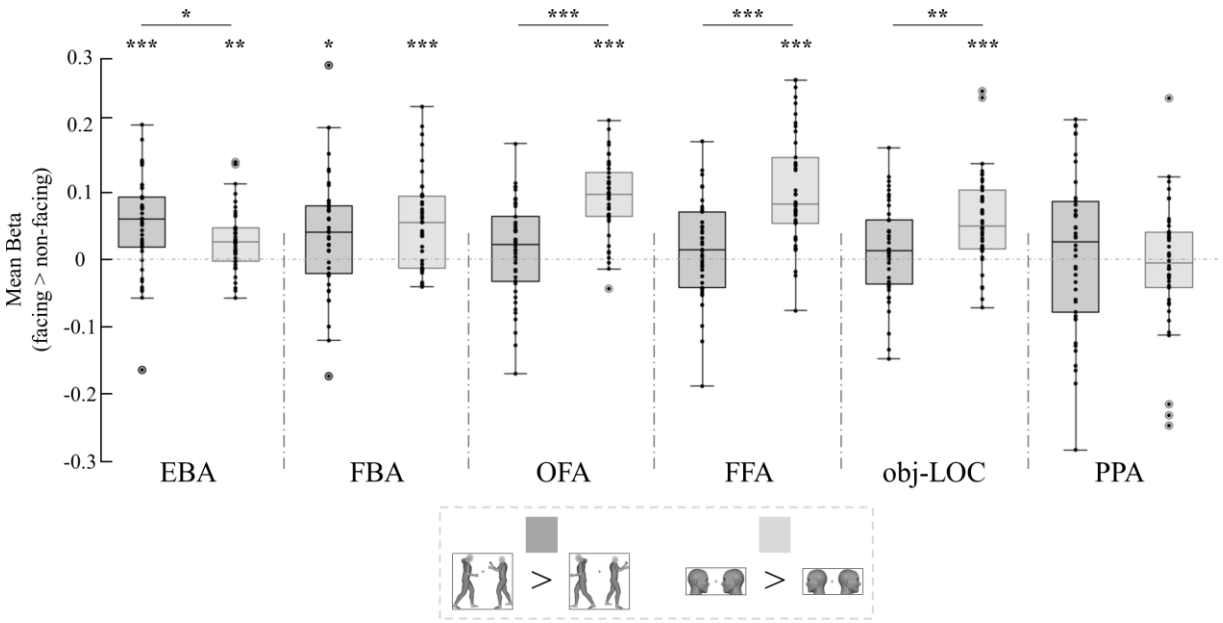

Figure 2. Results of ROIs analyses. The facing > non-facing effect for bodies in dark grey (*N*=42) and faces in light grey (*N*=42) in each ROI. Boxes depict the median, lower quartile and upper quartile while whiskers depict the nonoutlier minimum and maximum. Circles represent outliers and dots represent all subjects datapoints. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

