

Symposium

Neural Encoding of Bodies for Primate Social Perception

 Etienne Abassi,^{1*}  Anna Bognár,^{2,3*}  Bea de Gelder,^{4*}  Martin Giese,^{5*}  Leyla Isik,^{6*}  Alexander Lappe,^{5*}  Albert Mukovskiy,^{5*}  Marta Poyo Solanas,^{4*}  Jessica Taubert,^{7*} and  Rufin Vogels^{2,3*}

¹The Neuro, Montreal Neurological Institute-Hospital, McGill University, Montréal, QC H3A 2B4, Canada, ²Department of Neuroscience, KU Leuven, Leuven 3000, Belgium, ³Leuven Brain Institute, KU Leuven, Leuven 3000, Belgium, ⁴Cognitive Neuroscience, Maastricht University, Maastricht 6229 EV, Netherlands, ⁵Section Computational Sensomotrics, Hertie Institute for Clinical Brain Research & Centre for Integrative Neuroscience, University Clinic Tuebingen, Tuebingen D-72076, Germany, ⁶Cognitive Science, Johns Hopkins University, Baltimore, Maryland 21218, and ⁷The School of Psychology, University of Queensland, St Lucia, QLD 4072, Australia

Primates, as social beings, have evolved complex brain mechanisms to navigate intricate social environments. This review explores the neural bases of body perception in both human and nonhuman primates, emphasizing the processing of social signals conveyed by body postures, movements, and interactions. Early studies identified selective neural responses to body stimuli in macaques, particularly within and ventral to the superior temporal sulcus (STS). These regions, known as body patches, represent visual features that are present in bodies but do not appear to be semantic body detectors. They provide information about posture and viewpoint of the body. Recent research using dynamic stimuli has expanded the understanding of the body-selective network, highlighting its complexity and the interplay between static and dynamic processing. In humans, body-selective areas such as the extrastriate body area (EBA) and fusiform body area (FBA) have been implicated in the perception of bodies and their interactions. Moreover, studies on social interactions reveal that regions in the human STS are also tuned to the perception of dyadic interactions, suggesting a specialized social lateral pathway. Computational work developed models of body recognition and social interaction, providing insights into the underlying neural mechanisms. Despite advances, significant gaps remain in understanding the neural mechanisms of body perception and social interaction. Overall, this review underscores the importance of integrating findings across species to comprehensively understand the neural foundations of body perception and the interaction between computational modeling and neural recording.

Key words: action recognition; body patches; body perception; extrastriate body area; social interactions

Primates are social beings who live in complex cooperative groups (Tomasello, 2014). The primate brain evolved in response to selective pressures exerted by the need to function in social societies (Dunbar, 1998). Pressure from the requirements to deal with complex social situations explains the presence of brain mechanisms specialized in processing social information (Deen et al., 2023). Indeed, much of what primates need to understand about others, their intentions and actions, individually and as a group, is acquired through observing other individuals and their interactions. From early infancy, we know that the interpretation and learning of third-party social interactions play a fundamental role in social, cognitive, and moral development (Rogoff et al., 2003; Hamlin et al., 2007; Tomasello, 2019; Thomas et al., 2022).

While face perception is critical for the recognition of emotional expression and personal identity at close distance, social signals

from the body play a crucial role in nonverbal communication, underlying much of social interaction. Across the primate order, signals provided by body posture and movements promote the recognition of individuals and help discriminate different genders and emotional states. As a result, the primate brain evolved to easily detect conspecific actions (de Gelder, 2006) and interactions (Su et al., 2016; Papeo, 2020; McMahon and Isik, 2023) and quickly unpack their meaning. Despite this importance, in the past the neural processing of body stimuli has received less attention than faces (de Gelder et al., 2010) and our understanding of the behavioral and neural basis of bodily communication is only just starting.

In this brief review, we will first summarize work on the visual representations of single bodies in the macaque and human visual system. Then, we will examine the processing of multiple interacting social agents, followed by an overview of the relevant computational models. Most computational models of social interactions have employed a two-stream framework, in which static and dynamic features are analyzed by separate processing pipelines, which was inspired by the distinction between the form and motion processing streams in the primate visual cortex.

Visual Body Representations in Nonhuman Primates

In macaques, early recordings of neurons in the superior temporal sulcus (STS) and ventral inferotemporal (IT) cortex found selective responses to body parts (Gross et al., 1969, 1972; Bruce et al., 1981).

Received June 28, 2024; revised July 22, 2024; accepted July 23, 2024.

Author contributions: E.A., A.B., B.G., M.G., L.L., A.L., A.M., M.P.S., J.T., and R.V. wrote the paper.

This work was supported by grant National Institutes of Health R01MH132826 (L.L.), IMPRS for Intelligent Systems, Tübingen, Germany (A.L.), Fonds voor Wetenschappelijk Onderzoek (FWO) Vlaanderen G0E0220N (R.V.), KU Leuven grant C14/21/111 (R.V.) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, grant agreement no. 856495 (A.B., B.G., M.G., A.L., M.P.S., R.V.).

*All authors, listed in alphabetical order, contributed equally to this review.

The authors declare no competing financial interests.

Correspondence should be addressed to Rufin Vogels at rufin.vogels@kuleuven.be or Bea de Gelder b.degelder@maastrichtuniversity.nl.

<https://doi.org/10.1523/JNEUROSCI.1221-24.2024>

Copyright © 2024 the authors

Later, using fMRI, body category-selective regions known as body patches were identified (Tsao et al., 2003), organized in a local network (Premereur et al., 2016). Across multiple monkey fMRI studies using static images, at least two body category-selective regions can be reliably identified: (1) the middle STS body patch (MSB; Jastorff et al., 2012; Popivanov et al., 2012) which is situated in the medial part of the ventral bank of the middle STS, neighboring the Middle Lateral face patch (Pinsk et al., 2009; Bao et al., 2020), and, (2) the anterior STS body patch (ASB; Kumar et al., 2019), located in the lateral part of the ventral bank of the STS, with its anterior/posterior location varying between subjects and medial to the Anterior Lateral face patch (Bao et al., 2020). A smaller anterior ventral body (AVB) patch has been reported in ventral IT, dorsolateral to the anterior medial face patch (for review, see Vogels, 2022). These studies relied on static images but since bodily movements play an important role in emotion and action recognition (Giese and Poggio, 2003; de Gelder et al., 2015), body activations might differ for static and dynamic displays, as seen in humans (Grezes et al., 2007). Indeed, by using short videos of macaques, a more extended body-selective network was revealed (Fig. 1a), with additional body patches in the dorsal bank of the STS (Bognár et al., 2023).

