

Classification: Biological Sciences (Neuroscience)

# **Whole-agent selectivity within the macaque face-processing system**

Short Title: Whole-agent selectivity in the face patch system

Authors: Clark Fisher\* and Winrich A. Freiwald\*

Author Affiliations:

\*Laboratory of Neural Systems, The Rockefeller University, 1230 York Avenue, New York, NY 10065

Corresponding Author:

Clark Fisher

The Rockefeller University

1230 York Avenue #211

New York, NY 10065

(412) 496-4433

cfisher@rockefeller.edu

Keywords: face patches, body patches, visual integration, visual context

## **Abstract**

The primate brain contains a set of face-selective areas, which are thought to extract the rich social information that faces provide, such as emotional state and personal identity. The nature of this information raises a fundamental question about these face-selective areas: do they respond to a face purely because of its visual attributes, or because the face embodies a larger social agent? Here, we used functional magnetic resonance imaging to determine whether the macaque face patch system exhibits a whole-agent response above and beyond its responses to individually presented faces and bodies. We found a systematic development of whole-agent preference through the face patches, from subadditive integration of face and body responses in posterior face patches to superadditive integration in anterior face patches. Superadditivity was not observed for faces atop non-body objects, implying categorical specificity of face-body interaction. Furthermore, superadditivity was robust to visual degradation of facial detail, suggesting whole-agent selectivity does not require prior face recognition. In contrast, even the body patches immediately adjacent to anterior face areas did not exhibit superadditivity. This asymmetry between face- and body-processing systems may explain why observers attribute bodies' social signals to faces, and not vice versa. The development of whole-agent selectivity from posterior to anterior face patches, in concert with the recently described development of natural motion selectivity from ventral to dorsal face patches, identifies a single face patch, AF, as a likely link between the analyses of facial shape and semantic inferences about other agents.

## **Significance Statement**

From the ever-changing patterns of light captured by the eyes, our brains construct a meaningful model of the world. While we do not know the mechanisms that accomplish this feat, important clues are provided by a series of areas in the primate brain called face patches, which take visual inputs and sequentially process information to extract facial identity. This system is widely considered to be specialized exclusively for faces. However, this study shows that with every processing step, the system increasingly integrates face and body information to represent whole agents. Nearby body patches, in contrast, do not engage in this process. These findings advance our understanding of the neural circuitry that enables the understanding of our social world.

## Introduction

Making sense of the world around us is a computationally daunting challenge, and one of the brain's great feats. The face patch system of the macaque monkey (1; Fig. 1A) is a uniquely valuable model for understanding how the brain constructs a meaningful representation of the visual world (2). This network of areas along the superior temporal sulcus (STS) of the temporal lobe responds selectively to the sight of faces, parses their visual properties, and supports their perception (3, 4). However, face recognition is just one part of a greater perceptual goal. Faces are singularly rich sources of social information: they convey the identities (5), emotions (6), and intentions (7) of entire agents. This raises a critical question: do face patches respond to a face because they are fundamentally selective for facial shapes, or because these areas represent the larger agent embodied by the face?

Currently, it is thought that two distinct temporal lobe networks – the face patches and a parallel system of body-selective patches (8-10) – separately process the visual features of social agents. However, the sight of a body can alter the perception of a face (11), and some neurons in the upper STS have been found to integrate face and body information (12). While face patch neurons respond poorly to images of headless bodies (3), an untested possibility exists that they integrate information from a paired face and body to represent an agent as a whole.

Learning whether face patches represent whole agents would address the larger question of where the machinery of facial perception transitions into the machinery of social cognition – the ability to combine posture, motion, and other social information to understand and interact with others (13, 14). One indication that this transition may occur within the face system is that dorsal patches are specifically sensitive to the naturalness of facial motion (15), an important quality for social understanding (16). Integration of face and body is an ideal signature with which to confirm whether social processing begins in the face patches because – unlike traits such as “theory of mind” that are challenging to test with precision (17) – it is easily operationalized. The hypothesis that dorsal face areas represent social agents, and not merely facial features, predicts a ventral-to-dorsal development of whole-agent integration across the temporal lobe. In contrast, the

prevailing view that face patches are strictly face-selective modules predicts that adding a body to a face will produce no effect in any patch. Identifying how each face patch responds to whole agents would test these two possibilities, and clarify if and how social knowledge evolves within the face-processing system.

In this study, we set out to determine whether macaque face or body patches integrate face and body information to represent whole animals. We did this by pairing pictures of faces with pictures of bodies or matched non-body objects (Fig. S1), then using functional magnetic resonance imaging (fMRI) to observe the response signatures that these stimuli evoked (Fig. 1B). If face patches were strictly category-selective for faces, they would respond to pictures of whole monkeys just as much as they responded to faces alone. If, alternatively, face patches were composed of independently face- and body-selective cells, their response would be additive, the sum of the separate responses to lone faces and lone bodies. However, a superadditive response to whole monkeys would argue *against* these models of strict category-selectivity and *for* synergistic whole-animal processing within the face patches (18). By observing whether the whole-animal responses differed from the responses to faces on top of non-body objects, we determined the selectivity of face patches' responses to whole animals. Finally, we tested whether the face patches integrate the form of whole animals even in the absence of identifiable facial features, observing how they responded to visually obscured faces paired with bodies and non-body objects. Comparing the results of these three experiments, we could determine whether face patches may be strictly concerned with the visual details of faces, or whether they sometimes process social information gathered from across an agent.

## Results

**Synergistic Form Integration in Part of the Face Patch System.** To determine how face and body patches respond to the conjunction of face and body, we first used fMRI to identify face patches PL, ML, MF, MD, AL, and AF (see 15), and body patches midSTS and antSTS (9) in four monkeys. We measured average activity within these areas while the monkeys viewed pictures of lone faces, lone bodies, and whole monkeys comprised of face and body (Fig. 2A).

The face patches showed a range of responses to whole animals. At the network's posterior end, patch PL's response to whole monkeys was indistinguishable ( $\tilde{p} > 0.05$ ) from its response to lone faces. More anteriorly, ML responded more to whole monkeys than it did to lone faces, but still less than the sum of the response to lone faces and the response to headless bodies (subadditively). Patches MF and MD demonstrated nearly additive whole-animal responses (Fig. 2B), and AL's response - though it showed a superadditive trend - was not significantly greater than additive. Remarkably, anterior face patch AF showed a significantly superadditive response to pictures of whole monkeys. 62% of this patch's response to whole monkeys could be attributed to faces, and 18% to bodies. Thus 20% of AF's response to whole monkeys was a synergistic result of combining face and body. The body patches varied little in their response signatures, with both midSTS and antSTS responding subadditively to pictures of whole monkeys. Thus, while both body patches responded as traditional models of strict category selectivity would predict, activity in certain face patches was driven not solely by faces, but could be synergistically increased by the addition of bodies.

**Anterior Face Patches Show Category-Specific Whole-Agent Integration.** Any interaction between face and body responses could be one of two types: either category-specific, reflecting the special relevance of face-body pairing, or category-agnostic, resulting from pairing a face with any other object. In fact, both types of interaction could co-occur in the same voxel. In general object-object interactions, the non-preferred stimulus is frequently found to reduce the overall response (19-21), and this could mask simultaneously occurring category-specific face-body synergy. Therefore, we next examined how each functional area responded when a face was shown not with its accompanying body, but with an inanimate object chosen to match the overall shape and scale of its body (Fig. S1, bottom). Unlike adding bodies to faces, adding non-body objects to faces did not significantly increase the response of most face patches (except PL; Fig. 3A), and no face or body area showed a superadditive response when faces were viewed with non-body objects (Fig. 3B, top).