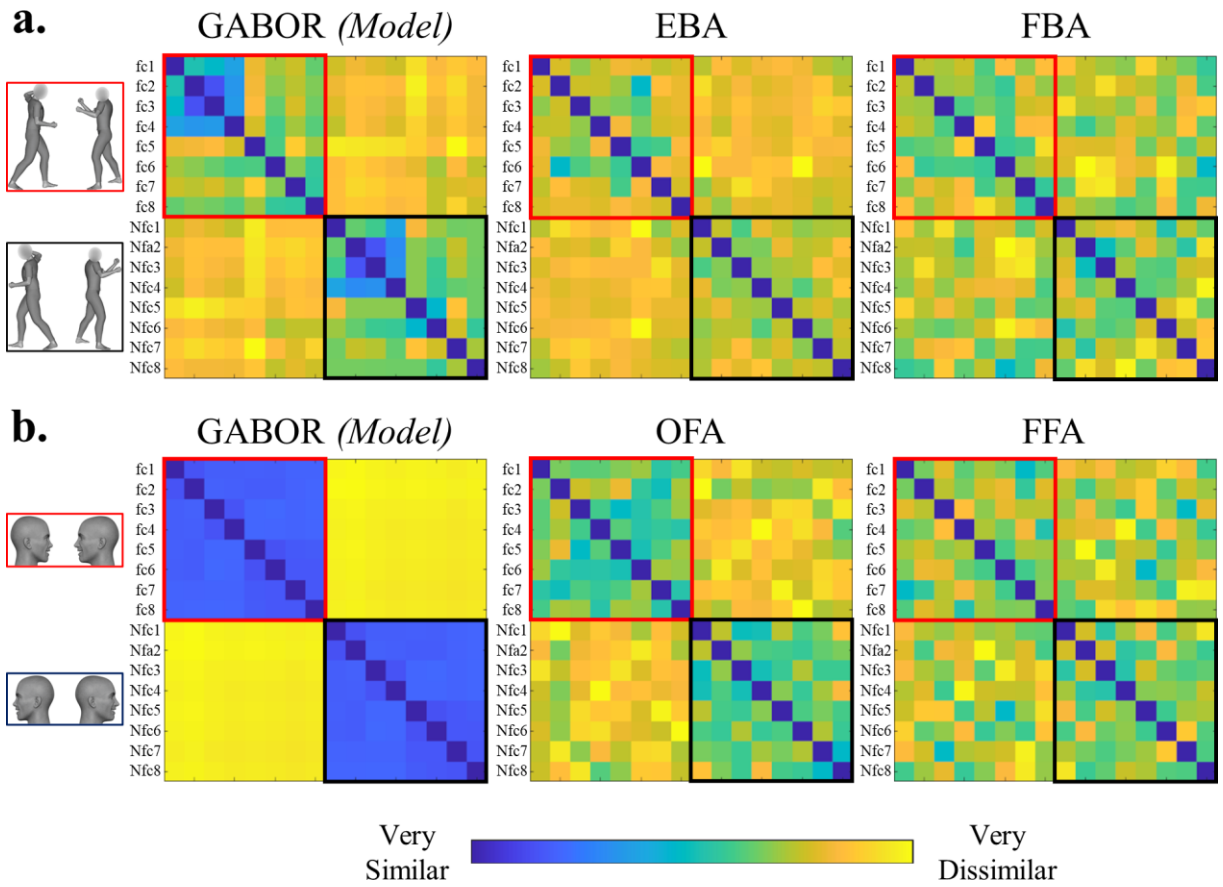

Figure 3. Representational Dissimilarity Matrices (RDMs) representing stimulus dissimilarities based on low-level visual features and neural activity patterns. **(a)** RDMs for body dyads based on V1-like visual features measured with a Gabor-jet model (Gabor model) and based on neural activity patterns extracted from EBA and FBA; **(b)** RDMs for face dyads based on V1-like visual features measured with a Gabor-jet model (Gabor model), and based on neural activity patterns extracted from OFA and FFA. For illustrations purposes, each RDM has been separately rescaled between 0 (very similar) and 2 (very dissimilar) values. The labels fc1 to fc8 correspond to the 8 facing pairs; the labels Nfc1 to Nfc8 correspond to the 8 non-facing pairs.

