Functional Networks for Social Communication in the Macaque Monkey

Stephen V. Shepherd[1] and **Winrich A. Freiwald**[2]

[1]Lead Contact

[2]The Laboratory of Neural Systems, The Rockefeller University, New York, New York, 10065, USA

Summary

All primates communicate. To dissect the neural circuits of social communication, we used fMRI to map non-human primate brain regions for social perception, 2nd person (interactive) social cognition, and orofacial movement generation. Face perception, 2nd person cognition, and face motor networks were largely non-overlapping and acted as distinct functional units rather than an integrated feedforward-processing pipeline. While 2nd person context selectively engaged one region in medial prefrontal cortex, production of orofacial movements recruited distributed subcortical and cortical areas in medial and lateral frontal and insular cortex. These areas exhibited some specialization, but not dissociation, of function along the medio-lateral axis. Production of lipsmack movements recruited areas including putative homologs of Broca's area. These findings provide a new view of the neural architecture for social communication, and suggest expressive orofacial movements generated by lateral premotor cortex as a putative evolutionary precursor to human speech.

Graphical Abstract

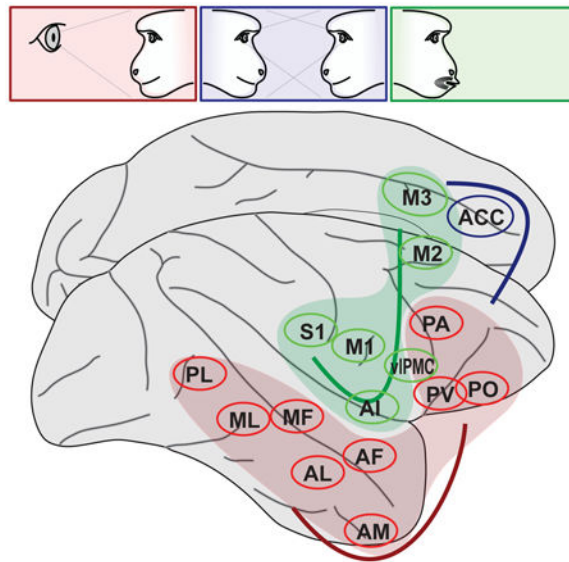Corresponding Emails: stephen.v.shepherd@gmail.com, wfreiwald@rockefeller.edu.

Author Contributions

SVS and WAF conceived experiments. SVS performed and analyzed experiments. SVS and WAF wrote the manuscript.

Declaration of Interests

The authors declare no conflicts of interest.

eTOC:

Shepherd & Freiwald examine the neural correlates of communication in monkeys during a simulated social interaction, discovering networks in the monkey brain for social cognition and social signal production, with surprising similarities to those producing human speech.

Introduction

As a rule, animals do not speak. However, all primates communicate, making prominent use of vocalizations and facial expressions. Facial expressions likely derived from evolutionary ritualization of noncommunicative postures indicative of arousal, aggressive or defensive movements, and ingestive actions (Andrew, 1963) and, as Charles Darwin famously noted, reveal otherwise hidden internal emotional states to others (Darwin, 1872). As such, facial expressions form an important category of visual signals in primate societies (van Hooff, 1967). To function as communicative signals, facial expressions must be perceived by a receiver, interpreted, and, at times, answered by a response. Because of their role in communication, facial expressions are more likely to be produced when there is an audience (Fernandez-Dols and Crivelli, 2013), and thus depend on the sender's awareness of social context. Despite their importance for emotional processing, communication, and social coordination, little is known about the neural circuits controlling facial expressions and their relationship to circuits of face recognition and social cognition.

Social perception, in the visual domain, relies on face recognition. Face recognition is supported by a network of selectively interconnected temporal and prefrontal face areas with unique functional specializations (Moeller, Freiwald and Tsao, 2008; Tsao *et al.*, 2008; Tsao, Moeller and Freiwald, 2008; Fisher and Freiwald, 2015; Grimaldi, Saleem and Tsao, 2016). While some of the outputs of the system are known, how outputs from this system are used for subsequent behavior, remains unknown. In particular, how the face-perception system relays information to facial motor areas during social communication is not understood. An

important clue to this questions is that facial expression exchange is partly reflexive. For example, humans (Dimberg, Thunberg and Elmehed, 2000) and perhaps macaques (Mosher, Zimmerman and Gothard, 2011) automatically mimic expressions, much as both primate species automatically follow gaze (Shepherd, 2010). This suggests that dedicated feedforward pathways may link these percepts to their respective reflexive responses. A major output of the face-perception network is the basolateral amygdala (Moeller, Freiwald and Tsao, 2008). The basolateral amygdala is also a major input into anterior cingulate facial motor area M3 (Morecraft, Stilwell–Morecraft and Rossing, 2004; Morecraft *et al.*, 2007). Area M3, in turn, projects directly to the facial nucleus (Morecraft, Stilwell–Morecraft and Rossing, 2004), which harbors motor neurons directly controlling facial musculature. It is thus a plausible hypothesis that face perception circuits drive emotional facial expressions in a feed-forward circuit from the face perception areas through the amygdala to the anterior cingulate cortex.