Although they are unnatural to some degree, one typically presents bodyless faces and headless bodies as stimuli to isolate

the category-specific networks. Given that faces are usually attached to bodies, researchers have searched for evidence of superadditive activations (i.e., higher responses to the whole agent than the sum of the responses to the body alone and the face alone), and there has been some indication that an anterior dorsal STS face patch prefers whole agents (Fisher and Freiwald, 2015; but see Zafirova et al., 2022). Although anterior IT patches activated by monkey images show stronger responses to natural configurations of faces and bodies than to unnatural ones (Zafirova et al., 2022, 2024), superadditive responses were rare at the single neuron level (Zafirova et al., 2024). The face–body configuration sensitivity was also present when faces were replaced by roundish objects (Zafirova et al., 2024), suggesting that bodies and head-like features, and not necessarily inner facial features, are encoded in a configuration-specific manner. Further evidence for the interaction of face and body processing are responses in posterior face patches to occluded faces when the body is present (Arcaro et al., 2020). These and multiple human fMRI studies (for review see Hu et al., 2020; Taubert et al., 2022b) support interactions between face and body processing. However, other findings argue for a functional distinction between the processing of faces and bodies (Premereur et al., 2016; Waidmann et al., 2022). For example, the sensitivity to affective states and actions might distinguish the processing of

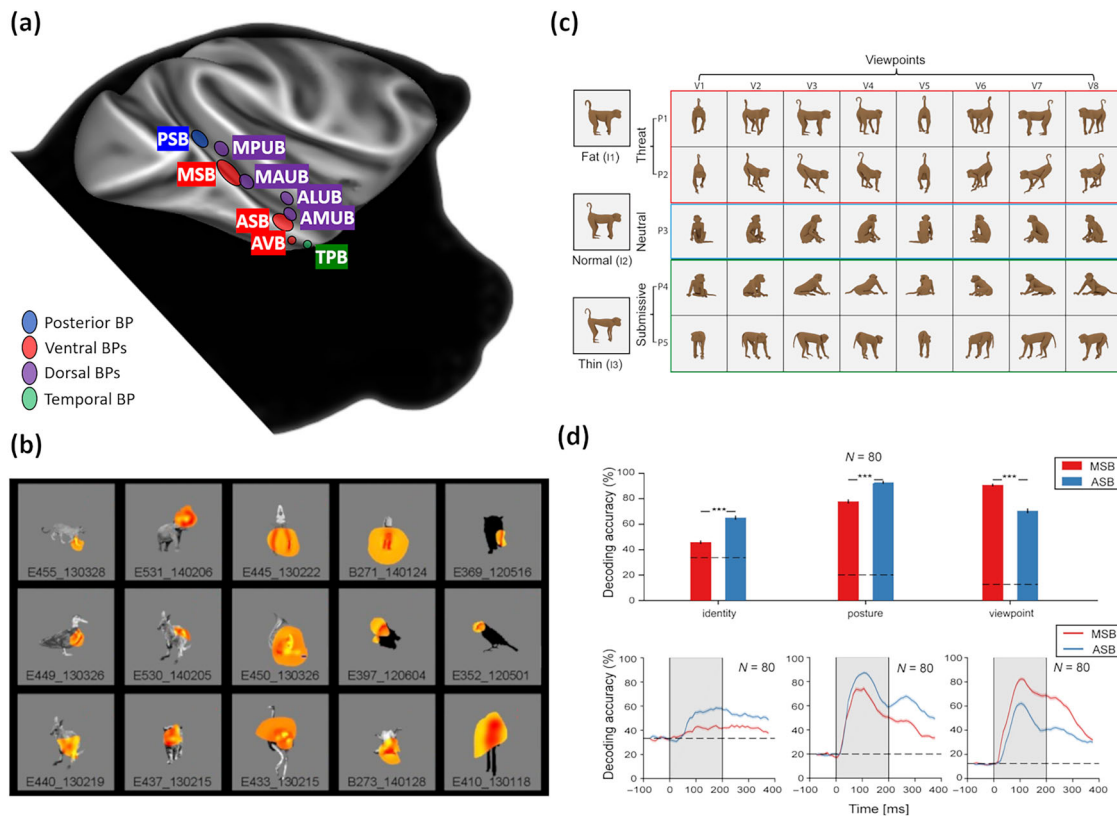


Figure 1. Monkey body patch network and selectivity of ventral STS body patch neurons. **a**, Schematic of the extended body patch (BP) network in macaques, revealed by dynamic stimuli (Bognár et al., 2023). The body patches typically localized with static images are shown in red (MSB, ASB, AVB). Four additional patches (purple), identified by using acting monkeys, are found in the middle and rostral dorsal bank of the STS [middle posterior upper bank STS body patch (MPUB), middle anterior upper bank body patch (MAUB), anterior lateral upper bank STS body patch (ALUB), anterior medial upper bank body patch]. Body-selective activations are present more posteriorly in the STS (PSB) and the temporal pole (TPB). **b**, Example MSB neurons organized by their preferred body fragment class, as revealed by the Bubbles technique (modified from Popivanov et al., 2016). Excitatory fragments are indicated in orange overlaid on the original image. **c**, Monkey avatar set with orthogonally manipulated identity, posture, and viewpoint of monkey bodies, used to assess selectivity differences between ASB and MSB patches for these attributes (modified from Kumar et al., 2019). Images depicting eight viewpoints (rows) for each posture (columns) are shown for identity I2 (“normal”). Three identities (I1–I3) are shown for one posture (a threat; P1) and one profile viewpoint (V3) to the left. **d**, Decoding of identity, posture, and viewpoint from MSB and ASB neuronal responses (modified from Kumar et al., 2019). Top panel, Decoding accuracy for identity (left), posture (middle), and viewpoint (right) for MSB (red) and ASB (blue) neurons. Lower panel, Decoding accuracy time courses for identity (left), posture (middle), and viewpoint (right). The dashed line indicates the chance level. Gray shaded region marks stimulus presentation (0 = stimulus onset).

faces and bodies in the macaque. While fMRI studies have reliably shown that the STS body patches are more activated by body postures signaling fear than body postures signaling no particular emotion (de Gelder, 2009; Taubert et al., 2022a), this is not the case for the face patches (Hadj-Bouziane et al., 2008; Zhu et al., 2013; Taubert and Japee, 2024). That said, these findings lack confirmation from direct neuronal recordings.

In line with fMRI findings, neurons of the ventral STS body patches respond, on average, stronger to bodies, regardless of whether a head is visible or not (Bell et al., 2011; Popivanov et al., 2014; Kumar et al., 2019; Bao et al., 2020). Single neurons respond selectively to bodies of monkeys, humans, or four-legged mammals, and even to birds. They can be quite selective, some responding only to a small proportion of body images, and show selectivity for body posture (Popivanov et al., 2014; Kumar et al., 2019). Thus, their responses can be used to decode body identity, posture, and view. Their selectivity is largely invariant to changes in image position and size, indicating a robust body preference (Popivanov et al., 2015, 2016; Kumar et al., 2019). Additionally, there is evidence of higher planar orientation and viewpoint tolerance, as well as identity representations, in anterior body patches compared with MSB (Fig. 1*c,d*; Kumar et al., 2019; Bao et al., 2020). This follows the general principle of object coding in IT (Bao et al., 2020).

Current evidence suggests that the body-selective ventral STS neurons are not semantic body detectors per se but, rather, are driven by features that are prevalent in bodies (Fig. 1*b*), related to limbs or parts of the torso (Popivanov et al., 2016; Lappe et al., 2024). Indeed, these neurons can respond to inanimate objects, when an object possesses the visual features that the neuron responds to (Popivanov et al., 2016; Kalfas et al., 2017; Lappe et al., 2024). Many MSB neurons respond to body parts (Popivanov et al., 2014), but as yet it is unclear whether exact body-part configuration matters. Although it has been proposed that body patch neurons are part of a broader IT object feature map, being selective for “spiky” animate object features (Bao et al., 2020), a definite characterization of the features driving body selectivity, in particular to differentiate postures and identity, is still lacking.