To determine which areas showed category-specific synergy in response to whole agents, and to account for category-agnostic interaction effects, we subtracted the nonlinear effects of pairing faces with objects (Fig. 3B, top) from the nonlinear effects of pairing faces with bodies (Fig. 2B) to measure form-specific whole-agent

responses (Fig. 3B, bottom). The posterior and middle face patches PL, ML, MF, and MD, showed no significant difference in nonlinear effects, no matter whether faces were paired with bodies or non-body objects. On the other hand, anterior patches AL and AF increased their activity *specifically* in response to whole agents, with significantly greater nonlinear responses when faces were paired with bodies rather than non-body objects. Notably, neither body patch showed such a change in response to whole agents. In fact, antSTS's response to whole monkeys was more *subadditive* than its response to face-object pairs. These results reveal a category-specific face-body synergy that is unique to anterior face patches AL and AF.

**Whole-Agent Synergy in the Absence of a Clear Face.** While results thus far showed that the body was necessary to evoke whole agent synergy, the role of the face remained to be investigated. Although a body and head can convey social information in the absence of a clear face, faces (and the selective responses they elicit in AF and AL) may have been prerequisite for the whole-agent synergy we observed. Alternatively, faces may have *inhibited* whole agent-synergy in certain patches; previous human work has suggested that bodies influence face areas only when the presence of a face is ambiguous (22). We examined these possibilities by measuring responses to stimuli with heavily pixelated faces (Fig. S2A). The face patches responded much less to these blurred faces than to unobscured faces (Fig. S2A, top red vs. bottom red), allowing us to observe whether and how whole-agent responses depended on the presence of a concomitant face response.

Bodies with blurred faces produced synergistic activity (Fig. 4, top) similar to that of bodies with unobscured faces (Fig. 2B). Once again, posterior patches PL and ML showed subadditive responses, while MF and MD showed nearly additive responses. More anteriorly, AL again showed a superadditive trend and AF showed a significantly superadditive response. Body patch responses were, once more, subadditive. The superadditivity that bodies and blurred faces produced in the anterior patches was specific to the presence of bodies: blurred faces paired with non-body objects produced no such significantly superadditive response (Fig. 4, bottom). Directly subtracting the nonlinear effects of pairing blurred faces with objects from the effects of pairing blurred faces with bodies revealed a small agent-specific synergy in all face patches (reaching significance in MF; Fig. S2B, top). Subtracting this synergy from the agent-specific synergy seen with *clear*

faces (Fig. 3B, bottom) showed that, across all areas, whole-agent synergy with obscured faces was statistically indistinguishable from whole-agent synergy with unobscured faces (Fig. S2B, bottom). Thus the superadditivity of combined face-body responses in anterior face patches was robust to facial blurring, and, more generally, obscuring facial details and reducing face-specific activity had little effect on the selective whole-agent responses in any face or body patch.

## Discussion

When we observe a face atop a body, these two forms represent a single agent, a single identity, and single sets of feelings and motivations. This explains why poker players scrutinize both faces and bodies for tells, why it is alarming to recognize a friend from behind but find yourself face to face with a stranger when she turns around, and why, presumably, the brain uses both the face and the body to interpret emotion and identity (23, 24). This study shows that anterior face patches respond not only to the sight of faces, but also to the simultaneous observation of face and body, specifically integrating information from both into a unified representation of a single agent.

We found three different activity signatures when we measured face patch responses to whole monkeys: subadditive activity in PL and ML, additive activity in MF and MD, and superadditive activity in AF and possibly AL (Fig. 2B). The subadditive and additive responses align with the prevailing view that face patches consist of category selective neurons; unadulterated face-selective neurons would lead to the former and a mix of selective face and body neurons would lead to the latter. These responses were largely expected, as viewing multiple objects commonly evokes subadditive or additive fMRI signals (25-27) and previous studies of face-body integration in the human face areas have found responses in this subadditive to additive range (22, 28, 29). The superadditive fMRI signals in the anterior patches, however, are a surprising finding, at odds with the idea that these areas are strictly domain-specific. fMRI superadditivity is a strict criterion for sensory integration, with high specificity but low sensitivity, as integrative single neuron activity is easily obscured in fMRI signals (30). In fact, many fMRI experimenters hoping to identify neural integration have turned away from superadditivity and towards largely unselective integration criteria (18). In light of this challenge, the positive



synergy between face and body responses in the anterior face patches demonstrates a link between faces and bodies in these areas that is both previously unrecognized and remarkably robust.

Supporting such a unique link, superadditive responses in AF and AL depended *specifically* on a body; faces shown atop non-body objects in these patches elicited merely additive responses (Fig. 3B, top). In contrast to this whole-agent selectivity in AF and AL, responses to faces-object pairs in the other patches mirrored the responses to face-body pairs (Fig. 3B, bottom). This reinforces the idea that whole-animal responses in anterior patches are fundamentally different from the responses in other patches, and results from a unique relationship between face and body representations in these regions.

Notably, the superadditive response to whole animals is not ubiquitous across anterior areas sensitive to biological form, as body patch antSTS which neighbors AL and AF showed a *negative* whole-agent response (Figs. 2B, 3B). This functional asymmetry between face and body areas may explain why people sometimes misattribute bodies' social signals to faces (23, 24): bodies alter activity within the face-processing system, but faces do not evoke reciprocal activity within the body-processing system.

When we examined the dependence of whole-agent integration on face responses, we found that the superadditive effect of pairing bodies with faces survives even when a face is obscured and its evoked response is drastically reduced (Figs. S2A, 4). This remaining synergy is remarkable: it is a face patch response in the absence of a clearly defined face. This is reminiscent of the psychological finding that a faceless head paired with a body is often perceived as having a face (31), and suggestions that human face-processing mechanisms can be engaged by appropriate context alone (22). Studies have even found that the human “face-processing” system can be activated in the absence of human-like surface features by stimuli suggesting agency (32-35). No matter whether this superadditive activity represents a specific relationship between neural models of head and body or a more abstract agency signal, these findings show that face-body integration is a highly robust phenomenon that, unlike the classical response to solitary faces (Figs. S2A, red), can be invoked with just the suggestion of a face.

Although whole-agent integration emerges across the face patches – bolstering recent evidence that these patches represent social information beyond face form (15, 36) – this integration does not develop along the same ventral-to-dorsal trajectory as natural motion selectivity, but rather the posterior-to-anterior axis of the face processing system (Fig. 5): subadditive responses in posterior patches PL and ML gave way to additive responses in MD and MF, and eventually a superadditive response in AF (and likely AL). Thus two important facets of social cognition, natural motion processing and whole-animal integration, form along anatomically perpendicular axes. Only face patch AF, at the terminus of both axes, shows selectivity for both whole agents and natural facial motion (15). This convergence of streams representing complementary aspects of social agents is reminiscent of a recent theory of the human brain organization, which suggests that multiple functional pathways separately process static and dynamic animacy cues before converging at a site of integrated agent representation (35, 37). The idea that AF might be such a site in macaques is bolstered by evidence that the anterior STS integrates many forms of social information. Electrophysiological studies near the location of AF have revealed a wide range of specializations for biological perception – ranging from neurons that respond to the paired sight and sound of a person’s actions to those that respond to the presence of a visually obscured person – and led to speculation that the anterior STS is a hub of social cognition (38, 39). A recent study of AF neurons has furthered this notion, suggesting that this face patch is specialized for interpreting faces within a social scene. (40). The realization that this social-processing hub lies at the intersection of two identified functional axes provides the opportunity to explore how social cognition is fashioned.