The central role this hypothesis ascribes to area M3 is in line with results from human neuropsychology, which has found a double dissociation between medial and lateral cortical areas in facial motor control. While lesions to medial frontal cortex impair affective expressions and vocalizations, lesions to lateral frontal cortex impair voluntary gestures and speech (Hopf, Müller-Forell and Hopf, 1992; Morecraft, Stilwell–Morecraft and Rossing, 2004; cf. Hage and Nieder, 2016). Despite the line of continuity between human and nonhuman expressions, human language has been argued to lack clear precedent in orofacial communication. Specifically, it has been argued that orofacial communication in non-human primates is too reflexive and simple to constitute precursors of speech, and that speech instead must derive from ape innovations in gesture and simulation (e.g. see Arbib, 2005). This reasoning suggests the hypothesis that emotional communications are shared with animals and mediated by medial frontal and subcortical structures, while volitional communication is unique to humans and mediated by lateral (and lateralized) structures (e.g. for vocal behavior; Wheeler and Fischer, 2012). However, unlike vocalizations, the neural substrates of primate facial expressions have received almost no attention from neuroscientists—thus the basic assumption of this hypothesis has not been tested, and the cortical substrates for naturalistic primate communication remains uncertain.

Once produced, facial expressions serve a communicative function only when there is an audience. Interestingly, social context impacts the production of human facial expressions: more are produced when there is an audience (Fernandez-Dols and Crivelli, 2013). Communication thus takes place in the so-called 2nd person context, in which the perceived individual is interacting with the subject. 2nd person contexts, it has been argued, elicit fundamentally different cognitive processes than the 3rd person contexts (the noninteractive observation of others) which have been traditionally used in cognitive science (Mosher, Zimmerman and Gothard, 2011; Schilbach *et al.*, 2013; Ballesta and Duhamel, 2015; Dal Monte *et al.*, 2016; Schilbach, 2016). Crucially, it is these understudied 2nd person contexts for which social communication evolved. Therefore, the study of social communication provides an opportunity to identify the neural circuits of 2nd person cognition and their relationship to the circuits of social signal production. The minimal instantiating condition for 2nd person contexts is a sense of mutual perception: We therefore simulated 2nd person

interactions in a controlled 'minimal interaction' behavioral framework, testing the hypothesis that these contexts activate circuits specialized for coordinating interactions.

Thus, to understand the neural processing of facial expression production, we need to understand the functional organization of facial motor circuits, their interaction with facial perception circuits, and their sensitivity to 2nd person context. Using a novel experimental approach, we aimed to address the following questions: What is the functional organization of face motor areas? In particular, are facial expressions generated by medial cortical areas, and voluntary face movements by lateral ones? What is the functional relationship between face perception and face motor circuits? Does 2nd person context facilitate the generation of facial expressions, and how is it represented in the brain?

Results

Bringing Social Behavior into the Laboratory

We tackled these questions with two innovations. First, we adapted whole-brain fMRI, which had been so successful unraveling the functional organization of the face-perception network, to the domain of social communication. Second, we developed a paradigm for eliciting social facial movements from macaques within the MR scanner. This allowed us to image neural activity across the brain during social perception, during 2nd and 3rd person social information processing, and during social signal production (Fig. 1).

The paradigm utilized videos of monkeys' dynamic facial displays taken from three angles (Fig. 1B, Fig. S1A, see STAR Methods): two from either side of the interaction (generating a 3rd person perspective) and one from the front, simulating direct eye contact (generating a 2nd person perspective). Showing the same facial movements from different directions kept all intrinsic properties of the real-world scene identical, but changed its social relevance to the subject (Schilbach *et al.*, 2006). As a low-level perceptual control, we also generated videos with systematically phase-scrambled frames, preserving spectral content and motion, but destroying shapes. During video presentation, monkeys seated in a scanner (Fig. 1C) spontaneously produced affiliative signals, particularly the 'lipsmack', whose rhythmic features bear similarities with human speech (Ghazanfar and Takahashi, 2014). These lipsmacks were generated selectively during 2nd person contexts (Fig. 1D). We scored subjects' facial movements within video segments corresponding to each 2-second fMRI frame: Out of a total of 16224 segments, 57% showed no facial movement, 26% drinking, 1% (225) lipsmacks, and 1% nonlipsmack expressions. (The remaining 15% were indeterminately ingestive or communicative, or included visible body or hand motion). Lipsmack versus nonlipsmack responses (see STAR Methods) varied significantly across experimental conditions, with 185 segments of lipsmacking observed during subject-directed social video, 25 in the subsequent blank period, and 15 total across all other conditions (X^2[df=27, N=13826]=2473, p<<0.001). Thus, 2nd person context was a nearly obligate prerequisite for the generation of this stereotypic facial movement.

Brain Responses to Social Stimuli in Social Context

Brain activity was modeled as a function of stimulus condition and parametrically of extraocular facial movement and background body movement as scored by a computer algorithm. Social videos, compared to scrambled controls, activated large parts of cortex (Fig. 2A, top; calculation based on brain responses to all social video less that to all scrambled video). These activations included visual, temporal and prefrontal face patches, as expected, as well as medial frontal areas near the rostral cingulate sulcus, areas PITd and LIP (implicated in attentional control and perceived gaze; Shepherd, 2010; Marciniak *et al.*, 2014; Stemmann and Freiwald, 2016), and several thalamic and amygdalar nuclei (Fig. 2A, bottom). Activity was reduced in frontal areas 46, 11 and 10m and large parts of insular, motor, and cingulate cortex. These effects were likely driven by the presence of visual forms and social content. In contrast, when we compared activation between 2nd and 3rd person social contexts, activity was much more restricted (Fig. 2B; calculation based on brain responses to all direct-gaze unscrambled video less that to all averted-gaze unscrambled video): A medial frontal cluster of areas around the rostral tip of the anterior cingulate cortex (ACC) was significantly more active in 2nd person context. In fact, the cluster was activated about twice as strongly in 2nd than in 3rd person contexts, a pattern rivaled only subcortically, by the periaqueductal grey (PAG). Thus, social context, a prerequisite for social signal production (Fig. 1D), appears to be signaled by a spatially-confined circuit.