In principle, a neuron that differentiates between different bodies (e.g., different response levels for bodies of different animal species or the different postures of the same conspecific) can contribute to the encoding of bodies, irrespective of its response to other categories like faces and objects. Neuronal responses selective for body parts, e.g., hands (Gross et al., 1969) or animals (Kar et al., 2019), have been observed also outside the STS body patches. This raises the question of whether one should restrict the investigation of body representations to body patches or, instead, also consider body-responsive neurons in other IT regions. This question is one about the specificity of the causal contribution of body patch neurons to body perception and, more broadly, the causal role of domain-specific modules in perception. There have been indications that the link between body perception and body patch activity is not straightforward. For example, while electrical microstimulation of ASB impacts the categorization of bodies, it also impacts behavior toward some other categories of visual stimuli. Indeed, the behavioral effects could not be explained simply by body selectivity (Kumar et al., 2022). Further, the selective inactivation of small parts of ventral IT, well outside the STS, results in discrimination deficits of animal species (e.g., dog vs bear; Rajalingham and DiCarlo, 2019). This suggests that neurons outside the body patch system contribute to the categorization of different animal species and that we should not limit the

examination of the neural representation of bodies to fMRI-defined body patches.

Given that the majority of single-unit studies have employed static images, even less is known about the processing of dynamic body displays. Older studies indicated that STS neurons, mainly of the dorsal bank, are sensitive to the sequences of poses and integrate motion and form cues to distinguish, e.g., walking direction (Perrett et al., 1985; Oram and Perrett, 1996; Barraclough et al., 2006; Jellema and Perrett, 2006; Vangeneugden et al., 2011), which inspired computational models of action recognition (see Computational Models of the Recognition of Social Interactions). More recent single-unit recordings examining the responses to dynamic and static stimuli showed the contribution of motion to the body selectivity of dorsal bank STS neurons (Raman et al., 2023). Several dorsal bank STS neurons respond also to static features, but how their static body selectivity relates to that of ventral bank STS neurons is still an open question.

The stronger contribution of dynamics to body selectivity in the dorsal STS is in line with the proposal of a motion-sensitive, social pathway (Pitcher and Ungerleider, 2021), distinct from the classical ventral and dorsal visual pathways. However, the distinction in processing of bodily actions between the macaque dorsal and ventral bank STS requires more attention given the contribution of dynamics also to responses in the ventral bank of the STS (Raman et al., 2023).

Visual Body Representations in Humans

Research on the perception of social stimuli and their affective significance in humans has now moved beyond its traditional focus on faces to include body postures and movements. Early investigations into the neural correlates of human body perception were motivated by localizing body category-selective areas, typically by contrasting headless bodies with other categories such as bodyless faces and inanimate objects. These studies identified the extrastriate body area (EBA) in the middle occipital/temporal gyrus (Downing et al., 2001), and later the fusiform body area (FBA) in the fusiform cortex (Peelen and Downing, 2005; Schwarzlose et al., 2005), as well as a body-selective region

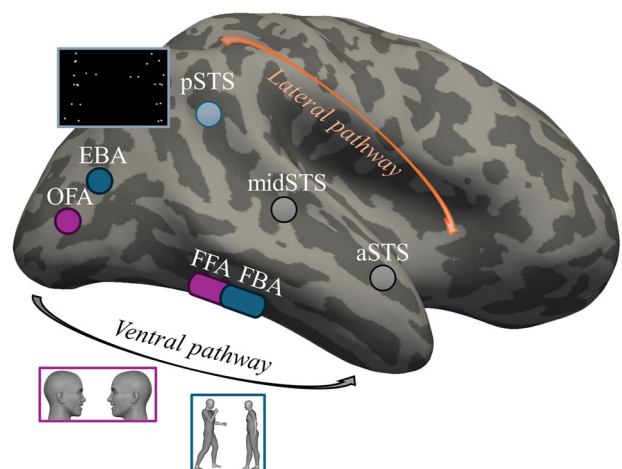


Figure 2. Ventral and lateral visual pathways for processing of social interactions. The ventral pathway for processing interacting bodies includes the extrastriate body area (EBA) and the fusiform body area (FBA). Areas processing interacting faces [the occipital face area (OFA) and the fusiform face area (FFA)] are shown also. The lateral pathway for processing interacting bodies includes the posterior superior temporal sulcus (pSTS). Localization of ROIs for the ventral pathway is from Abassi and Papeo (2020) and localizations for the lateral pathway from Isik et al. (2017) and Pinsk et al. (2009).

of the posterior superior temporal sulcus (pSTS; Fig. 2; Pinski et al., 2009).

Beyond that, some central issues await clarification. First, concerning the division of labor between EBA and FBA, it is still unclear whether EBA is primarily involved in processing body parts, encoding part-based features related to the shape and posture of the body. FBA, on the other hand, may then have a relative bias for holistic or configuration-based processing of whole-body images (Taylor et al., 2007; Downing and Peelen, 2011). Secondly, other brain regions beyond these body category-selective areas are also shown to contribute to body perception (de Gelder et al., 2010; de Gelder and Poyo Solanas, 2021). This matter is particularly relevant since, similar to the non-human primate studies above, most human studies on category selectivity used static images leaving it open what the contribution is of movement to body selectivity. For example, the pSTS has a key role in biological movement processing (Allison et al., 2000; Grossman et al., 2005; Fig. 2). Thirdly, it is not yet clear what specific computations take place in each body-selective area and what the functional connectivity of the areas is. Finally, it is currently an open question how these body-selective areas contribute to processing whole-body information for real-life tasks like emotional expression or identity recognition (de Gelder and Poyo Solanas, 2021). Taken together, these central questions, in line with monkey research as described above, suggest that body-selective areas do not consist of semantic body detectors per se. Rather, their activity may be driven by visual features that are statistically more prevalent in bodies but also occur in some objects.

Some new directions of human body perception research are not limited to body category-selective areas. For example, Li et al.

(2023) used data-driven methods aimed at discovering body perception networks at the whole brain level and computing network and node activity. They identified two distinct body-selective networks contributing differently to species-specific social perception. One network, which was not species-specific, was dominated by the lateral occipital cortex and the species-specific network included the EBA, the right pSTS, temporo-parietal junction, premotor cortex, and inferior frontal gyrus. Next, starting from these networks Li et al. (2024) investigated how different affective whole-body movements impacted network activity. This approach throws light on the important anatomical and functional complexity of EBA itself (Weiner and Grill-Spector, 2011) as Li et al. (2024) revealed four different EBA nodes. Depending on the specific action considered (aggressive, defensive, etc.) each EBA node had its separate network, e.g., defensive actions showing an enhanced connectivity between a specific EBA node and the precuneus and caudate nucleus.

There is a pressing need to investigate the processes bridging low-level and high-level perceptual processes. A crucial issue is to define the intermediate feature levels and this search can be guided by evolutionary and ethological observations (de Gelder and Poyo Solanas, 2021). A focus on features rather than on semantic categories is in line with current findings from nonhuman primates summarized above. Along these lines, the computations underlying EBA's selectivity for body information were explored using encoding models. Marrazzo et al. (2023) compared encoding models and found that the EBA encodes bodies through a combination of low-level visual features and postural features, independent of high-order body categorization processes (Fig. 3).