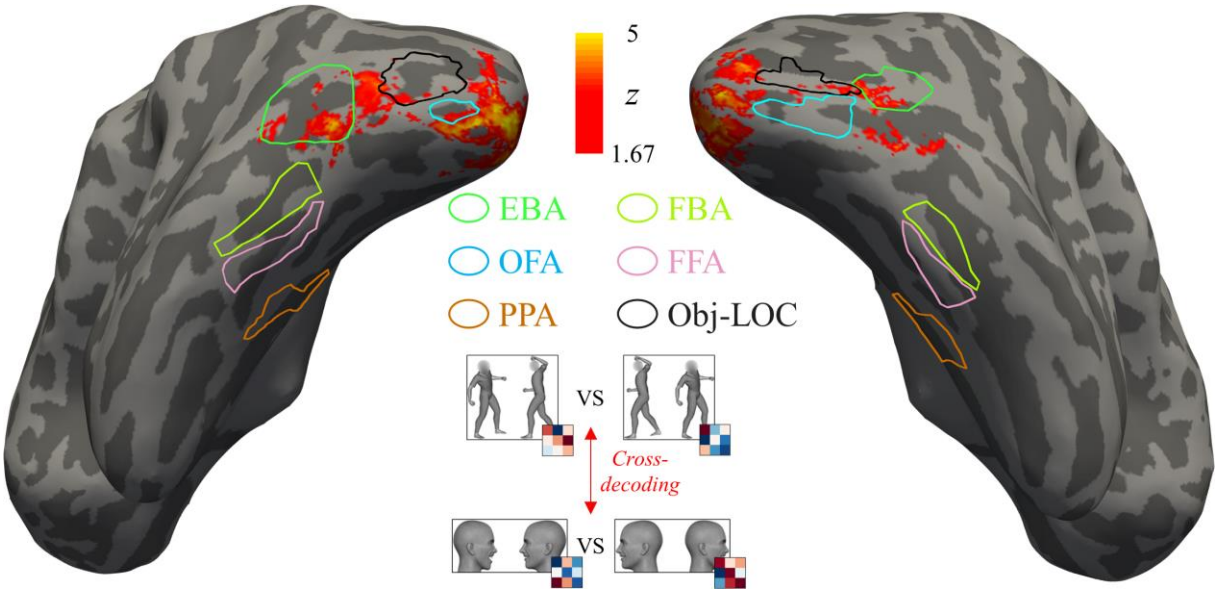

Figure 4. Cross-decoding of spatial relations (facing vs. non-facing) across dyads of bodies and faces. Results of the cross-decoding analysis using searchlight MVPA to identify areas that encode spatial relations across face and body dyads. Areas bounded by colored lines correspond to the ROIs defined with the group-level random-effect contrasts (N = 42), bodies > [objects+faces+places] for the EBA and FBA, faces > [objects+faces+places] for the OFA and FFA, objects > scrambled objects for the obj-LOC, and places > [bodies+faces+objects] for the PPA. Statistical maps are corrected for multiple comparisons using FDR at the cluster level.

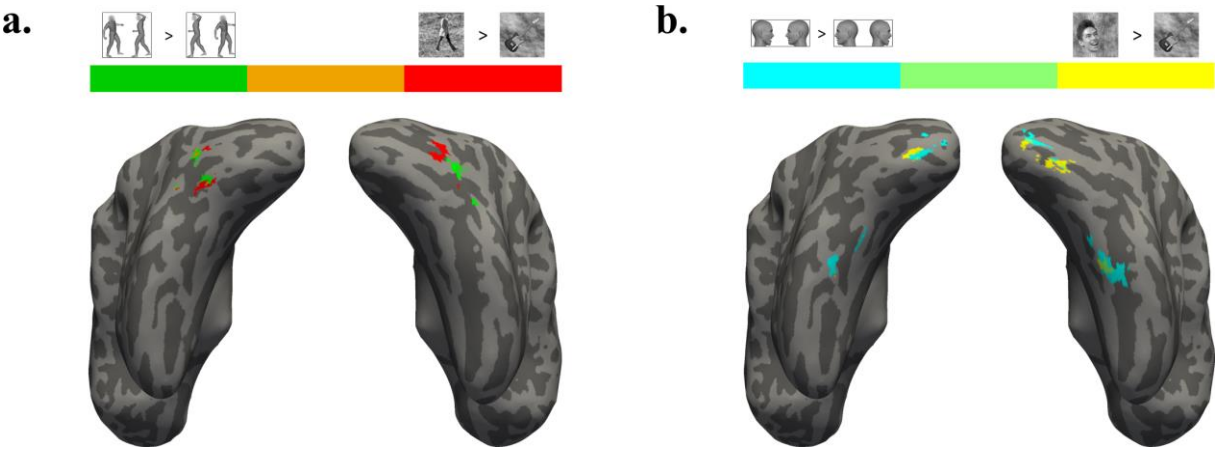

Figure 5. Peaks of (group-level) effects of body/face perception and relation perception. a. Peaks (100 best voxels) for the group-level effects of the contrasts facing > non-facing bodies and bodies > [objects+faces+places]. **b.** Peaks (100 best voxels) for the group-level effects of the contrasts facing > non-facing faces and faces > [objects+bodiess+places].