Brain Activity During the Production of Communicative Social Signals

Facial movements yielded a separate activation pattern (Figs. 3A,B). Lipsmacks (Fig. 3A; calculation based on brain response during manually-annotated lipsmacks less that during manually-annotated nonmovement, with parametric evaluation of extraocular facial movement ignored) engaged a specific cortical and subcortical motor network including lateral primary somato-motor cortex (M1 and sM1), ventrolateral premotor cortex (VLPMC), anterior and dorsal insula (AI and DI), an ACC area (M3) just posterior to the 2nd person context cluster, and subcortical areas including thalamus, striatum, and brainstem (specifically the facial nucleus, trigeminal motor nucleus, and associated reticular network). A largely identical suite of activations was found for facial movements overall (Fig. 3B; calculation based on brain correlates with computer-scored extraocular face movement). Activation in the facial portion of M1 was so strong, it was often detectable during a single facial movement event (Video S1). Thus, in contrast to social perception of 2nd person contexts, social signal production is supported by a distributed but spatially-specific facial motor control system.

In monkeys, the facial movement network exhibited functional specialization, but not dissociation, between communicative and non-communicative movements. Facial-motor cortex (M1, M2, M3, and VLPMC, but *not* M4; cf. Morecraft et al., 2004) was strongly activated by all facial movements across social stimulus context (Fig. 4A, top) and movement type (Fig. 4A, bottom). Across all conditions, the lateral motor cortices were most strongly and consistently activated. However, lateral motor cortices were relatively more activated by non-communicative than communicative facial actions, while the reverse was true for the medial facial-motor regions, including putative area M4 (Fig. 4A, see also Fig. S3-S4). Two structures previously implicated in affective signaling, the amygdala and

PAG, showed only limited evidence for involvement in lipsmack exchange. Hence, neural control of different facial movements, including expressions, appears to be supported by a single facial control network with quantitative rather than qualitative division of labor between medial and lateral motor areas.

The Organization of Brain Circuits for Social Communication

Humans and monkeys tend to match facial expressions, a phenomenon referred to as facial mimicry (Lundqvist and Dimberg, 1995; Mosher, Zimmerman and Gothard, 2011). This simple and strong linkage of facial perception to signal production suggests a "vertically" integrated (modular, feed forward) circuit. A candidate pathway with strong anatomical support proceeds from the face-perception system through the amygdala to cingulate area M3, and from there, directly to the facial nucleus (Morecraft, Stilwell–Morecraft and Rossing, 2004; Livneh *et al.*, 2012; Gothard, 2014). A potential signature of such a feed-forward, vertically-integrated module is a gradual transition of selectivity from perception to movement generation. Comparing activation profiles for face perception, social context, and face movement across a large number of regions of interest (ROIs, Fig. 4B), we found ROIs activated by perception but only weakly by context or movement generation, areas activated by context but not movement, and areas activated by movement with little sensitivity to perception or context. A region dorsal to the amygdala, and more weakly the lateral and medial geniculate nuclei, showed mixed sensitivity to face perception and movement but no sensitivity to context. To capture the distribution of selectivity across ROIs, we performed multidimensional scaling (Fig. 5A). The overall pattern of selectivity indicates strong specialization for either of the three functions—face perception, social context, and face movement—but shows little evidence for gradual transitions between them.

A second signature of vertically-integrated modules is strong functional connectivity between areas of different specialization within the module. Functional connectivity analysis (Fig. 5A) provided little evidence for a strong connection of face perception and movement areas via an amygdala-M3 bridge, but instead suggested thalamic and striatal routes. Furthermore, strong functional coupling was found between face perception areas, between cortical and subcortical facial motor areas, and within medial frontal areas. Thus, rather than providing evidence for vertical integration across a social-perception-to-social-response pathway, our functional connectivity analysis suggests the dominance of "horizontal" integration within cortex—that is, within the levels of face perception, of second-person cognition, and of face movement— without strong connections between levels. This view was independently supported by a spatial independent component analysis (ICA, see STAR Methods). Comparing ICAs across subjects, four large-scale networks were consistently observed (Fig. 5B): The first captured activation during social stimulus presentation. The second captured a nonspecific motor network comprising broadly distributed frontal, cingulate and insular areas, typically suppressed during social stimulus presentation (Fig. 1A) and facial movement (Figs. 3A,B). The third network comprised lateral and medial facial motor areas, M1, sM1, VLPMC, AI, DI, and M3, but not the more anterior social-context sensitive cluster. The fourth network included this cluster and dorsolateral prefrontal areas, and may exert executive control. Together, these results suggest multiple functional

networks, with at least three functional specializations, engaged during facial communication (Fig. 5C).

