

Letter

What is abstract about seeing social interactions?

Liuba Papeo ^{1,*}

Social vision is a branch of vision science investigating the visual processing of socially relevant stimuli, primarily people. The focus of this research has recently moved from individuals (faces/bodies and their actions) to groups (two faces/bodies and their interactions). This new focus has revealed that the visual system is particularly sensitive to information that signals social interaction, or the social engagement of an agent. In a recent article in *TICS* [1], McMahon and Isik propose that the visual system uses this information to generate abstract representations of social interaction on an ‘exciting middle ground’ that goes beyond the input structure, but precedes cognitive processes, such as theory of mind. A hierarchical organization that supports increasingly abstract levels of representation is common to many large-scale brain networks [2], and is probably a good model to understand the processing of social interaction. However, the mapping of different representational levels into the visual cortex is more uncertain than McMahon and Isik suggest.

In the perception of two bodies, the visual system ‘sees’ distance, relative positioning (facing/non-facing), postural relations, synchrony, and possibly other relational properties (or social primitives) that distinguish related/interacting from unrelated/non-interacting people. Face-to-face interacting dyads (Figure 1A) are processed faster and better compared with non-interacting dyads [3]. McMahon and Isik argue that the computation of social primitives along the visual hierarchy results in the abstract representation of social interaction

in regions such as the extrastriate body area (EBA) and the posterior superior temporal sulcus (pSTS). This representation would capture the structure of the input event beyond the specific visual features, enabling category-level generalization (i.e., the recognition of a structure that is common to, say, all instances of helping, and that distinguishes helping from hindering events). However, current results show that the response to social interactions in those regions lacks the quality of generality that defines abstract representation, and remains compatible with the representation of visual correlates of social interaction.

Neurons in EBA and pSTS show selectivity for bodies/body parts and body motion, respectively. Both regions also respond more strongly to face-to-face interacting than to non-interacting dyads. Consistent with the functional properties of neurons in these regions, EBA shows an effect for both static and dynamic body dyads, whereas pSTS does so only for dynamic dyads [4–6].

Critically, static body dyads do not yield pSTS activity. If it does not even generalize between static and dynamic stimuli, in what sense is the representation of social interaction in pSTS abstract? Other effects that are interpreted as high-level feature representation in pSTS [1] can be explained by fine-tuning to kinematic aspects of the stimuli. In particular, neural patterns in pSTS discriminate between helping and hindering events, as depicted in coordinated motion patterns of two agents [7]. Before representation of social-goal compatibility (cooperation versus competition), discrimination could concern those motion patterns: in helping, the two agents move along the same trajectory, whereas, in hindering, they move in opposite directions (Figure 1B). Key evidence for abstraction of features as high level as goal compatibility would be that the pSTS response to helping in one situation (climbing up a hill)

generalizes to a different situation (opening a box). This kind of evidence is missing.

In EBA, increased response to both static and dynamic social interaction could suggest abstraction. However, such a general response would also be found if EBA responds to each static snapshot that forms a motion sequence, as appears to be the case [8]. In effect, the response to body dyads in EBA does not even generalize to dyads of faces: EBA responds more to facing (versus non-facing) bodies, but does not show a comparable effect for facing (versus non-facing) faces [9].

If we dismiss for a moment the hypothesis of abstraction, discrimination between categories of social interactions (e.g., arguing versus celebrating) in EBA (Figure 1C) [10] may reveal so-far uncharted visual sensitivities to relational properties of facing/related bodies [11,12]. For instance, recent work shows that EBA, similar to pSTS, is sensitive to alignment/synchrony between facing individuals [11] (Figure 1D). Thus, rather than categorization, discrimination of social interactions (arguing versus celebrating) could capture consistent variations in the level of synchrony/alignment between stimuli.

It is an empirical fact that the superior/middle temporal cortex is a hub for social-information processing, where the boundaries between perception and cognition become blurred. Measuring brain response during the presentation of large sets of naturalistic events, a trend in the field, highlights the extent of this territory. However, research on algorithmically controllable stimuli is still needed to define which features, and to what extent, give rise to an individual’s perception of social interaction. This painstaking work can benefit computer vision, where bottom-up models have already seen improvements toward human-level social-interaction recognition, with the introduction of visual social primitives in the inputs [1]. It can also help