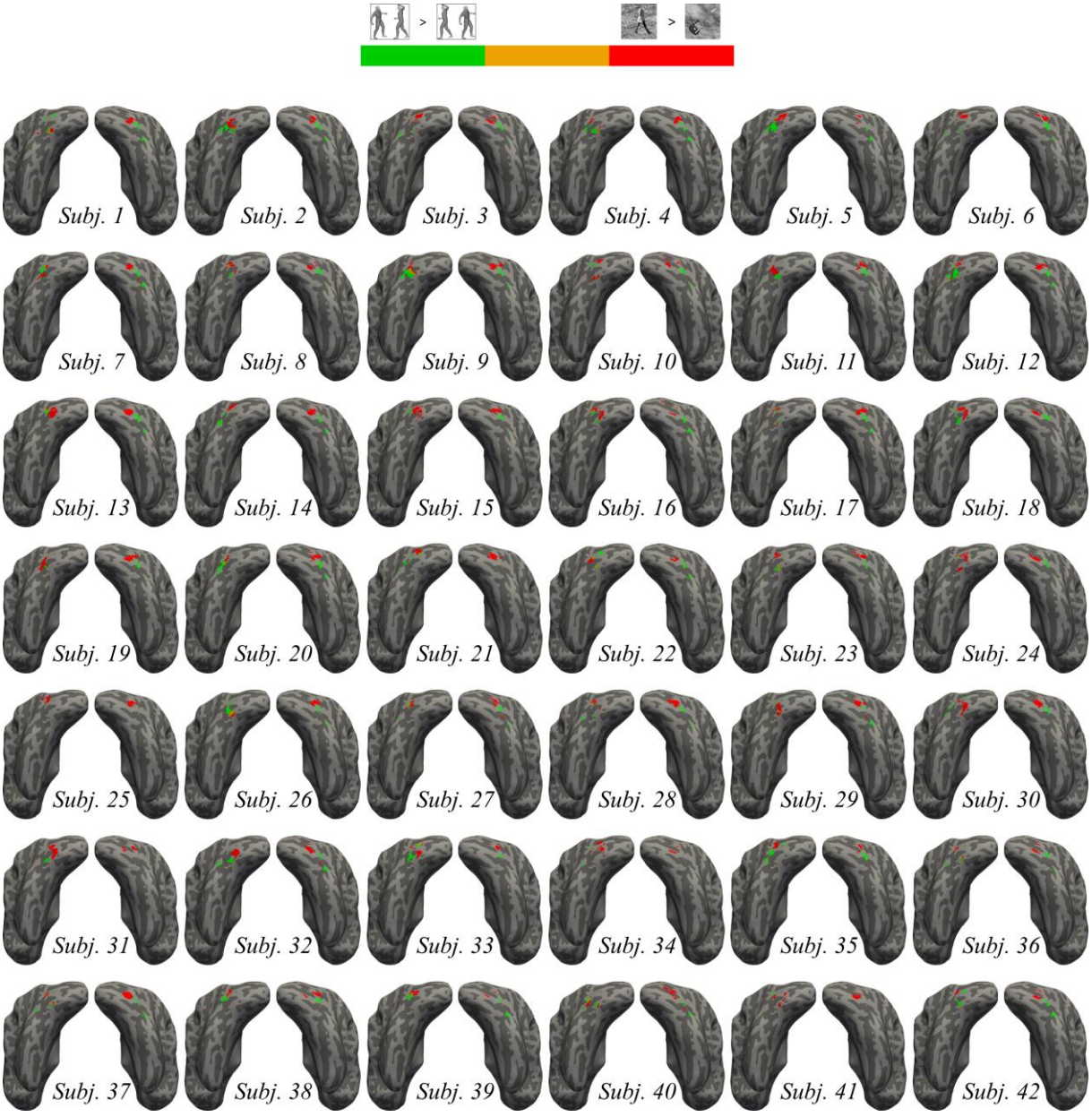

Figure 6. Subject-by-subject peaks of activity for body perception and the facing > non-facing effect for body dyads. For each of the 42 subjects, we show the best 100 voxels for the contrasts facing > non-facing bodies (green) and bodies > [objects+faces+places] (red) constrained within the anatomical middle occipitotemporal cortex. Overlapping voxels are highlighted in orange.

Figure 7. Subject-by-subject peaks of activity for face perception and the facing > non-facing effect for face dyads. For each of the 42 subjects, we show the best 100 voxels for the contrasts facing > non-facing faces (blue) and faces > [objects+bodies+places] (yellow) constrained within the anatomical inferior occipital cortex and fusiform gyrus. Overlapping voxels are highlighted in green.