Discussion

In this first fMRI study of the neural circuits of social communication in primates, we found a network of cortical and subcortical areas controlling a facial signal, the lipsmack. By systematically monitoring the entire brain, we established which candidate areas—shown by anatomy to project to the facial nucleus (Morecraft, Stilwell–Morecraft and Rossing, 2004), by microstimuation to support vocalization (Jürgens, 2009), or by electrophysiology to contain facial expression-related cells (Ferrari *et al.*, 2003; Petrides, Cadoret and Mackey, 2005; Coudé *et al.*, 2011; Livneh *et al.*, 2012; Hage and Nieder, 2013; cf. Caruana *et al.*, 2011)—are engaged in routine communication. It is widely held that affective signals are broadly conserved in primates while the more derived, voluntary systems, including speech, are unique to humans (e.g. Arbib, 2005; Wheeler and Fischer, 2012). Moreover, medial and lateral areas of facial motor control are expected to functionally dissociate (e.g. Morecraft, Stilwell–Morecraft and Rossing, 2004; Gothard, 2014; Müri, 2016; cf. Hage and Nieder, 2016). Contrary to these (and our own) expectations, the network supporting the generation of a canonical primate emotional expression—the lipsmack (van Hooff, 1967)—was not centered on medial cortical regions but, in fact, most strongly activated lateral motor and premotor areas with homologies to human speech control.

Yet a medial specialization, relevant for facial expression, does exist: a cluster of areas sensitive to 2nd person context, including parts of anterior cingulate gyrus, anterior cingulate sulcus, medial prefrontal cortex and dorsomedial prefrontal cortex. It is centered just anterior to cingulate face motor area M3 and next to areas implicated in high-level social cognition (Schwiedrzik *et al.*, 2015; Sliwa and Freiwald, 2017), revealing a hitherto unknown functional parcellation of macaque medial prefrontal cortex (Yoshida *et al.*, 2011, 2012; Chang, Gariépy and Platt, 2012; Haroush and Williams, 2015). This cluster may include a homolog to the region activated by 2nd person contexts in humans (Schilbach *et al.*, 2006).

Face motor areas were distributed, exhibited functional specializations, and were functionally interacting. These are all characteristics the face motor areas share with the face-perception network (e.g. Fisher and Freiwald, 2015; Schwiedrzik *et al.*, 2015). Yet while intra-network interactions were strong, functional interactions between the two networks were limited. Even when face-perception and face-motor areas were near one another, as in the frontal lobe, functional interactions were minimal. Thus we found little evidence for the hypothesis of a "vertical" perception-to-production pathway for facial signals. Vertical interactions might transiently occur between networks whose dominant mode of operation is within-network.

Given the apparent simplicity of the lipsmack motor pattern (Ghazanfar *et al.*, 2012), activation of an entire multi-area face motor network was surprising. Activation was not restricted to medial facial representations in the anterior cingulate, as predicted for merely affective signals (e.g. Morecraft, Stilwell–Morecraft and Rossing, 2004; Gothard, 2014; Hage and Nieder, 2016; Müri, 2016), but instead prominently included lateral facial

representations, associated in humans with voluntary movements including speech. Importantly, these activations included putative macaque homologues of Broca's area. Prior research has variously identified these homologues as F5/anteriolateral BA6, BA44, and/or BA45 (Petrides, Cadoret and Mackey, 2005; Neubert *et al.*, 2014). Portions of our motor-associated ventrolateral activations, specifically our ALPMC and OMC ROIs, overlap both F5 and BA44. Our data thus refute the argument that primate facial communication circuits could not have served as a precursor to human speech (e.g. see Arbib, 2005). In fact, lipsmack articulatory patterns are specialized for communication (Shepherd, Lanzilotto and Ghazanfar, 2012), and their rhythms are reminiscent of human speech syllables (Ghazanfar *et al.*, 2012; Ghazanfar and Takahashi, 2014). We thus show that, like human speech areas, macaque homologues in lateral frontal cortex support communicative articulation. This similarity in cortical function between humans and old-world monkeys suggests that brain networks supporting volitional communication have a common evolutionary origin which arose well prior to the hominin radiation.

STAR*Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Stephen V. Shepherd (stephen.v.shepherd@gmail.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects: Data was obtained from 10 subadult male rhesus macaques (*Macaca mulatta*), each about 4 years old. All subjects were maternally-reared, transfering to juvenile groups at weaning and into adult colonies at 2 years. The laboratory colony features auditory and visual contact with conspecific males, and study subjects were generally pair-housed. All subjects were fitted with head restraint prostheses using standard lab approaches (Fisher and Freiwald, 2015). Procedures conformed to applicable regulations and to NIH guidelines per 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.

METHOD DETAILS

Stimuli: Expressive behaviors were elicited by the first author from 7 stimulus individuals, male rhesus macaques of varying age and familiarity relative to the subjects. Expressions were predominantly lipsmacks, but included other expressions such as silent bared teeth displays, open mouth stares, and mere sustained gaze. Three cameras recorded the stimulus monkeys: two flanking the axis of interaction and one capturing direct 'eye contact' via a half-silvered mirror. Stimulus videos thus appeared to be oriented directly toward (2nd person context, or if you prefer, 2nd monkey) or away from (3rd person/monkey context) the subject (Schilbach *et al.*, 2006).