The results described here, as they highlight the anterior face patches, encourage a re-examination of all face patches. First, they demonstrate that body information must somehow enter the selective and largely self-connected (41) face patch system. This suggests that non-face information within the face patches (42), local connections outside the face patch system, or network-mediated interactions with the wider social brain have a larger influence than previously recognized. Additionally, by verifying that body information is integrated by the anterior patches, the observed superadditivity implies that such information may influence the posterior patches. fMRI superadditivity is a specific but insensitive criterion that obscures many interactions (see above),

so the fact that fMRI responses to face-body pairs were linear or sublinear in the posterior patches does not rule out the integration of body information within these areas. Instead, the evidence that body information is present and integrated in some nodes of this tightly connected system raises the possibility that the influence of bodies is present in subtler ways in other nodes.

In total, this examination of whole-agent representation expands our understanding of biological form-processing in the temporal lobe, revealing a functional organization that moves from apparently separate representations of faces and bodies in the body patches and posterior face patches to a synergistic response to whole monkeys – face and body – in the anterior face patches. The ability to examine differences between these regions has the potential to shed light on major problems of object vision: how selectivity for complex objects develops (43, 44), how spatial context facilitates object processing (45), and how social cognition is established (14). The functional specializations identified here provide a framework for understanding the neural mechanisms that integrate visual information from faces and bodies, and help to address the overarching questions of how the brain combines visual information from distinct but related objects, and how it forms rich and flexible representations of social agents.

## **Methods**

All procedures conformed to local and NIH guidelines, including the NIH Guide for Care and Use of Laboratory Animals. Experiments were performed with the approval of the Institutional Animal Care and Use Committees of The Rockefeller University and Weill Cornell Medical College.

This experiment was carried out in parallel with a previously reported experiment on facial motion representations (15), included the same set of subjects, and shared stimulation, data collection, and analysis methods.

**Subjects.** We studied four male rhesus monkeys (*Macaca mulatta*), aged 3.5 to 5 years old, weighing 5.5 to 7.5 kg, referred to here as M1–M4. Headpost design and implantation and the creation of individualized anatomical models of each monkey's brain have been described previously (15).

**fMRI Data Acquisition.** All fMRI data were acquired with a 3T Siemens Tim Trio MRI scanner, using an AC-88 gradient insert (Siemens) and a custom-designed 8-channel surface coil (Lawrence Wald, MGH Martinos Center). Intravenous Feraheme (AMAG Pharmaceuticals; M1-M3) or Molday ION (BioPAL; M4) was administered to each subject to increase functional contrast; positive signal values reported here reflect increased blood volume. Volumes covered the entire brain with an isotropic resolution of 1mm<sup>3</sup> and a TR of 2s (15).

During each functional scan, a head-fixed subject sat in the sphinx position inside the scanner, fixating on a point at the center of the stimulus screen; in return for fixation, he received a periodic reward of juice or water. Eye position was monitored by a video-based eye tracking system (ISCAN). The fixation window size was adjusted before each session to account for daily variation in the noise of the eye-tracking signal. The window ranged from 2.5° × 2.5° to 4° × 4° of reported visual angle. We analyzed only runs where fixation remained inside of this window for ≥ 90% of the entire run. Both fixation errors and blinks were counted against this measure. Fixation performance was similar across stimulus conditions (Fig. S3).

**Visual Stimuli.** Stimuli were projected at 60 Hz onto a screen placed 35 cm in front of the monkey's eyes. Three block design stimulus sets were presented to each monkey: the whole agent, object category, and object motion stimulus sets.

The whole agent stimulus set consisted of 8 conditions, in which grayscale exemplars of different categories were placed either above or below the point of fixation. The screen above the fixation point could contain either a monkey face ("face") or a pixelated version of the same monkey face ("blurred face"), while either a monkey body ("body") or an inanimate object ("object") was placed below the fixation point (Fig. S1). Each face (or its blurred version) was matched only with the body of the same monkey or an object chosen to match the body in size and shape. Monkeys portrayed in the stimuli were not personally known to the subjects. All pictures of monkeys and objects were histogram-normalized (46) before assembly into stimuli. All stimuli were presented on a middle gray background at 15° × 15° in the scanner. Each block of this stimulus set consisted of 15 exemplar images shown for 0.4 s, with each image repeated 3 times in a block. A Latin squares design was used, resulting in 8 counterbalanced orders. Each run of this stimulus set began with 24 s

of grey, the 8 stimulus blocks (18 s apiece), 24 s of grey, the 8 stimulus blocks reversed, and a final 24 s of grey. Palindromic design allowed for balance with matched sets of 4 orders when available runs were not balanced across all 8 orders. 48, 28, 24 and 40 runs of this stimulus set were analyzed, respectively, for M1 to M4.

The object category and object motion stimulus sets, used to identify regions of interest (ROIs), have been described previously (15). The object category set contained blocks of grayscale images; each block contained images from a single object category (faces, headless bodies, manmade objects, etc.). This was used to identify body patches. The object motion set contained grayscale movies; the two blocks used in this study were movies of familiar macaque faces and movies of familiar cage toys. This was used to identify face patches.

**Data Analysis.** fMRI data was preprocessed and analyzed using standard linear modeling techniques, as described in SI Methods.

Face patch and body patch ROIs were identified using functional and anatomical characteristics. We identified face patches with the object motion stimulus set as regions responding more to natural face movies than to natural object movies (15). Body patches were identified using the object category stimulus set to find regions that responded more to pictures of bodies than to matched pictures of manmade objects, and were named (following 9) based on anatomical location and relationship to identified face patches. Because the currently recognized midSTS and antSTS body patches often include multiple areas of body selectivity, we chose the body-selective activation between ML and MF to be midSTS, and the one medial to AL along the ventral bank of the STS to be antSTS, as these were prominent and reliably identifiable in all monkeys and all hemispheres. Specific voxel choice is described in SI Methods.

For group analysis, we normalized the time courses from each ROI of each monkey and subjected them to a fixed effects analysis (see SI Methods, 15). The time courses in each face and body patch of each monkey were normalized by the response to whole monkeys (face & body) in that patch. We report adjusted  $p$  values ( $\tilde{p}$ ), using the Holm-Bonferroni method to control for familywise error rate (47).

## Acknowledgements

We thank Akinori Ebihara, Rizwan Huq, Sofia Landi, Srivatsun Sadagopan, Caspar Schwiedrzik, Stephen Shepherd, Julia Sliwa, and Wilbert Zarco for help with animal training, data collection, and discussion of methods; veterinary services and animal husbandry staff for care of the subjects; Galit Yovel for project discussion; and Marya Fisher and Caspar Schwiedrzik for comments on the manuscript.

This work was supported by a Human Frontier Science Program Research Grant (RGP0015/2013, WAF), an Irma T. Hirschl/Monique Weill-Caulier Trusts Award (WAF), a Pew Scholar Award in the Biomedical Sciences (WAF), a McKnight Scholars Award (WAF), the New York Stem Cell Foundation (NYSCF-R-NI23, WAF), the National Eye Institute (1R01EY021594-01A1, WAF), the NSF Science and Technology Center for Brains, Minds, and Machines (CCF-1231216/5710003506, WAF), and a Medical Scientist Training Program grant (NIGMS T32GM007739, CF) to the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD Program. WAF is a New York Stem Cell Foundation-Robertson Investigator.