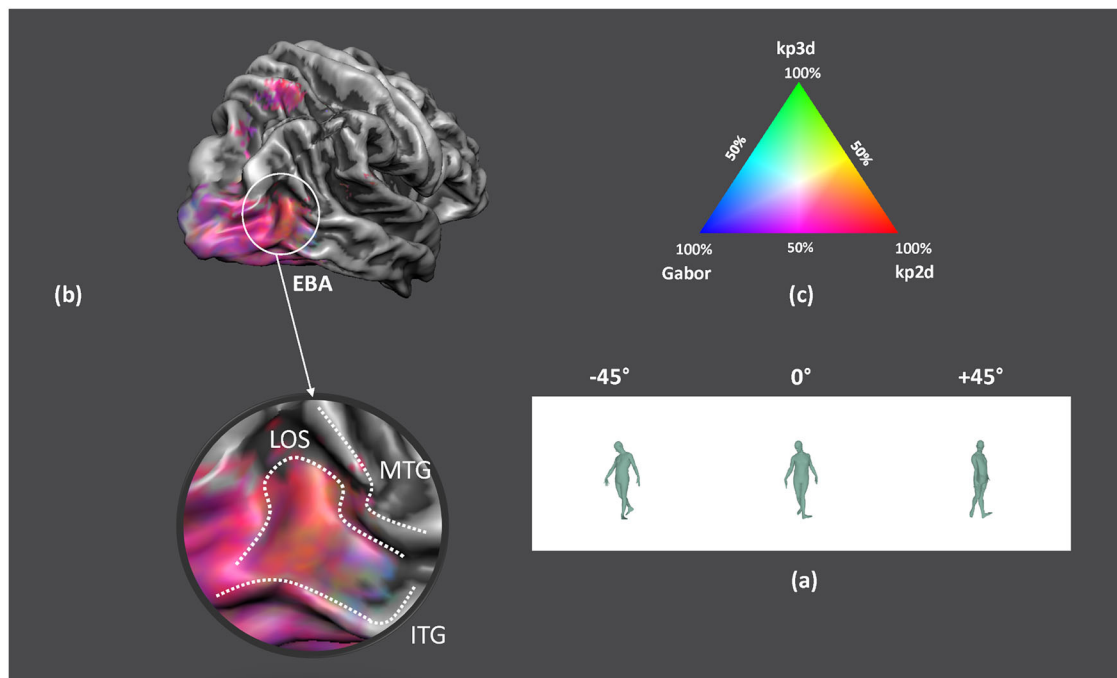


Figure 3. Information encoded in EBA for body stimuli. *a*, Stimuli of the 7T fMRI study by Marrazzo et al. (2023) were body postures presented in three viewpoints. Gabor filter responses and keypoint features were extracted from the stimuli and tested as hypotheses using linearized encoding. Specifically, keypoint features represent joint position in two (kp2d) and three dimensions (kp3d), with the latter being viewpoint. *b*, In EBA, the information contained in the joint model predictions which significantly correlates with BOLD activity follows a gradient from posterior (posterior ITG/LOS) to anterior (anterior LOS), with darker shades of magenta in the posterior part indicating a higher representation of low-level body features (Gabor), and lighter shades of magenta in the anterior part indicating a higher representation of mid-level features (kp2d–kp3d). ITG, inferior temporal gyrus; MTG, middle temporal gyrus; LOS, lateral occipital sulcus. *c*, RGB map in which each vertex is color coded according to the relative contribution of each model: Gabor filters, 2D body keypoint (kp2d), and 3D body keypoint (kp3d) to the accuracy of the joint model (red, 100% kp2d; blue, 100% Gabor; green, 100% kp3d). Modified from Marrazzo et al. (2023).

In social body communication posture and movement play a central role. The literature contains several detailed descriptions of how specific emotions and actions are conveyed through body posture or movement (James, 1890; Frijda, 1986; Wallbott, 1998). Recent methodological and technical advances for tracking and quantifying body movements and posture [e.g., OpenPose (Cao et al., 2017) or DeepLabCut (Mathis et al., 2018)], along with novel computational analysis methods such as Deep Neural Networks (Cichy and Kaiser, 2019), promise to make this field more systematic and objective. For example, a recent study computed kinematic and postural features from whole-body movements to investigate their relationship to the emotional category of the stimuli as well as to subjective ratings of similar feature descriptors (Poyo Solanas et al., 2020b). Key postural features, such as limb angles and symmetry, allowed to distinguish fearful movements from other emotional expressions. These features were encoded in areas associated with body and emotion perception but also in networks related to action representation and motor preparation (Poyo Solanas et al., 2020a).

Representations of Social Interactions

Several regions in the visual cortex are tuned to not only the perception of other agents and their bodies but also to the interactions between them. In particular, the EBA, pSTS, and nearby regions in the lateral occipital temporal cortex (LOTc; Fig. 2) are tuned to dyadic interactions (Isik et al., 2017; Wurm et al., 2017; Walbrin et al., 2018; Abassi and Papeo, 2020, 2024). For example, EBA represents not only the presence of single bodies, but also configural information about these bodies such as whether they are facing or in near proximity (Abassi and Papeo, 2022; McMahon and Isik, 2023), and disrupting EBA with transcranial magnetic stimulation selectively disrupts visual perception of social interactions (Gandolfo et al., 2024). Regions in surrounding LOTc differentiate between person-direction and object-direction actions (Wurm et al., 2017; Wurm and Caramazza, 2022), as well as information about the relative positioning of bodies in the scene (McMahon and Isik, 2023). The pSTS responds specifically to dynamic interactions, including agents moving toward versus away from each other (Bellot et al., 2021) or engaged in communicative interaction (Centelles et al., 2011; Isik et al., 2017; Walbrin et al., 2018; McMahon and Isik, 2023). The pSTS is also sensitive to some higher-level aspects of social interaction, such as distinguishing between teasing or threatening (Sinke et al., 2010) or helping and hindering interactions (Isik et al., 2017; Walbrin et al., 2018). All of these regions also contain information about contingent motion between agents, such as their interpersonal synchrony (Tsantani et al., 2024). Together, these regions comprise a “third visual pathway” specialized for social motion, localized in humans along the lateral surface of the brain (Fig. 2; Pitcher and Ungerleider, 2021). Beyond social interactions, this pathway might also be involved in dynamic face processing (Pitcher et al., 2011) and voice recognition (Deen et al., 2015; Pernet et al., 2015).

Some work has suggested that social interaction information in these regions is represented hierarchically, with information about the spatial configuration of agents primarily represented in EBA and nearby regions in LOTc, and more interpretive aspects of the interaction in pSTS (McMahon and Isik, 2023; Tsantani et al., 2024), though the differential roles of these regions is still an open question (Papeo, 2024). Though most neuroscience studies on social interaction have been conducted

in human adults, there is some preliminary evidence to suggest these neural mechanisms are early developing (Walbrin et al., 2020; Farris et al., 2022; McMahon and Isik, 2024) and shared with primate relatives (Sliwa and Freiwald, 2017; Goupil et al., 2024).