We extracted from each of these videos two 10-second video clips in which different social signals were produced in a consistent direction, while expressive content and intensity was allowed to vary. This duration seemed, to human observers, long enough to create a strong

sense of communicative intent by the depicted monkey, but short enough that a lack of responsiveness to a viewer could be overlooked. Additional videos were produced by digitally phase-scrambling the originals, using a random constant phase rotation across frames to preserve spatial frequencies and motion content while disrupting visual form. In total, we generated 12 videos each for the 7 stimulus individuals, comprising 3 different perspectives on 2 dynamic sequences of social expression as well as matched, phase-scrambled controls. Stimulus videos were thus were either facing toward the subject (2nd person), to his left or right (3rd person, high-order visual control), or were phase-scrambled (nonsocial, low-order visual control). Finally, we stitched videos together so that they appeared sparsely, with ample downtime between presentations and with minimal repeats. Each experimental run consisted of one pseudorandomly selected video from each of the 7 subjects, in pseudorandom order, with 14-second blank periods between each video. The typical recording session consisted of 12 runs with no videos repeated, and would thus include 14 videos of 2nd-person perspectives, 28 of 3rd-person perspectives, and 42 of phase-scrambled controls.

Half of the subjects (S06-10) additionally performed a brief oculomotor task with alternating periods in which they either sustained gaze toward a static visual target or rapidly shifted gaze to look at random targets to garner a fluid reward. Face movements were also recorded in this purely nonsocial task.

Finally, face and body responsive regions were mapped in each subject using a variants of a standard localizer task including static images of primate faces, nonprimate objects, and primate bodies (Fisher and Freiwald, 2015).

For details on subjects' participation in experiments, see Table S1.

fMRI Imaging: Each subject was previously implanted with an MR-compatible headpost (Ultem; General Electric Plastics) using MR-compatible zirconium oxide ceramic screws (Thomas Recording), medical cement (Metabond and Palacos) and standard anesthesia, asepsis, and post-operative treatment procedures (Wegener, 2004). Imaging was performed in a 3T Siemens Tim Trio scanner. Anatomical images were constructed by averaging high-resolution volumes gathered under anesthesia (ketamine and isoflurane, 8 mg/kg and 0.5-2% respectively) in a T1-weighted 3D inversion recovery sequence (Magnetization-Prepared RApid Gradient Echo, or MPRAGE, 0.5mm isometric) with a custom single-channel receive coil. Functional images were gathered using an AC88 gradient insert (quadrupling the scanner's slew rate) and 8-channel receive coil in echoplanar imaging sequences (EPI: TR 2 s, TE 16 ms, 1 mm^3 voxels, horizontal slices, 2x GRAPA acceleration). To increase the contrast-to-noise-ratio of functional recordings, we injected monocrystaline iron oxide nanoparticles (feraheme at 8-12 mg/kg for subjects S01-04 or Molday ion at 6-9 mg/kg for subjects S05-10) into the saphenous vein at the beginning of each scan session; in IRON-fMRI, decreases in signal strength correspond to increases in blood flow and hence to neural activity (Vanduffel, Fize and Mandeville, 2001).

Imaging Sessions: During each functional imaging session, subjects sat in sphinx position with heads fixed at isocenter. Eye position was measured at 60 Hz using a

commercial eye monitoring system (Iscan), and facial movements were captured at 30 Hz using a MR-compatible infrared video camera (MRC). Videos were played in Presentation (Neurobehavioral Systems) stimulus control software. The sparse presentation of social stimuli was designed to maximize the impact of each social interaction while minimizing the subject's exposure to each stimulus (and hence habituation). In some sessions, subjects received juice rewards: either sparsely and randomly or for looking toward the video screen (they were never required to fixate). This helped us to dissociate communicative from noncommunicative facial motion. In other sessions, the juice tube was removed to give an unobstructed view of the subject's mouth and to remove ingestive incentives for orofacial movement (helping us identify naturalistic communicative movements). As described above, half the subjects also participated in an additional nonsocial task during imaging sessions (see Supplemental Table S1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Subject Video Scoring: Facial movements produced by our monkey subjects were recorded in infrared and scored by computer-assisted video analysis, supplemented by targeted manual annotation (Fig. S1B). Specifically, the subject's facial behavior in each session was monitored through an MR-compatible camera, which was synched to the stimuli through the placement of a small plastic bicycle mirror behind the subject's head. Subject's facial behavior was extracted through simple computer vision techniques (Matlab 2015b, custom code) which evaluated local changes in luminance distribution over a manually-defined face-movement window; ovals around each eye were masked from consideration. Additionally, a background-movement window that excluded the face (and the bicycle mirror) attempted to control for body and limb movements; visibility varied across sessions. Finally, the first author and an undergraduate volunteer manually (and blindly, in random order) scored each 2-second period of video as depicting mostly (a) drinking (b) lipsmacking (c) other expressive movements including yawns (d) mixed/miscellaneous/manual movement (e) relative stillness.