## References

1. Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. *P Natl Acad Sci USA* 105(49):19514–19519.
2. Orban GA, Zhu Q, Vanduffel W (2014) The transition in the ventral stream from feature to real-world entity representations. *Front Psychol* 5:695.
3. Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330(6005):845–851.
4. Afraz A, Boyden ES, Aparicio P (2015) Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. *P Natl Acad Sci USA* 112(21):6730–6735.
5. O'Toole AJ, et al. (2011) Recognizing people from dynamic and static faces and bodies: dissecting identity with a fusion approach. *Vision Research* 51(1):74–83.
6. Ekman P (1993) Facial expression and emotion. *Am Psychol* 48(4):384–392.
7. Oosterhof NN, Todorov A (2008) The functional basis of face evaluation. *P Natl Acad Sci USA* 105(32):11087–11092.
8. Pinsk MA, et al. (2009) Neural representations of faces and body parts in macaque and human cortex: a

- comparative fMRI study. *J Neurophysiol* 101(5):2581–2600.
9. 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(2):723–741.
  10. Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2014) Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J Neurosci* 34(1):95–111.
  11. Aviezer H, et al. (2008) Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychol Sci* 19(7):724–732.
  12. Wachsmuth E, Oram MW, Perrett DI (1994) Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cereb Cortex* 4(5):509–522.
  13. Brothers L (1990) The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci* 1:27–51.
  14. Adolphs R (2010) Conceptual challenges and directions for social neuroscience. *Neuron* 65(6):752–767.
  15. Fisher C, Freiwald WA (2015) Contrasting specializations for facial motion within the macaque face-processing system. *Curr Biol* 25(2):261–266.
  16. Pavlova MA (2012) Biological motion processing as a hallmark of social cognition. *Cereb Cortex* 22(5):981–995.
  17. Schaafsma SM, Pfaff DW, Spunt RP, Adolphs R (2015) Deconstructing and reconstructing theory of mind. *Trends Cogn Sci* 19(2):65–72.
  18. James TW, Stevenson RA (2012) The use of fMRI to assess multisensory integration. *The Neural Bases of Multisensory Processes*, eds Murray MT, Wallace MM (CRC Press, Boca Raton), pp. 131–146.
  19. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18(1):193–222.
  20. Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19(5):1736–1753.
  21. Kastner S, Ungerleider LG (2001) The neural basis of biased competition in human visual cortex. *Neuropsychologia* 39(12):1263–1276.
  22. Cox D, Meyers E, Sinha P (2004) Contextually evoked object-specific responses in human visual cortex. *Science* 304(5667):115–117.
  23. Aviezer H, Trope Y, Todorov A (2012) Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338(6111):1225–1229.
  24. Rice A, Phillips PJ, Natu V, An X, O'Toole AJ (2013) Unaware person recognition from the body when face identification fails. *Psychol Sci* 24(11):2235–2243.
  25. MacEvoy SP, Epstein RA (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol* 19(11):943–947.
  26. Nagy K, Greenlee MW, Kovács G (2011) Sensory competition in the face processing areas of the human brain. *PLoS ONE* 6(9):e24450.

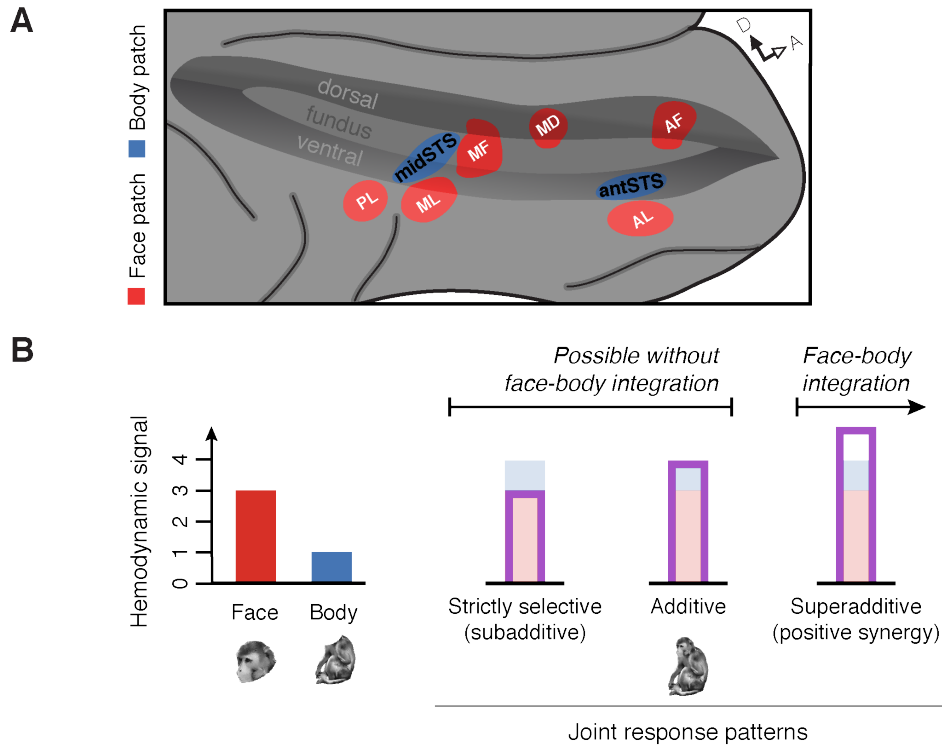
27. Baeck A, Wagemans J, Op de Beeck HP (2013) The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. *NeuroImage* 70:37–47.
28. Song Y, Luo YLL, Li X, Xu M, Liu J (2013) Representation of contextually related multiple objects in the human ventral visual pathway. *J Cogn Neurosci* 25(8):1261–1269.
29. Kaiser D, Strnad L, Seidl KN, Kastner S, Peelen MV (2014) Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *J Neurophysiol* 111(1):82–90.
30. Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res* 166(3-4):289–297.
31. Brandman T, Yovel G (2012) A face inversion effect without a face. *Cognition* 125(3):365–372.
32. Schultz RT, et al. (2003) The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philos Trans R Soc Lond, B, Biol Sci* 358(1430):415–427.
33. Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007) Two takes on the social brain: a comparison of theory of mind tasks. *J Cogn Neurosci* 19(11):1803–1814.
34. Gobbini MI, et al. (2011) Distinct neural systems involved in agency and animacy detection. *J Cogn Neurosci* 23(8):1911–1920.
35. Shultz S, McCarthy G (2012) Goal-directed actions activate the face-sensitive posterior superior temporal sulcus and fusiform gyrus in the absence of human-like perceptual cues. *Cereb Cortex* 22(5):1098–1106.
36. Russ BE, Leopold DA (2015) Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage* 109:84–94.
37. Shultz S, McCarthy G (2014) Perceived animacy influences the processing of human-like surface features in the fusiform gyrus. *Neuropsychologia* 60:115–120.
38. Brothers L (1996) Brain mechanisms of social cognition. *J Psychopharmacol* 10(1):2–8.
39. Keysers C, Perrett DI (2004) Demystifying social cognition: a Hebbian perspective. *Trends Cogn Sci* 8(11):501–507.
40. McMahon DBT, Russ BE, Elnaiem HD, Kurnikova AI, Leopold DA (2015) Single-unit activity during natural vision: diversity, consistency, and spatial sensitivity among af face patch neurons. *J Neurosci* 35(14):5537–5548.
41. Moeller S, Freiwald WA, Tsao DY (2008) Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320(5881):1355–1359.
42. Meyers EM, Borzello M, Freiwald WA, Tsao D (2015) Intelligent information loss: the coding of facial identity, head pose, and non-face information in the macaque face patch system. *J Neurosci* 35(18):7069–7081.
43. Connor CE, Brincat SL, Pasupathy A (2007) Transformation of shape information in the ventral pathway.