Until now, most of the studies on social interactions have approached the perception of social interaction as a function of goal-directed or reciprocal actions between two or more actors with a few studies investigating perception of the emotional contents of these interactions. In one of them, Clarke et al. (2005) examined whether the emotional content of behavior can be recognized from point-light displays where pairs of actors are engaged in interpersonal communication. They found that emotions could be correctly identified from these displays but that the performance was significantly diminished for some emotions when the displays were inverted, suggesting configural processing for dyadic emotions. They also showed that the recognition of emotions was impaired by the absence of the acting partner, suggesting a joint encoding of the body expressions of the participants of a dyad, a finding further supported by other studies also using point-light emotional dyads, e.g., Lorey et al. (2012). Using naturalistic videos, Sinke et al. (2010) showed that subtle differences in movement reliably convey differences between teasing and threatening interactions. More recently, researchers used computer animation to create scenes with two interacting avatars whose emotional style was independently controlled (Christensen et al., 2024). Here, participants had to report the emotional style of a single neutral agent while the emotion conveyed by the other agent was varying. The authors found converging evidence of a systematic influence of the emotion expressed by an agent on the perception of the other agent, further confirming the joint encoding of dyadic emotions. Extending from dyads to groups, an fMRI study used realistic videos of a large group of people expressing fearful, happy, or neutral emotions (Huis in 't Veld and de Gelder, 2015). Emotions were expressed by each individual in the group or by an interacting group. The results revealed that observing interacting individuals, more so than observing independently expressive individuals, activated brain regions related to the perception (e.g., the middle temporal gyrus; fusiform gyrus), execution and integration of action (e.g., cerebellum), and, critically, emotion (e.g., anterior cingulate). An important challenge in upcoming research on social interactions perception will be to disentangle different features of social interactions (e.g., limb synchronization and other motion features; McMahon and Isik, 2023) and investigate how they are processed in the brain.

Computational Models of the Recognition of Social Interactions

Another major line of investigation has focused on modeling of action recognition including the perception and recognition of biological motion. This work has employed specific computational approaches for extracting and processing the low-level features that define posture, body configuration, and kinematics, which can be further processed for inference about social interactions. A physiologically inspired neural two-stream model has been proposed by Giese and Poggio (2003) that accounts for the recognition of actions from moving bodies as well as from other biological stimuli. This model processes shape and motion features in parallel pathways akin to the ventral and dorsal streams. The shape pathway extracts increasingly complex shapes, mimicking the processing from V1 to V4, and results in neurons tuned to specific body posture snapshots as found

in macaque STS (Vangeneugden et al., 2011; Raman et al., 2023). The motion pathway consists of local motion detectors modeling direction-selective V1 neurons, local optic flow detectors to model MT neurons, and finally a set of neurons detecting optic flow patterns associated with a brief interval of a specific action. Subsequent work by Jhuang et al. (2007) expanded on this approach by equipping the model with a motion pipeline consisting of a hierarchy of alternating spatiotemporal template matching and pooling layers, enabling it to recognize actions in naturalistic videos. Lange and Lappe (2006) suggested a model based on shape cues alone, where each frame of an input sequence is matched to the closest template, and the resulting output sequence is processed by a leaky integrator. Thus, the model considers temporal aspects only in terms of high-level shape features, without encoding low-level motion. Work by Schindler et al. (2008) focused on recognizing emotional states from images of bodies, drawing on insights from object recognition. Classification is performed by linearly transforming principal components of max-pooled Gabor filters, which were learned by feeding the model with a set of body images.

A related physiologically inspired model for the recognition of goal-directed hand actions has been proposed by Fleischer et al. (2012) and Fleischer et al. (2013). This model combines keyframe-based mechanisms for the detection of hand shapes in terms of sequences of key shapes with mechanisms to process the relationship between the hand and the object, a property that critically influences the tuning of action-selective neurons in pre-motor and parietal cortex. A final classification layer integrates those two outputs with that of motion pattern neurons detecting the grip type of the hand to detect the presence of a particular action in the movie. The same type of model could also reproduce different aspects of the perception of causality from abstract stimuli, as investigated by Michotte (1963).

Since the advent of powerful deep-learning-based approaches in computer vision, a variety of methods have been proposed for estimating body posture or keypoints (Josyula and Ostadabbas, 2021; Chen et al., 2022) as well as for end-to-end action recognition (Le et al., 2022). These architectures are not devised to constitute accurate models of the brain but instead are optimized for task performance. However, pretrained convolutional neural networks (CNNs) can be used to replace the shape processing pathways from classical work, yielding a richer feature set and higher robustness of the models.

Extending the hand action recognition model by Fleischer et al. (2013), Hovaidi-Ardestani et al. (2018) presented a simple neural model that reproduces some of the key observations in psychophysical experiments about the perception of animacy and social interactions from abstract highly simplified stimuli similar to the ones of Heider and Simmel (1944). The model comprises a form and a motion pathway, each consisting of a hierarchy of feature detectors. The form pathway detects the instantaneous positions and orientations of agents of abstract shapes in a decontextualized scene. The motion pathway analyzes the 2D motion and the relative motion of the moving agents. It is based on relative position maps followed by series of sequential motion energy detectors performing temporal differentiation. This biologically plausible architecture allows the parallel estimation of relative positions, velocities, and accelerations of the two interacting agents. Based on the outputs of the motion pathway, a simple classifier in the output layer of the model robustly categorizes classes of social interactions with two moving agents. In follow-up work (Mukovskiy et al., 2022), the model was extended to exploit the standard CNN architecture as the form pathway, followed by robust shape classifiers and neural field representations for the robust tracking of the natural animal shapes in native cluttered environments (Fig. 4).

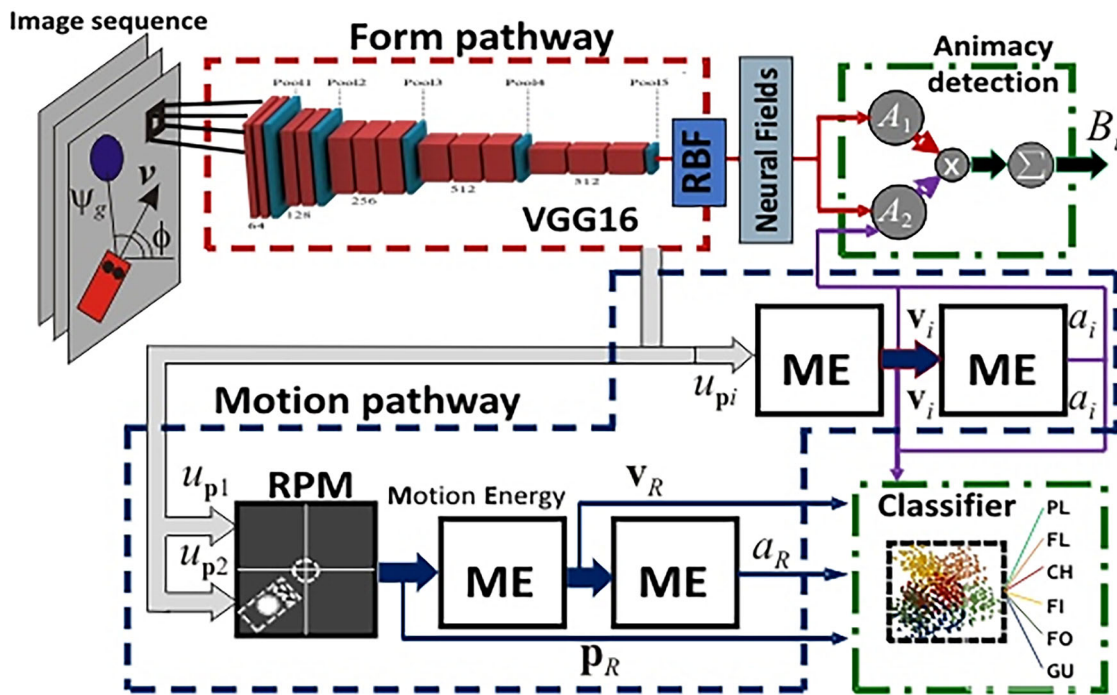


Figure 4. Neural model architecture extending the model of the visual pathway that analyzes perceived agency and classifies social interactions (Mukovskiy et al., 2022). Midlevel features are processed by the first 5 layers of the CNN VGG16. A radial basis function (RBF) network recognizes the position and orientation of agents for specific keyframes. Neural field/recurrent neural network (RNN) is used for the stabilization of agent tracking in the video sequence. The motion pathway performs biologically plausible computations of absolute velocity and acceleration of agents, relative distance, velocity, and acceleration. The motion pathway is followed by classifiers of the perceived interaction (RPM, relative position map; ME, motion energy estimator).