fMRI Analysis: MRI data analysis was performed in FreeSurfer and FS-FAST (http://surfer.nmr.mgh.harvard.edh/) and with custom Linux shell and Matlab scripts. Raw image volumes were preprocessed through slice-by-slice motion- and slice-time correction (AFNI), aligned to anatomicals and unwarped using JIP Analysis Toolkit (http://www.nmr.mgh.harvard/~jbm/jip/). smoothed (Gaussian kernel, 1mm FWHM), and masked. Data were quadratically detrended in time and analyzed with respect to video stimulus type and facial movement (computer-scored or manually-categorized), generating significance maps by calculating the mean and variance of the response within each voxel to these conditions. GLMs were constructed with binary variables representing stimulus conditions and continuous (computer-scored) variables representing subjects' facial and background movements. Contrasts examined the response to social stimuli (response to social less that to scrambled video), the response to 2nd vs. 3rd person context (response to direct-gaze less that to averted-gaze video), and the neural correlates of facial movement (indexed by response to the computer-scored continuous variable). To further parse types of motor response, we substituted binary hand-scored variables for the computer-scored continuous variables, contrasting lipsmack and drinking with stillness (responses during lipsmack/drinking less

that during stillness). As an alternate means of probing communicative versus noncommunicative movement, we broke apart computer-scored facial responses by context (e.g. computer-scored facial movement during social video, during scrambled video, or during a nonsocial task). Individual subjects' data were registered to the Reveley/McLaren atlases (McLaren *et al.*, 2009; Reveley *et al.*, 2016) and across hemispheres using FSL's FLIRT and FNIRT alignment tools. Regional differences were summarized by averaging unsmoothed voxelwise data, for each subject and at each time point, within each post-hoc ROI as defined in common McLaren space and exported to each subject's individual anatomical space. Data were aggregated across subjects' hemispheres (and ROI, see Figure S2) using weighted least squares estimation; voxelwise data was FDR-corrected for multiple comparisons. The extent to which ROI signal correlated with facial movement across conditions (Fig. 3D) was tested for significance using a bootstrap procedure on WLS differences between the beta value subscores for facial movement within specific contexts versus the beta value for overall computer-scored facial movement (10 subjects) and for the WLS difference between hand-categorized drinking and lipsmack beta values (in 4 subjects: S01-02 and S06-07).

Network Analysis: Spatial ICA was conducted using MELODIC (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MELODIC) on the fMRI signal residuals after a blank FS-FAST analysis consisting only of quadratic detrending and the nuisance regression of estimated head motion. Individual ICs were compared across subjects using factor analysis with 'Promax' rotation (Matlab 2015b) based on the covariance of spatial components in a standardized space, with the most strongly shared factors kept as 'consensus ICs', reconstructed by signed averaging of all spatial components with common-factor loadings greater than 0.5. Functional connectivity analysis was conducted on the fMRI signal residuals of the full FS-FAST analysis, including stimulus video type and autocoded face and background motion parameters as well as motion nuisance regressors. First, correlation coefficients were calculated within each session between each pair of ROI in each hemisphere. Second, we took the median correlation coefficient for each ROI in each hemisphere across runs, tested significance by ttest of the Fischer-transformed correlation values across the 20 hemispheres, and finally used the median correlation value across hemispheres for display. ROI coordinates were illustrated through multidimensional scaling of their signed responses to task conditions, weighting equally the contrasts of social versus scrambled, of 2nd- versus 3rd-person perspective, of background movement, and of facial movement (as aggregated across context-and motion-type subscores). Finally, the color of each ROI was chosen based on the absolute level of modulation of the ROI due to seen faces versus scrambles (red), due to 2nd versus 3rd person social perspective (blue), or due to produced facial movement (green), relatively to the maximal observed modulation across recorded ROI for each contrast.

DATA AND SOFTWARE AVAILABILITY

Further information and requests for software and datasets should be directed to and will be fulfilled by the Lead Contact, Stephen V. Shepherd (stephen.v.shepherd@gmail.com).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Andrew RJ (1963) 'Evolution of facial expression.', Science (New York, NY), 142(3595), pp. 1034–1041.

Arbib MA (2005) 'From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics.', The Behavioral and brain sciences, 28(2), pp. 105–24; discussion 125-67. [PubMed: 16201457]

Ballesta S and Duhamel J-R (2015) 'Rudimentary empathy in macaques' social decision-making', Proceedings of the National Academy of Sciences, 112(50), pp. 15516–15521. doi: 10.1073/pnas.1504454112.

Caruana F, Jezzini A, Sbriscia-Fioretti B, Rizzolatti G, Gallese V and Neuroscienze D (2011) 'Emotional and social behaviors elicited by electrical stimulation of the insula in the macaque monkey', Current Biology. Elsevier Ltd, 21(3), pp. 195–199. doi: 10.1016/j.cub.2010.12.042. [PubMed: 21256020]

Chang SWC, Gariépy J-F and Platt ML (2012) 'Neuronal reference frames for social decisions in primate frontal cortex', Nature Neuroscience, 16(December), pp. 243–50. doi: 10.1038/nn.3287. [PubMed: 23263442]

Coudé G, Ferrari PF, Rodà F, Maranesi M, Borelli E, Veroni V, Monti F, Rozzi S and Fogassi L (2011) 'Neurons controlling voluntary vocalization in the macaque ventral premotor cortex', PLoS ONE. Edited by Bartolomucci A, 6(11), p. e26822. doi: 10.1371/journal.pone.0026822. [PubMed: 22073201]

Dal Monte O, Piva M, Morris JA and Chang SWC (2016) 'Live interaction distinctively shapes social gaze dynamics in rhesus macaques', Journal of Neurophysiology, 116(4), pp. 1626–1643. doi: 10.1152/jn.00442.2016. [PubMed: 27486105]

Darwin C (1872) The Expression of the Emotions in Man and Animals. London: John Murray, Albemarle Street.