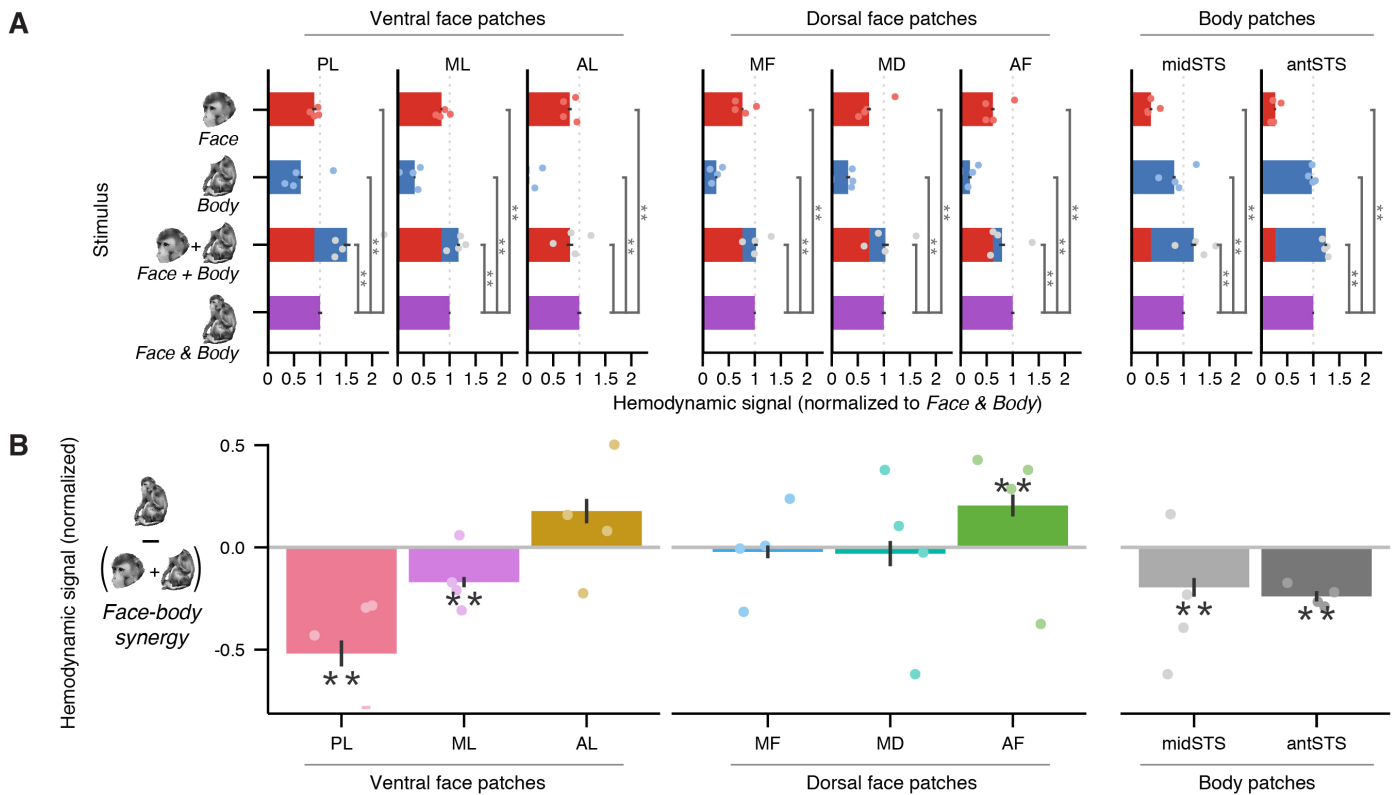
*Curr Opin Neurobiol* 17(2):140–147.

44. Rust NC, DiCarlo JJ (2010) Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30(39):12978–12995.
45. Bar M (2004) Visual objects in context. *Nat Rev Neurosci* 5(8):617–629.
46. Willenbockel V, et al. (2010) Controlling low-level image properties: The SHINE toolbox. *Behav Res Methods* 42(3):671–684.
47. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6(2):65–70.
48. Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RBH (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6(9):989–995.
49. Fischl B (2012) FreeSurfer. *NeuroImage* 62(2):774–781.
50. 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(2):911–921.
51. Polosecki P, et al. (2013) Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *J Neurosci* 33(29):11768–11773.
52. Leibovici DG, Smith S (2000) *Comparing groups of subjects in fMRI studies: a review of the GLM approach* (Oxford, UK) Available at:  
<http://www.fmrib.ox.ac.uk/analysis/techrep/tr00dl1/tr00dl1/index.html>.
53. Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? *NeuroImage* 10(1):1–5.

# Figures



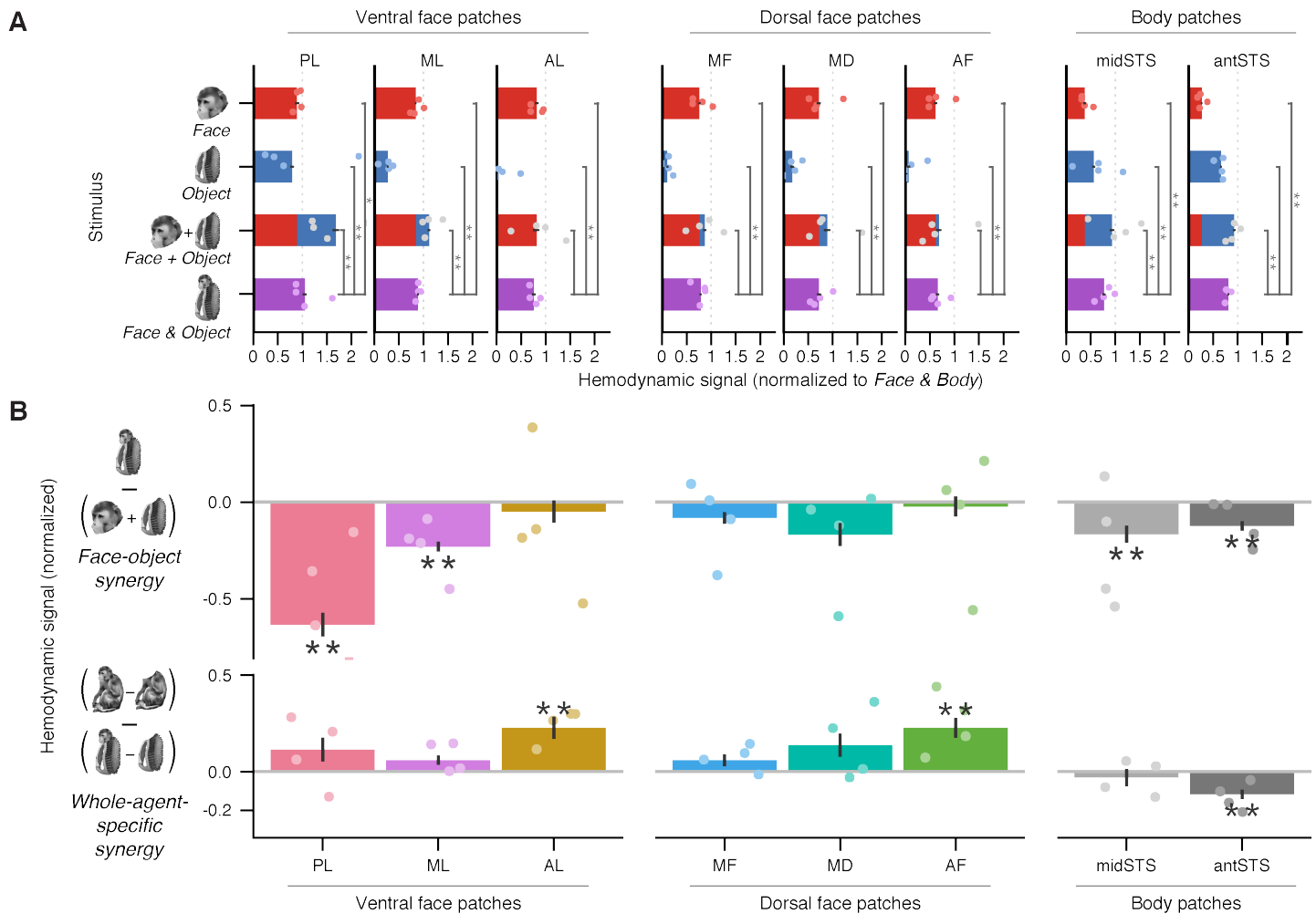
**Fig. 1.** Experimental plan. (A) Schematic of the macaque face and body patches along the superior temporal sulcus (STS) of a right hemisphere, with the anterior temporal pole on the right. (B) Three hypothetical responses of a model face patch to a face and body viewed together. Red and blue bars indicate the response of the patch to lone faces and headless bodies, respectively. Purple outlines show possible whole-agent response signatures given the responses to face and body. A strictly selective response to whole monkeys would be equivalent to the response to isolated faces, an additive response would equal the sum of the responses to individually presented faces and bodies, and a superadditive response would exceed this sum. Note that while a strictly selective or additive response could be created by a mix of selective face and body cells, a superadditive response is highly indicative of joint representation of face and body by single neurons (see 18).