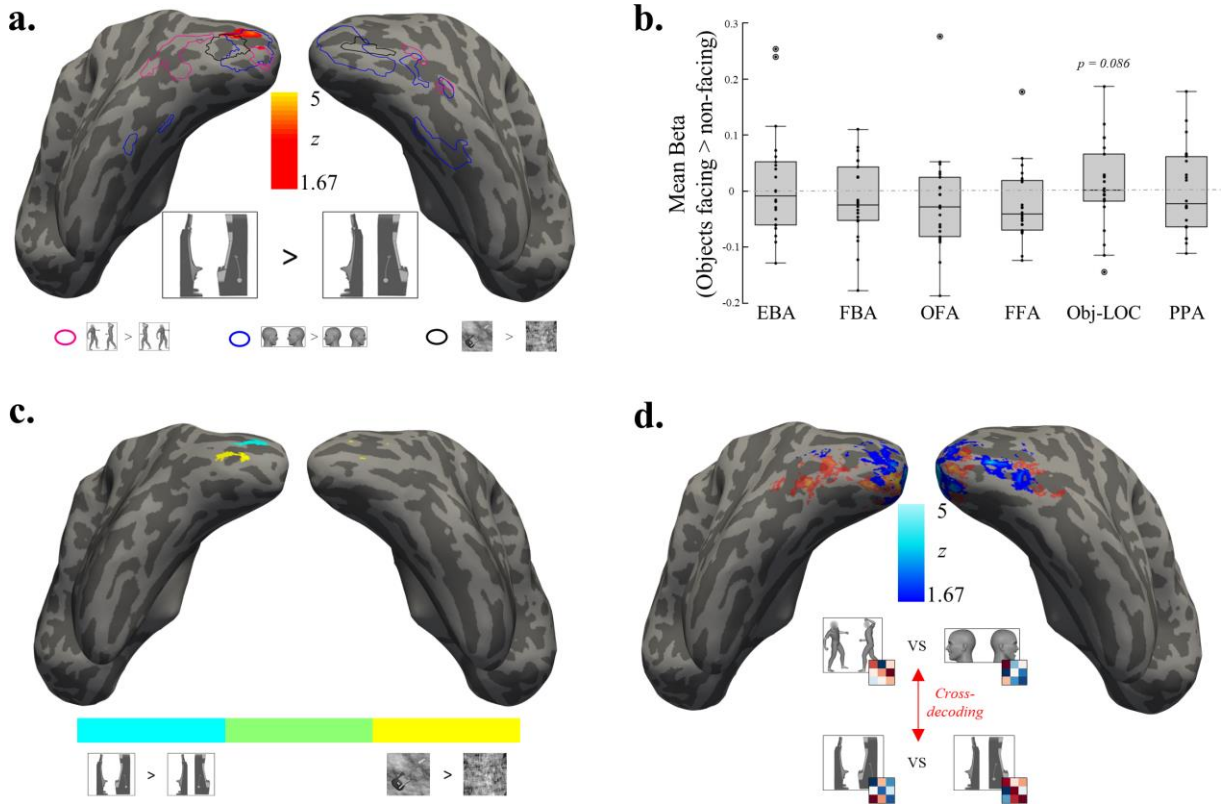

Figure 8. Results for non-social object pairs. a. Whole brain univariate analyses for non-social object dyads. Left and right group random-effect map (N = 22) for the contrast facing > non-facing non-social object dyads (machines and chairs). The results of the group-level (N = 42) random-effect contrasts of facing > non-facing body dyads (in pink), facing > non-facing face dyads (in blue) and objects > scrambled objects for the obj-LOC (in black) are also highlighted. **b. Results of the ROIs analyses.** The *facing>non-facing* effect (N=22) in each functionally-localized ROI. Boxes depict the median, lower quartile and upper quartile while whiskers depict the nonoutlier minimum and maximum. Circles represent outliers and dots represent all subjects datapoints. **c. Peaks of (group-level) effects of object perception and relation perception.** Peaks (100 best voxels) for the group-level effects of the contrasts facing > non-facing object dyads and objects > scrambled objects. **d. *Cross-decoding of spatial relations (facing vs. non-facing) across dyads of people (bodies and faces) and objects*.** Results of the cross-decoding analysis using searchlight MVPA to identify areas that encode spatial relations across people (bodies and faces) and objects. For comparison, areas represented in red-to-yellow are the same as shown in Fig. 4 (cross-decoding of face and body dyads). All statistical maps are corrected for multiple comparisons using FDR at the cluster level.