Dimberg U, Thunberg M and Elmehed K (2000) 'Unconscious facial reactions to emotional facial expressions', Psychological Science, 11(1), pp. 86–89. [PubMed: 11228851]

Fernandez-Dols J-M and Crivelli C (2013) 'Emotion and Expression: Naturalistic Studies', Emotion Review, 5(1), pp. 24–29. doi: 10.1177/1754073912457229.

Ferrari PF, Gallese V, Rizzolatti G and Fogassi L (2003) 'Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex', European Journal of Neuroscience. John Wiley & Sons, 17(8), pp. 1703–1714. doi: 10.1046/j.1460-9568.2003.02601.x. [PubMed: 12752388]

Fisher C and Freiwald WA (2015) 'Contrasting Specializations for Facial Motion within the Macaque Face-Processing System', Current Biology. Elsevier Ltd, 25(2), pp. 261–266. doi: 10.1016/j.cub.2014.11.038. [PubMed: 25578903]

Ghazanfar AA and Takahashi DY (2014) 'Facial expressions and the evolution of the speech rhythm.', Journal of cognitive neuroscience, 26(6), pp. 1196–207. doi: 10.1162/jocn_a_00575. [PubMed: 24456390]

Ghazanfar AA, Takahashi DY, Mathur N and Fitch WT (2012) 'Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics.', Current biology : CB, 22(13), pp. 1176–82. doi: 10.1016/j.cub.2012.04.055. [PubMed: 22658603]

Gothard KM (2014) 'The amygdalo-motor pathways and the control of facial expressions.', Frontiers in neuroscience, 8(March), p. 43. doi: 10.3389/fnins.2014.00043. [PubMed: 24678289]

Grimaldi P, Saleem KS and Tsao DY (2016) 'Anatomical Connections of the Functionally Defined "Face Patches" in the Macaque Monkey', Neuron. Elsevier Inc., 90(6), pp. 1325–1342. doi: 10.1016/j.neuron.2016.05.009. [PubMed: 27263973]

Hage SR and Nieder A (2013) 'Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations.', Nature communications. Nature Publishing Group, 4, p. 2409. doi: 10.1038/ncomms3409.

Hage SR and Nieder A (2016) 'Dual Neural Network Model for the Evolution of Speech and Language', Trends in Neurosciences. Elsevier Ltd, 39(12), pp. 813–829. doi: 10.1016/j.tins.2016.10.006. [PubMed: 27884462]

Haroush K and Williams ZM (2015) 'Neuronal Prediction of Opponent's Behavior during Cooperative Social Interchange in Primates', Cell. Elsevier Inc., 160(6), pp. 1233–1245. doi: 10.1016/j.cell.2015.01.045. [PubMed: 25728667]

van Hooff JARAM (1967) 'The Facial Displays of the Catarrhine Monkeys and Apes', in Morris D (ed.) Primate Ethology. 2006th edn. London: Wiedenfeld & Nicolson, pp. 7–68.

Hopf HC, Müller-Forell W and Hopf NJ (1992) 'Localization of emotional and volitional facial paresis.', Neurology, 42(10), pp. 1918–23. [PubMed: 1407573]

Jurgens U (2009) 'The neural control of vocalization in mammals: a review.', Journal of voice : official journal of the Voice Foundation, 23(1), pp. 1–10. doi: 10.1016/j.jvoice.2007.07.005. [PubMed: 18207362]

Livneh U, Resnik J, Shohat Y and Paz R (2012) 'Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex.', Proceedings of the National Academy of Sciences of the United States of America, 109(46), pp. 18956–61. doi: 10.1073/pnas.1207662109. [PubMed: 23112157]

Lundqvist L-O and Dimberg U (1995) 'Facial expressions are contagious.', Journal of Psychophysiology, 9, pp. 203–11.

Marciniak K, Atabaki A, Dicke PW and Thier P (2014) 'Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus.', eLife, 3(2011), p. e03222. doi: 10.7554/eLife.03222.

McLaren DG, Kosmatka KJ, Oakes TR, Kroenke CD, Kohama SG, Matochik JA, Ingram DK and Johnson SC (2009) 'A population-average MRI-based atlas collection of the rhesus macaque', NeuroImage, 45(1), pp. 52–59. doi: 10.1016/j.neuroimage.2008.10.058. [PubMed: 19059346]

Moeller S, Freiwald WA and Tsao DY (2008) 'Patches with links: a unified system for processing faces in the macaque temporal lobe.', Science (New York, N.Y.), 320(5881), pp. 1355–9. doi: 10.1126/science.1157436.

Morecraft RJ, Mcneal DW, Stilwell–morecraft KS, Gedney M, Ge J, Schroeder CM and Hoesen GWVAN (2007) 'Amygdala Interconnections with the Cingulate Motor Cortex in the Rhesus Monkey', J Comp Neurol, 165(April 2006), pp. 134–165. doi: 10.1002/cne.