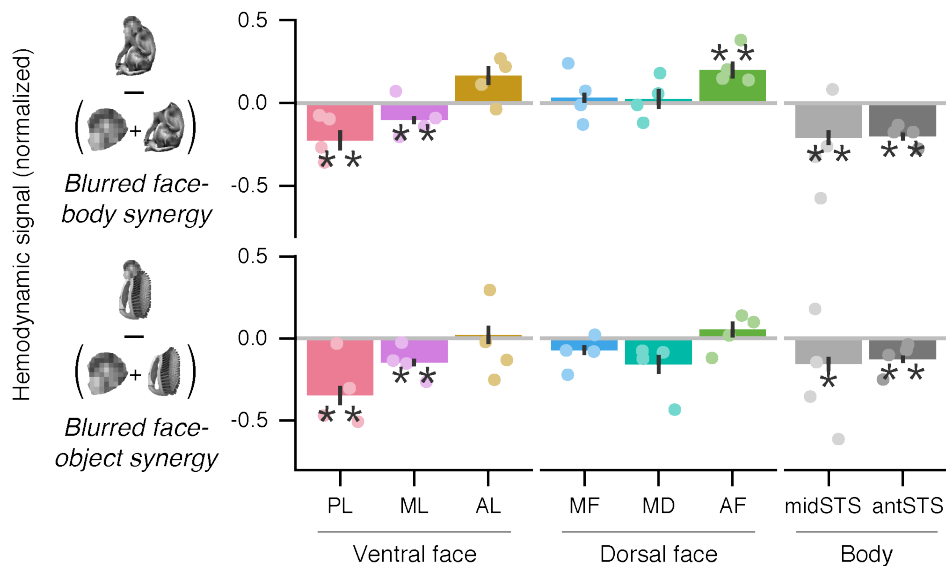
**Fig. 2.** Response to faces with bodies. (A) Response to lone faces (“face”, red), lone bodies (“body”, blue), and whole monkeys (“face & body”, purple) in 6 face patches and 2 body patches. Two-tone blue and red bars (“face + body”) represent the sum of the measured response to lone bodies and the measured response to lone faces, not the response to an independent stimulus condition. Dots on bar plots represent values for individual subjects; dashes represent dots outside of plotted bounds. Comparing responses to faces alone (red) and whole monkeys (purple, dotted vertical line), all face patches except PL responded significantly more to whole monkeys. The midSTS body patch responded significantly more to whole monkeys than to bodies alone (blue). (B) Additivity (linearity) of the joint response to viewing faces and bodies together: (face & body) – (face + body). This is the comparison between the purple bars and the red/blue bars in panel A. PL and ML were activated significantly less by viewing faces and bodies together than by viewing them separately. MF and MD showed an almost exactly additive response. Face patch AF was activated significantly more by viewing faces and bodies together than by viewing them separately (3 of 4 subjects). Although AL and AF showed similar degrees of superadditivity, AL’s did not reach statistical significance ( $\tilde{p} > 0.05$ ). Both body patches were activated significantly less by whole animals than was predicted by their responses to faces and bodies alone.

\* =  $\tilde{p} < 0.05$  and \*\* =  $\tilde{p} < 0.01$ , adjusted using Holm-Bonferroni method for 88 tests (8 ROIs × 11 measures [across Figs. 2, 3, 4, and S2]). Comparisons without an asterisk did not reach significance. Error bars

represent standard error. Signal change in bar plots is normalized per ROI to the response to “face & body” stimuli.

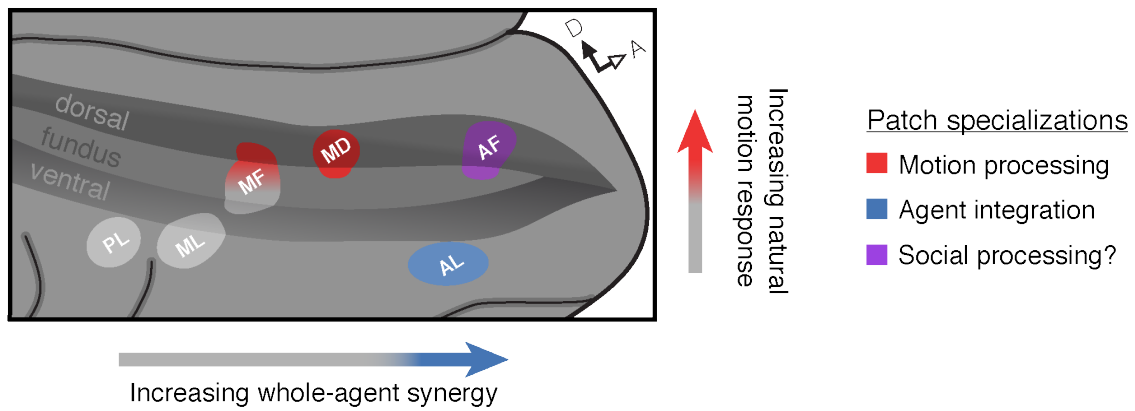


**Fig. 3.** Response to faces with non-body objects. (A) Response to monkey faces (“face”, red), non-body objects (“object”, blue), and their conjunction (“face & object”, purple) in 6 face patches and 2 body patches. Two-tone blue and red bars represent the sum of response to objects and the response to faces (“face + object”). (B, top) Additivity of the joint response to faces paired with non-body objects, i.e., (face & object) – (face + object). This is the comparison between the purple bar and the red/blue bar in panel A. All face and body patches were activated less when viewing faces and objects together than when viewing them separately, a difference that was significant in PL, ML, and the two body patches. (B, bottom) Form-specific whole-agent synergy, i.e., ((face & body) – body) – ((face & object) – object). This is the difference between Fig. 2B and the top plot in this panel. Both AL and AF showed significantly more benefit when viewing a face with a body than when viewing a face with a matched non-body object. Plotting conventions as in Fig. 2.