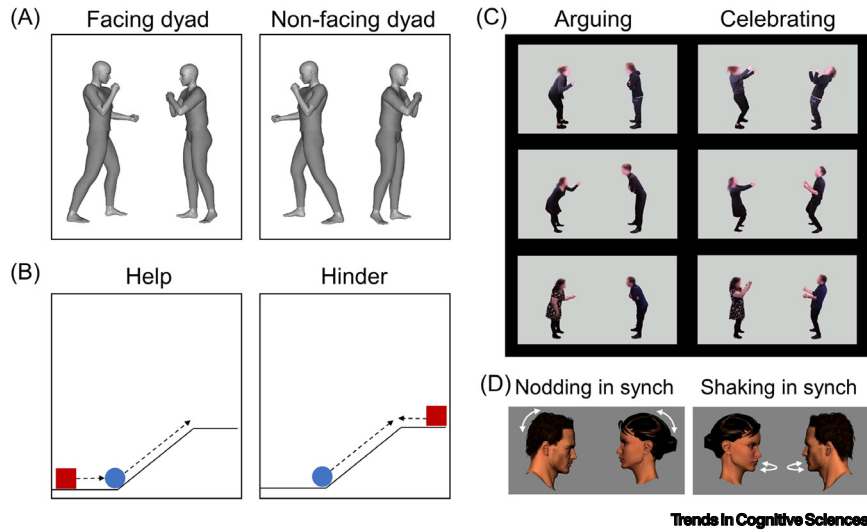


Figure 1. Examples of stimuli used to represent social interaction. Effects of social interaction perception in the lateral occipitotemporal visual areas are observed with stimuli that can be discriminated based on high-level as well as lower-level visual features. (A) In the posterior superior temporal sulcus (pSTS) and extrastriate body area (EBA), the stronger response to interacting (versus non-interacting) bodies can be reduced to the effect of spatial positioning: a stronger response to facing, seemingly interacting, than to non-facing bodies [4,6,9]. (B) The pSTS responds to animated shapes that act as social agents, and discriminates between events differing for goal compatibility (helping versus hindering) as well as motion patterns (same trajectories versus opposite and colliding trajectories) [7]. (C) In the EBA, discrimination between social interaction categories (arguing versus celebrating) [10] remains compatible with the effect of systematic visual differences, such as in postural relations, alignment, and synchrony. (D) Recent findings show sensitivity to postural alignment and/or synchrony of face-to-face individuals in EBA and pSTS [11].

References

1. McMahon, E. and Isik, L. (2023) Seeing social interactions. *Trends Cogn. Sci.* 27, 1165–1179
2. Margulies, D.S. et al. (2016) Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. USA* 113, 12574–12579
3. Papeo, L. (2020) Twos in human visual perception. *Cortex* 132, 473–478
4. Bellot, E. et al. (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, 2670–2268
5. Landsiedel, J. et al. (2022) The role of motion in the neural representation of social interactions in the posterior temporal cortex. *NeuroImage* 262, 119533
6. Gandolfo, M. et al. (2024) Converging evidence that left extrastriate body area supports visual sensitivity to social interactions. *Curr. Biol.* 34, 343–351
7. Isik, L. et al. (2017) Perceiving social interactions in the posterior superior temporal sulcus. *Proc. Natl. Acad. Sci. USA* 114, E9145–E9152
8. Giese, M.A. and Poggio, T. (2003) Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* 4, 179–192
9. Abassi, E. and Papeo, L. (2024) Category-selective representation of relationships in visual cortex. *J. Neurosci.* 44, e0250232023
10. Walbrin, J. and Koldewyn, K. (2019) Dyadic interaction processing in the posterior temporal cortex. *NeuroImage* 198, 296–302
11. Tsantani, M. et al. Neural representations of observed interpersonal synchrony in the social perception network. *PsyArXiv*, <https://doi.org/10.31234/osf.io/pjvke>
12. Wurm, M.F. and Caramazza, A. (2019) Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract representations of sociality. *NeuroImage* 202, 116153

isolate truly abstract representations of social interactions in the brain, those that might or might not correlate with the visual properties of the stimuli, but are independent from them and from the objects that instantiate the interaction, generalizing across features, exemplars, and stimulus modalities.

valuable comments on an early version of this article, and to Sofie Vettori for help in preparing Figure 1.

Declaration of interests

The author declares no competing interests.

¹Institute of Cognitive Sciences Marc Jeannerod –UMR5229, Centre National de la Recherche Scientifique (CNRS) and Université Claude Bernard Lyon 1, France

*Correspondence: liuba.papeo@isc.cnrs.fr (L. Papeo). <https://doi.org/10.1016/j.tics.2024.02.004>

© 2024 Elsevier Ltd. All rights reserved.

Acknowledgments

This work was supported by a European Research Council Starting Grant to L.P. (Grant number: THEMPO-758473). The author is grateful to Jean-Rémy Hochmann, Moritz Wurm, and Manuel Mello for