Alternative modeling approaches are based on Bayesian inference (Baker et al., 2008, 2009) and showed that Bayesian inverse planning allows one to make inferences about the latent goals of animated agents that were more similar to human inferences than bottom-up models (Fleischer et al., 2013) based on the recognition of low level features. They presented a computational model of social goal inference that takes as input observations of multiple agents moving in the same context. Thus, the authors argue for the importance of the inference of internal states based on the rationality of the other agents, i.e., the theory of mind concept.

Tauber and Steyvers (2011) used Bayesian inference on graphical models to interpret the movement goals of abstract animated agents (Heider & Simmel type). They showed that human observers use line-of-sight cues to assign the belief states to agents and that these belief states can be used to interpret the agent's behavior. This explanatory Bayesian model outperformed other observer models in describing human responses, also supporting the theory of mind assumptions. Shu et al. (2018) proposed a three-layered hierarchical Bayesian architecture (graphical model), where the first layer estimates the spatiotemporal motion patterns of the decontextualized agent (Heider and Simmel type) in short time periods. The second layer represents the spatiotemporal patterns over longer time periods. Finally, the third layer of hidden variables encodes the presence of the interactivity between two agents. The model trained with top-view aerial videos of the real but decontextualized scenes shows some generalization to the original Heider and Simmel animations. Using a similar Bayesian inference framework, Shu et al. (2021) proposed two underlying generative processes on the lowest level of representation of spatiotemporal patterns of the agents. The first one is based on the potential functions that capture the physical laws, constraints, and interactions and the second one of the value functions that describe the goals of agents. The authors show the superiority of their model in predicting social interactions, compared with existent bottom-up, data-driven deep neural network approaches.

Finally, Malik and Isik (2023) proposed a relational graphical network model that accurately predicts human interaction judgments across both animated and natural videos. Opposite to the previous results (Baker et al., 2008; Ullman et al., 2009), this new study suggests that these judgments are made without explicit simulations or inferences about agents' mental states.

Future Perspectives

In this brief review, we highlighted several directions of research regarding the representation of bodies and social interactions in the nonhuman primate and human brain. It is clear that still more work is needed to further our understanding of how the brain represents the body and how body perception depends on these representations. A more detailed understanding of the anatomical and functional complexity of body-selective areas is required, and the interactions between different body-selective (sub)regions need to be established in addition to their connectivity with other cortical and subcortical brain structures. The body representations in visual cortex can be viewed as entry points for further processing of body information for social and affective behavior, involving parietal, frontal, and subcortical areas. The integration of the visual cortical body network with these other networks needs to be investigated further. The body-selective regions might be best viewed as nodes within broader circuits responsible for social cognition and the perception of emotion in others. This also needs to be done for areas

representing social interactions. One important but challenging question concerns the homology of the body category-selective network in macaques and humans. MSB might be the homolog of EBA and ASB a homolog of FBA, but this homology is tentative. A comparison between monkey and human imaging and recording data suggested that the dorsal, but not ventral, bank of the macaque STS (Fig. 1) corresponds to the human STS (Jastorff et al., 2012) and thus might correspond to the proposed third visual pathway for social motion in humans (Fig. 2). The visual feature selectivity of the neurons in each of these regions still needs to be understood in both species. Single-unit recordings in these areas in human patients, when possible, using similar manipulations as employed in nonhuman primates will be informative. We expect that the combination and the improved integration of computational modeling and recordings using a cross-species perspective will advance our understanding of the neural mechanisms of body recognition and social interactions in the future.

References

- Abassi E, Papeo L (2020) The representation of two-body shapes in the human visual cortex. *J Neurosci* 40:852–863.
- Abassi E, Papeo L (2022) Behavioral and neural markers of visual configural processing in social scene perception. *Neuroimage* 260:119506.
- Abassi E, Papeo L (2024) Category-selective representation of relationships in the visual cortex. *J Neurosci* 44:e0250232023.
- Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. *Trends Cogn Sci* 4:267–278.
- Arcaro MJ, Ponce C, Livingstone M (2020) The neurons that mistook a hat for a face. *Elife* 9:e53798.
- Baker CL, Goodman ND, Tenenbaum JB (2008) Theory-based social goal inference. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* 30:1447–1452.
- Baker CL, Saxe R, Tenenbaum JB (2009) Action understanding as inverse planning. *Cognition* 113:329–349.
- Bao P, She L, McGill M, Tsao DY (2020) A map of object space in primate inferotemporal cortex. *Nature* 583:103–108.
- Barracough NE, Xiao D, Oram MW, Perrett DI (2006) The sensitivity of primate STS neurons to walking sequences and to the degree of articulation in static images. *Prog Brain Res* 154:135–148.
- Bell AH, Malecek NJ, Morin EL, Hadj-Bouziane F, Tootell RB, Ungerleider LG (2011) Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *J Neurosci* 31:12229–12240.
- 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. *Cereb Cortex* 31:2670–2685.
- Bognár A, Raman R, Taubert N, Zafirova Y, Li B, Giese M, De Gelder B, Vogels R (2023) The contribution of dynamics to macaque body and face patch responses. *Neuroimage* 269:119907.
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369–384.
- Cao Z, Simon T, Wei S-E, Sheikh Y (2017) Realtime multi-person 2d pose estimation using part affinity fields. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 7291–7299.
- Centelles L, Assaiante C, Nazarian B, Anton JL, Schmitz C (2011) Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: a neuroimaging study. *PLoS One* 6:e15749.
- Chen H, Feng R, Wu S, Xu H, Zhou F, Liu Z (2022) 2D human pose estimation: a survey. *arXiv*.
- Christensen A, Taubert N, Huis In 't Veld EMJ, de Gelder B, Giese MA (2024) Perceptual encoding of emotions in interactive bodily expressions. *iScience* 27:108548.
- Cichy RM, Kaiser D (2019) Deep neural networks as scientific models. *Trends Cogn Sci* 23:305–317.
- Clarke TJ, Bradshaw MF, Field DT, Hampson SE, Rose D (2005) The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception* 34:1171–1180.