**Fig. 4.** Response to ambiguous faces with either bodies or non-body objects. (*top*) Response synergy between blurred faces and bodies, i.e. (blurred face & body) – (blurred face + body). This is the comparison tested in the top row of Fig S2A. PL and ML showed a subadditive response, MF and MD showed a nearly additive response, and AL showed a superadditive trend that does not reach significance. AF was activated significantly more by viewing blurred faces and bodies together than by viewing them separately. (*bottom*) Response synergy between ambiguous faces and bodies, i.e., (blurred face & object) – (blurred face + object). This is the comparison tested in the 2<sup>nd</sup> row of Fig S2A. PL and ML showed significantly subadditive responses, MD showed a subadditive trend that didn't reach significance, and AL, MF, and AF showed largely additive responses. In short, no face or body patch was activated significantly more by viewing blurred faces and objects together than by viewing them separately.

Plotting conventions as in Fig. 2.



**Fig. 5.** Correlates of social cognition across the face patch system. Different face patches within the STS have been shown to process two different facets of social cognition. *Dorsal* face patches (MD, AF, and sometimes MF) respond to the naturalness of facial motion (15) while *anterior* face patches (AF and AL) integrate face and body representations into a synergistic whole-agent response (this study). Face patch AF, at the dorsal and anterior extreme of this system, is unique in demonstrating both of these responses, and may therefore serve as an early hub of social cognition.

## SI Methods

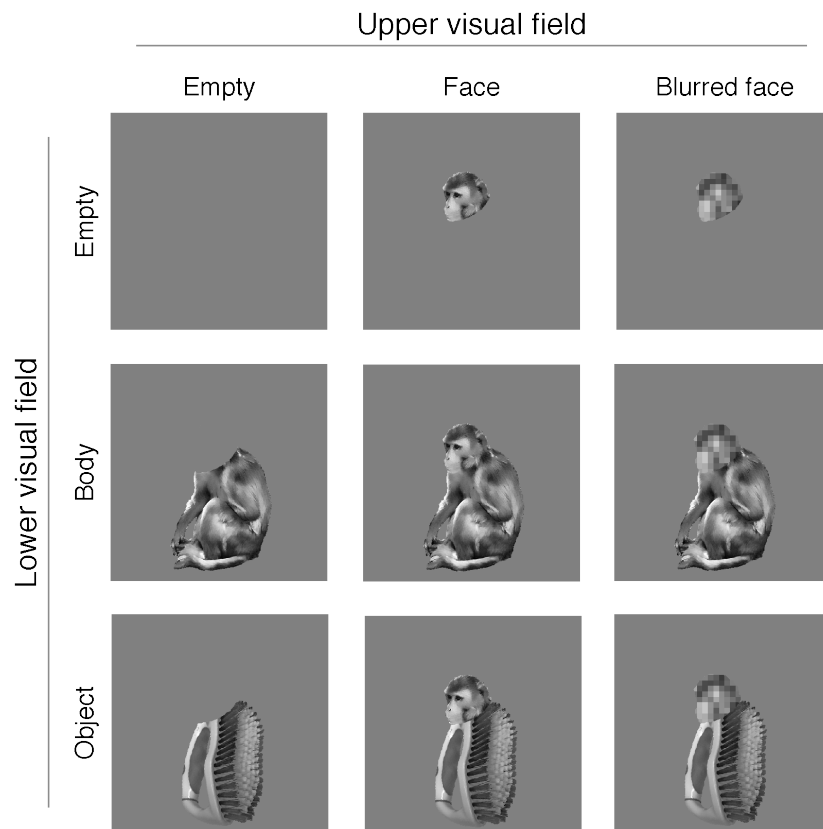
**Data Analysis.** We pre-processed the fMRI data with slice-wise motion correction, slice-time correction, and nonlinear alignment to an anatomical template, as previously described (15). We fit a general linear model to the preprocessed data with FS-FAST (<http://surfer.nmr.mgh.harvard.edu/fswiki/FsFast/>) to identify the effects of different stimulus conditions (48, 49).

ROIs were first defined in each hemisphere and then combined into bilateral ROIs. Within a given hemisphere, we positioned each ROI by finding a voxel showing a local maximum signal change in an area, selecting a 3 x 3 x 3 voxel region centered on this point, and excluding all voxels that did not meet a  $q$  (Benjamini-Hochberg false discovery rate)  $< 0.05$  significance criterion. If multiple maxima were found which could represent a given area of interest, we chose the one that maximized the symmetry between the hemispheres of a subject.

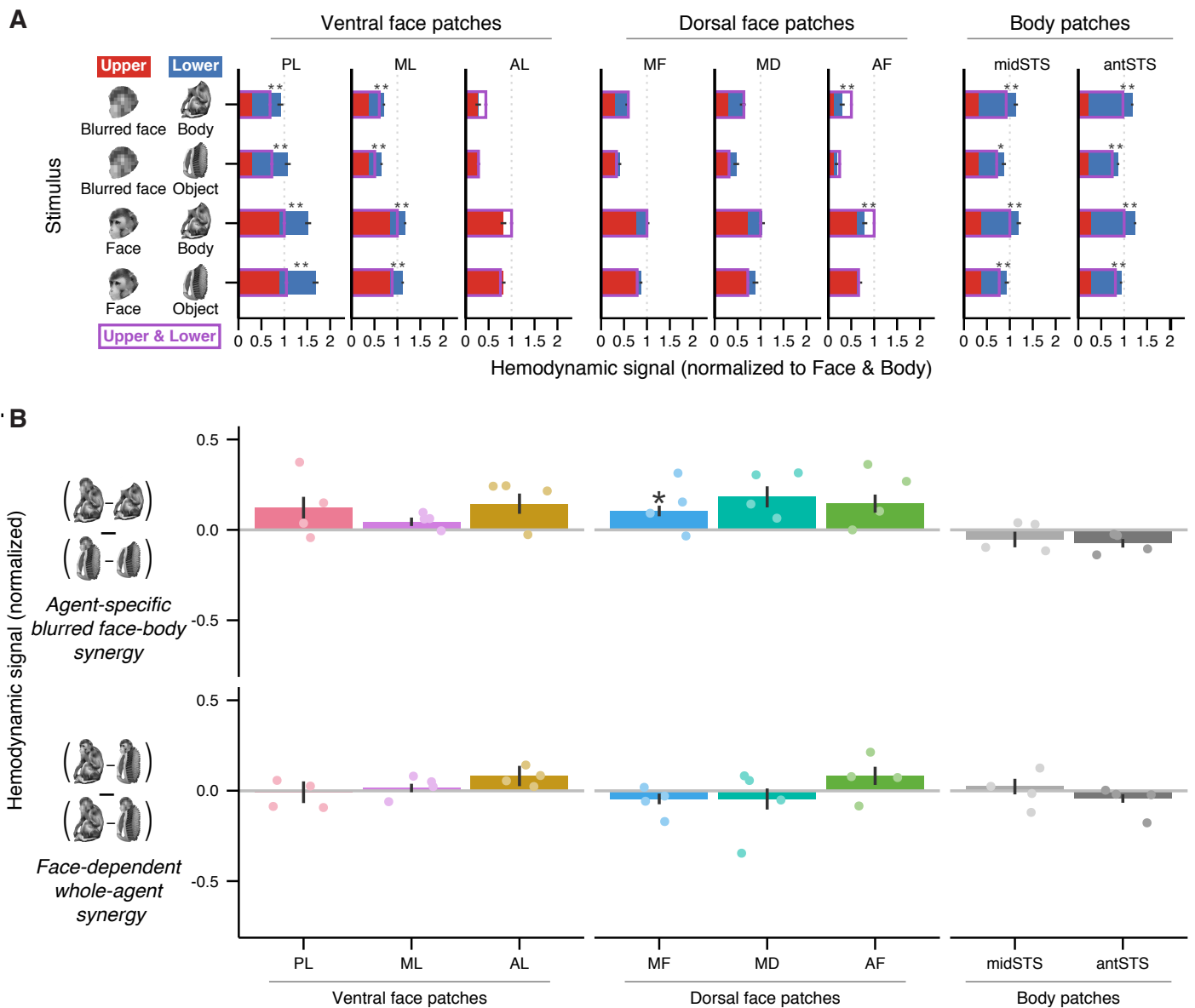
Because each ROI was independently normalized, comparisons between ROIs should be drawn with care, keeping in mind that the magnitude of each signal change is dependent upon the response of that ROI to whole monkeys. We used a fixed effects analysis (as is common in monkey fMRI experiments, e.g., 50, 51) because our sample size was too small for effective power in a random effects analysis (52) despite our comparatively large number of subjects. Statistical conclusions drawn from fixed effects analyses generalize only to the sample under study (see discussion in 53).