- Deen B, Koldewyn K, Kanwisher N, Saxe R (2015) Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb Cortex* 25:4596–4609.
- Deen B, Schwiedrzik CM, Sliwa J, Freiwald WA (2023) Specialized networks for social cognition in the primate brain. *Annu Rev Neurosci* 46:381–401.
- de Gelder B (2006) Towards a biological theory of emotional body language. *Biol Theory* 1:130–132.
- de Gelder B (2009) Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philos Trans R Soc Lond B Biol Sci* 364:3475–3484.
- de Gelder B, de Borst AW, Watson R (2015) The perception of emotion in body expressions. *Wiley Interdiscip Rev Cogn Sci* 6:149–158.
- de Gelder B, Poyo Solanas M (2021) A computational neuroethology perspective on body and expression perception. *Trends Cogn Sci* 25:744–756.
- de Gelder B, Van den Stock J, Meerem HK, Sinke CB, Kret ME, Tamietto M (2010) Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neurosci Biobehav Rev* 34:513–527.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470–2473.
- Downing PE, Peelen MV (2011) The role of occipitotemporal body-selective regions in person perception. *Cogn Neurosci* 2:186–203.
- Dunbar RIM (1998) The social brain hypothesis. *Evol Anthropol Issues News Rev* 6:178–190.
- Farris K, Kelsey CM, Krol KM, Thiele M, Hepach R, Haun DB, Grossmann T (2022) Processing third-party social interactions in the human infant brain. *Infant Behav Dev* 68:101727.
- Fisher C, Freiwald WA (2015) Whole-agent selectivity within the macaque face-processing system. *Proc Natl Acad Sci U S A* 112:14717–14722.
- Fleischer F, Caggiano V, Thier P, Giese MA (2013) Physiologically inspired model for the visual recognition of transitive hand actions. *J Neurosci* 33:6563–6580.
- Fleischer F, Christensen A, Caggiano V, Thier P, Giese MA (2012) Neural theory for the perception of causal actions. *Psychol Res* 76:476–493.
- Frijda NH (1986) *The emotions*. Cambridge: Cambridge University Press; Editions de la Maison des Sciences de l'Homme.
- Gandolfo M, Abassi E, Balgova E, Downing PE, Papeo L, Koldewyn K (2024) Converging evidence that left extrastriate body area supports visual sensitivity to social interactions. *Curr Biol* 34:343–351.e5.
- Giese MA, Poggio T (2003) Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4:179–192.
- Goupil N, Rayson H, Serraille E, Massera A, Ferrari PF, Hochmann JR, Papeo L (2024) Visual preference for socially relevant spatial relations in humans and monkeys. *Psychol Sci* 35:681–693.
- Grezes J, Pichon S, De Gelder B (2007) Perceiving fear in dynamic body expressions. *Neuroimage* 35:959–967.
- Gross CG, Bender DB, Rocha-Miranda C (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* 166:1303–1306.
- Gross CG, Rocha-Miranda C, Bender D (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol* 35:96–111.
- Grossman ED, Battelli L, Pascual-Leone A (2005) Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Res* 45:2847–2853.
- Hadj-Bouziane F, Bell AH, Knusten TA, Ungerleider LG, Tootell RB (2008) Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. *Proc Natl Acad Sci U S A* 105:5591–5596.
- Hamlin JK, Wynn K, Bloom P (2007) Social evaluation by preverbal infants. *Nature* 450:557–559.
- Heider F, Simmel M (1944) An experimental study of apparent behavior. *Am J Psychol* 57:243–259.
- Hovaidi-Ardestani M, Saini N, Martinez AM, Giese MA (2018) Neural model for the visual recognition of animacy and social interaction. 27th International Conference on Artificial Neural Networks, Rhodes, Greece, Oct. 4–7, 2018, Proceedings Part III:168–177.
- 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.
- Huis in 't Veld EM, de Gelder B (2015) From personal fear to mass panic: the neurological basis of crowd perception. *Hum Brain Mapp* 36:2338–2351.
- Isik L, Koldewyn K, Beeler D, Kanwisher N (2017) Perceiving social interactions in the posterior superior temporal sulcus. *Proc Natl Acad Sci U S A* 114:E9145–E9152.
- James W (1890) *The principles of psychology*. Vol. 1. New York: Henry Holt and Company.
- Jastorff J, Popivanov ID, Vogels R, Vanduffel W, Orban GA (2012) Integration of shape and motion cues in biological motion processing in the monkey STS. *Neuroimage* 60:911–921.
- Jellema T, Perrett DI (2006) Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia* 44:1535–1546.
- Jhuang H, Serre T, Wolf L, Poggio T (2007) A biologically inspired system for action recognition. In: 2007 IEEE 11th International Conference on Computer Vision, pp 1–8.
- Josyula R, Ostadabbas S (2021) A review on human pose estimation. arXiv.
- Kalfas I, Kumar S, Vogels R (2017) Shape selectivity of middle superior temporal sulcus body patch neurons. *eNeuro* 4:ENEURO.0113-17.2017.
- Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat Neurosci* 22:974–983.
- Kumar S, Mergan E, Vogels R (2022) It is not just the category: behavioral effects of fMRI-guided electrical microstimulation result from a complex interplay of factors. *Cereb Cortex Commun* 3:tgac010.
- Kumar S, Popivanov ID, Vogels R (2019) Transformation of visual representations across ventral stream body-selective patches. *Cereb Cortex* 29:215–229.
- Lange J, Lappe M (2006) A model of biological motion perception from configural form cues. *J Neurosci* 26:2894–2906.
- Lappe A, Bognár A, Nejad GG, Mukovskiy A, Martini L, Giese MA, Vogels R (2024) Parallel backpropagation for shared-feature visualization. arXiv preprint arXiv:240509827.
- Le VT, Tran-Trung K, Hoang VT (2022) A comprehensive review of recent deep learning techniques for human activity recognition. *Comput Intell Neurosci* 2022:8323962.
- Li B, Poyo Solanas M, Marrazzo G, de Gelder B (2024) Connectivity and functional diversity of different temporo-occipital nodes for action perception. *bioRxiv:2024.2001.2012.574860*.
- Li B, Poyo Solanas M, Marrazzo G, Raman R, Taubert N, Giese M, Vogels R, de Gelder B (2023) A large-scale brain network of species-specific dynamic human body perception. *Prog Neurobiol* 221:102398.
- Lorey B, Kaletsch M, Pilgramm S, Bischoff M, Kindermann S, Sauerbier I, Stark R, Zentgraf K, Munzert J (2012) Confidence in emotion perception in point-light displays varies with the ability to perceive own emotions. *PLoS One* 7:e42169.
- Malik M, Isik L (2023) Relational visual representations underlie human social interaction recognition. *Nat Commun* 14:7317.
- Marrazzo G, De Martino F, Lage-Castellanos A, Vaessen MJ, de Gelder B (2023) Voxelwise encoding models of body stimuli reveal a representational gradient from low-level visual features to postural features in occipitotemporal cortex. *Neuroimage* 277:120240.
- Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 21:1281–1289.
- McMahon E, Isik L (2023) Seeing social interactions. *Trends Cogn Sci* 27:1165–1179.
- McMahon E, Isik L (2024) Abstract social interaction representations along the lateral pathway. *Trends Cogn Sci* 28:392–393.
- Michotte A (1963) *The perception of causality*, Ed 1. London: Routledge.
- Mukovskiy A, Hovaidi-Ardestani M, Salatiello A, Stettler M, Vogels R, Giese MA (2022) Physiologically-inspired neural model for social interaction recognition from abstract and naturalistic videos. In: VSS Annual Meeting 2022, Abstracts.
- Oram MW, Perrett DI (1996) Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol* 76:109–129.
- Papeo L (2020) Twos in human visual perception. *Cortex* 132:473–478.
- Papeo L (2024) What is abstract about seeing social interactions? *Trends Cogn Sci* 28:390–391.
- Peelen MV, Downing PE (2005) Selectivity for the human body in the fusiform gyrus. *J Neurophysiol* 93:603–608.
- Pernet CR, et al. (2015) The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* 119:164–174.
- Perrett D, Smith P, Mistlin A, Chitty A, Head A, Potter D, Broennimann R, Milner A, Jeeves MA (1985) Visual analysis of body movements by

- neurons in the temporal cortex of the macaque monkey: a preliminary report. *Behav Brain Res* 16:153–170.
- Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, Kastner S (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *J Neurophysiol* 101:2581–2600.
- Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011) Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* 56:2356–2363.
- Pitcher D, Ungerleider LG (2021) Evidence for a third visual pathway specialized for social perception. *Trends Cogn Sci* 25:100–110.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2012) Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *Neuroimage* 63:723–741.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2014) Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J Neurosci* 34:95–111.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2015) Tolerance of macaque middle STS body patch neurons to shape-preserving stimulus transformations. *J Cogn Neurosci* 27:1001–1016.
- Popivanov ID, Schyns PG, Vogels R (2016) Stimulus features coded by single neurons of a macaque body category selective patch. *Proc Natl Acad Sci U S A* 113:E2450–E2459.
- Poyo Solanas M, Vaessen MJ, de Gelder B (2020a) The role of computational and subjective features in emotional body expressions. *Sci Rep* 10:6202.
- Poyo Solanas M, Vaessen M, de Gelder B (2020b) Computation-based feature representation of body expressions in the human brain. *Cereb Cortex* 30:6376–6390.
- Premereur E, Taubert J, Janssen P, Vogels R, Vanduffel W (2016) Effective connectivity reveals largely independent parallel networks of face and body patches. *Curr Biol* 26:3269–3279.
- Rajalingham R, DiCarlo JJ (2019) Reversible inactivation of different millimeter-scale regions of primate IT results in different patterns of core object recognition deficits. *Neuron* 102:493–505.e5.
- Raman R, Bognár A, Nejad GG, Taubert N, Giese M, Vogels R (2023) Bodies in motion: unraveling the distinct roles of motion and shape in dynamic body responses in the temporal cortex. *Cell Rep* 42:113438.
- Rogoff B, Paradise R, Arauz RM, Correa-Chávez M, Angelillo C (2003) Firsthand learning through intent participation. *Annu Rev Psychol* 54:175–203.
- Schindler K, Van Gool L, de Gelder B (2008) Recognizing emotions expressed by body pose: a biologically inspired neural model. *Neural Netw* 21:1238–1246.
- Schwarzlose RF, Baker CI, Kanwisher N (2005) Separate face and body selectivity on the fusiform gyrus. *J Neurosci* 25:11055–11059.
- Shu T, Peng Y, Fan L, Lu H, Zhu S-C (2018) Perception of human interaction based on motion trajectories: from aerial videos to decontextualized animations. *Top Cogn Sci* 10:225–241.
- Shu T, Peng Y, Zhu S-C, Lu H (2021) A unified psychological space for human perception of physical and social events. *Cogn Psychol* 128:101398.
- Sinke CB, Sorger B, Goebel R, de Gelder B (2010) Tease or threat? Judging social interactions from bodily expressions. *Neuroimage* 49:1717–1727.
- Sliwa J, Freiwald WA (2017) A dedicated network for social interaction processing in the primate brain. *Science* 356:745–749.
- Su J, Van Boxtel JJ, Lu H (2016) Social interactions receive priority to conscious perception. *PLoS One* 11:e0160468.
- Tauber S, Steyvers M (2011) Using inverse planning and theory of mind for social goal inference. *Proc 33th Annual Conference of the Cognitive Science Society* 33.
- Taubert J, Japee S (2024) Real face value: the processing of naturalistic facial expressions in the macaque inferior temporal cortex. *J Cogn Neurosci* 1–17.
- Taubert J, Japee S, Patterson A, Wild H, Goyal S, Yu D, Ungerleider LG (2022a) A broadly tuned network for affective body language in the macaque brain. *Sci Adv* 8:eadd6865.
- Taubert J, Ritchie JB, Ungerleider LG, Baker CI (2022b) One object, two networks? Assessing the relationship between the face and body-selective regions in the primate visual system. *Brain Struct Funct* 227:1423–1438.
- 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.
- Thomas AJ, Saxe R, Spelke ES (2022) Infants infer potential social partners by observing the interactions of their parent with unknown others. *Proc Natl Acad Sci U S A* 119:e2121390119.
- Tomasello M (2014) The ultra-social animal. *Eur J Soc Psychol* 44:187–194.
- Tomasello M (2019) *Becoming human: a theory of ontogeny*. Cambridge: Harvard University Press.
- Tsantani M, Yon D, Cook R (2024) Neural representations of observed interpersonal synchrony/asynchrony in the social perception network. *J Neurosci* 44:e2009222024.
- Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6:989–995.
- Ullman T, Baker C, Macindoe O, Evans O, Goodman N, Tenenbaum J (2009) Help or hinder: Bayesian models of social goal inference. In, pp 1874–1882.
- Vangeneugden J, De Maziere PA, Van Hulle MM, Jaeggli T, Van Gool L, Vogels R (2011) Distinct mechanisms for coding of visual actions in macaque temporal cortex. *J Neurosci* 31:385–401.
- Vogels R (2022) More than the face: representations of bodies in the inferior temporal cortex. *Annu Rev Vis Sci* 8:383–405.
- Waidmann EN, Koyano KW, Hong JJ, Russ BE, Leopold DA (2022) Local features drive identity responses in macaque anterior face patches. *Nat Commun* 13:5592.
- Walbrin J, Downing P, Koldewyn K (2018) Neural responses to visually observed social interactions. *Neuropsychologia* 112:31–39.
- Walbrin J, Mihai I, Landsiedel J, Koldewyn K (2020) Developmental changes in visual responses to social interactions. *Dev Cogn Neurosci* 42:100774.
- Wallbott HG (1998) Bodily expression of emotion. *Eur J Soc Psychol* 28:879–896.
- Weiner KS, Grill-Spector K (2011) Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage* 56:2183–2199.
- Wurm MF, Caramazza A (2022) Two ‘what’ pathways for action and object recognition. *Trends Cogn Sci* 26:103–116.
- Wurm MF, Caramazza A, Lingnau A (2017) Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *J Neurosci* 37:562–575.
- Zafirova Y, Bognár A, Vogels R (2024) Configuration-sensitive face-body interactions in primate visual cortex. *Prog Neurobiol* 232:102545.
- Zafirova Y, Cui D, Raman R, Vogels R (2022) Keep the head in the right place: face-body interactions in inferior temporal cortex. *Neuroimage* 264:119676.
- Zhu Q, Nelissen K, Van den Stock J, De Winter F-L, Pauwels K, de Gelder B, Vanduffel W, Vandenbulcke M (2013) Dissimilar processing of emotional facial expressions in human and monkey temporal cortex. *Neuroimage* 66:402–411.