## SI Figures



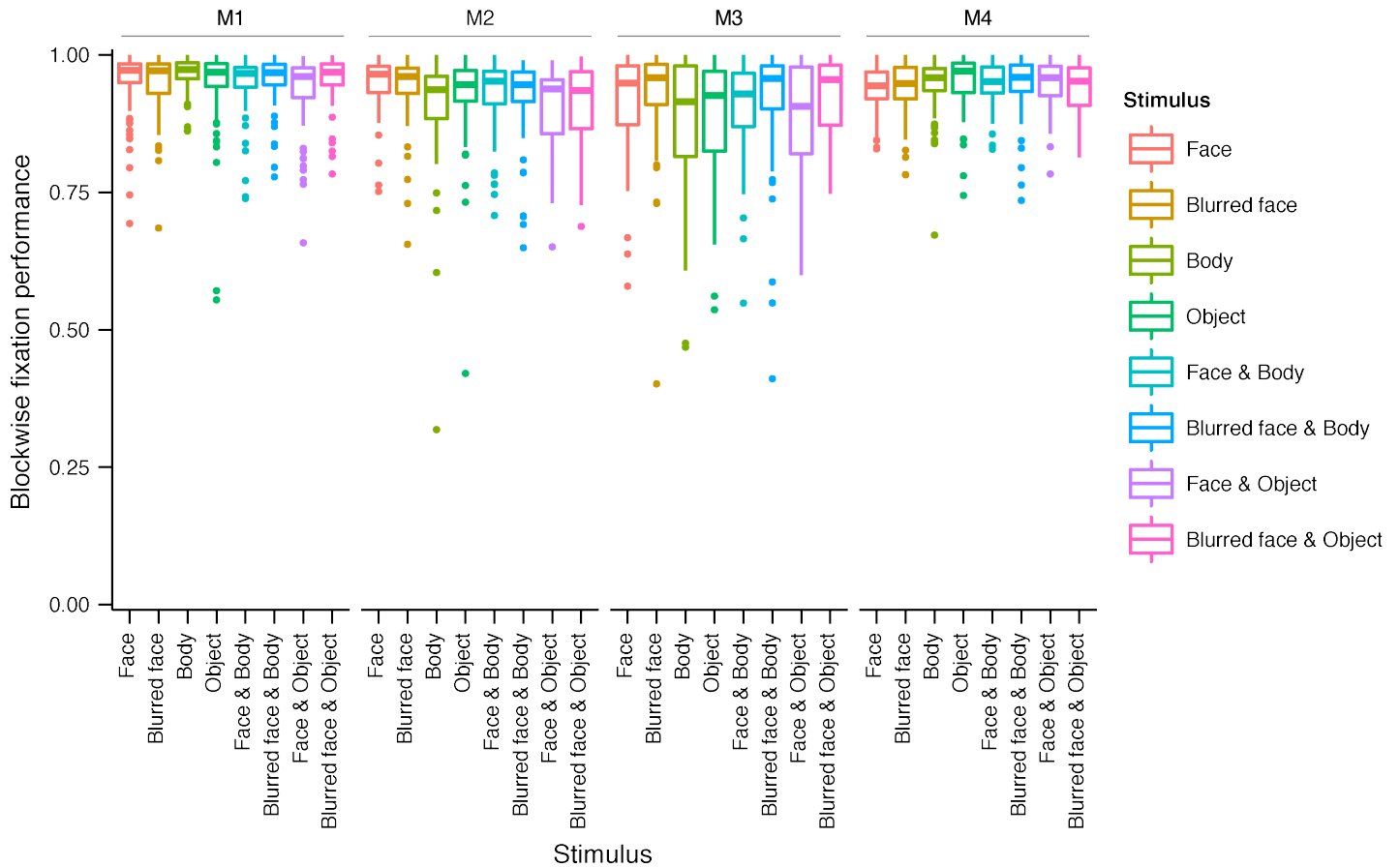
**Fig. S1.** Sample stimuli from the whole agent stimulus set. These stimuli represent all of the stimuli derived from the picture of a single monkey; the full stimulus set contained 14 additional examples of each stimulus type. Objects presented in the lower visual field were chosen to match the overall shape and size of the bodies. The empty/empty condition was presented in baseline blocks.



**Fig. S2.** Whole-agent synergy with an obscured face. (A) Effects of viewing 4 different pairs of upper/lower visual field stimuli, either singly or in concert, in 6 face patches and 2 body patches. Responses to lone upper field stimuli (faces or blurred faces) are represented by the red portion of each two-tone bar and responses to lone lower field stimuli (bodies or objects) are represented by the blue portion. Responses to viewing the pairs together are represented by the purple outlines. Asterisks mark cases where there is a significant difference between the response to viewing upper and lower field stimuli together and the sum of the responses to viewing them separately. The bottom two rows show the same data as Figs. 2A and 3A for the sake of comparison. (B, top) Difference between blurred face-body response nonlinearity and blurred face-object response nonlinearity, i.e.  $((\text{blurred face \& body}) - \text{body}) - ((\text{blurred face \& object}) - \text{object})$ . This is the difference between Fig. 4 (top) and Fig. 4 (bottom). (B, bottom) Specific whole-agent synergy that is dependent on the clarity of internal facial features, i.e.,  $((\text{face \& body}) - (\text{face \& object})) - ((\text{blurred face \& body}) - (\text{blurred face \& object}))$ .

face & object)). This is the difference between Fig. 3B (bottom) and the top plot of this panel. This effect did not reach significance in any examined area.

Plotting conventions as in Fig. 2.



**Fig. S3.** Fixation performance by subject and stimulus type. Fixation performance is reported as fraction of time within each 18 s stimulus block where a subject’s gaze remained within the fixation window; both fixation errors and blinks reduced this fraction. Horizontal lines in each boxplot represent the 1<sup>st</sup> quartile, median, and 3<sup>rd</sup> quartile of the distribution of fixation performance for a given subject and stimulus type. Vertical whiskers extend from the closest quartile to the last value within 150% of the interquartile range. Dots denote outliers beyond the whiskers. Across subjects, there was no pattern suggesting that fixation performance was consistently better or worse for a given stimulus or set of stimuli. Even within individual subjects, fixation performance was fundamentally consistent across stimulus categories: ANOVA analyses of the fixation performance of M1 ( $F_{(760, 7)} = 2.3$ ,  $\tilde{p} = 0.075$ ), M2 ( $F_{(470, 7)} = 2.5$ ,  $\tilde{p} = 0.063$ ), M3 ( $F_{(376, 7)} = 1.2$ ,  $\tilde{p} = 0.60$ ), and M4 ( $F_{(632, 7)} = 0.84$ ,  $\tilde{p} = 0.60$ ) across stimulus category did not reveal variance that reached significance when adjusted for 4 comparisons.