Morecraft RJ, Stilwell-Morecraft KS and Rossing WR (2004) 'The Motor Cortex and Facial Expression: New Insights From Neuroscience', The Neurologist, 10(5), pp. 235–249. doi: 10.1097/01.nrl.0000138734.45742.8d. [PubMed: 15335441]

Mosher CP, Zimmerman PE and Gothard KM (2011) 'Videos of conspecifics elicit interactive looking patterns and facial expressions in monkeys.', Behavioral neuroscience, 125(4), pp. 639–52. doi: 10.1037/a0024264. [PubMed: 21688888]

Müri RM (2016) 'Cortical control of facial expression', Journal of Comparative Neurology, 524(8), pp. 1578–1585. doi: 10.1002/cne.23908. [PubMed: 26418049]

Neubert F-X, Mars RB, Thomas A, Sallet J and Rushworth MFS (2014) 'Comparison of Human Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal Cortex', Neuron. Elsevier Inc., 81(3), pp. 700–713. doi: 10.1016/j.neuron.2013.11.012. [PubMed: 24485097]

Petrides M, Cadoret G and Mackey S (2005) 'Orofacial somatomotor responses in the macaque monkey homologue of Broca's area.', Nature, 435(7046), pp. 1235–8. doi: 10.1038/nature03628. [PubMed: 15988526]

Reveley C, Gruslys A, Ye FQ, Glen D, Samaha J, Russ BE, Saad Z, Seth AK, Leopold DA and Saleem KS (2016) 'Three-Dimensional Digital Template Atlas of the Macaque Brain', Cerebral Cortex, pp. 1–15. doi: 10.1093/cercor/bhw248. [PubMed: 25139941]

Schilbach L (2016) 'Towards a second-person neuropsychiatry', Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1686), p. 20150081. doi: 10.1098/rstb.2015.0081.

Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T and Vogeley K (2013) 'Toward a second-person neuroscience.', The Behavioral and brain sciences, 36(4), pp. 393–414. doi: 10.1017/S0140525X12000660. [PubMed: 23883742]

Schilbach L, Wohlschlaeger AM, Kraemer NC, Newen A, Shah NJ, Fink GR and Vogeley K (2006) 'Being with virtual others: Neural correlates of social interaction', Neuropsychologia. Elsevier, 44(5), pp. 718–730. doi: 10.1016/j.neuropsychologia.2005.07.017. [PubMed: 16171833]

Schwiedrzik CM, Zarco W, Everling S and Freiwald W. a. (2015) 'Face Patch Resting State Networks Link Face Processing to Social Cognition', PLOS Biology, 13(9), p. e1002245. doi: 10.1371/journal.pbio.1002245. [PubMed: 26348613]

Shepherd SV (2010) 'Following gaze: Gaze-following behavior as a window into social cognition.', Frontiers in integrative neuroscience, 4(March), p. 5. doi: 10.3389/fnint.2010.00005. [PubMed: 20428494]

Shepherd SV, Lanzilotto M and Ghazanfar AA (2012) 'Facial muscle coordination in monkeys during rhythmic facial expressions and ingestive movements.', The Journal of neuroscience : the official journal of the Society for Neuroscience, 32(18), pp. 6105–16. doi: 10.1523/JNEUROSCI.6136-11.2012. [PubMed: 22553017]

Sliwa J and Freiwald WA (2017) 'A dedicated network for social interaction processing in the primate brain', Science, 356(6339), pp. 745–749. doi: 10.1126/science.aam6383. [PubMed: 28522533]

Stemmann H and Freiwald WA (2016) 'Attentive Motion Discrimination Recruits an Area in Inferotemporal Cortex', Journal of Neuroscience, 36(47), pp. 11918–11928. doi: 10.1523/JNEUROSCI.1888-16.2016. [PubMed: 27881778]

Tsao DY, Moeller S and Freiwald WA (2008) 'Comparing face patch systems in macaques and humans.', Proceedings of the National Academy of Sciences of the United States of America, 105(49), pp. 19514–19519. doi: 10.1073/pnas.0809662105. [PubMed: 19033466]

Tsao DY, Schweers N, Moeller S and Freiwald WA (2008) 'Patches of face-selective cortex in the macaque frontal lobe.', Nature neuroscience, 11(8), pp. 877–9. doi: 10.1038/nn.2158. [PubMed: 18622399]

Vanduffel W, Fize D and Mandeville JB (2001) 'Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys', Neuron, 32, pp. 565–577. [PubMed: 11719199]

Wegener D (2004) 'The influence of sustained selective attention on stimulus selectivity in macaque visual area MT', The Journal of Neuroscience, 24(27), pp. 6106–14. doi: 10.1523/JNEUROSCI.1459-04.2004. [PubMed: 15240802]

Wheeler BC and Fischer J (2012) 'Functionally referential signals: A promising paradigm whose time has passed.', Evolutionary anthropology. Wiley Subscription Services, Inc., A Wiley Company, 21(5), pp. 195–205. doi: 10.1002/evan.21319. [PubMed: 23074065]

Yoshida K, Saito N, Iriki A and Isoda M (2011) 'Representation of Others' Action by Neurons in Monkey Medial Frontal Cortex', Current Biology. Elsevier Ltd, 21(3), pp. 249–253. doi: 10.1016/j.cub.2011.01.004. [PubMed: 21256015]

Yoshida K, Saito N, Iriki A and Isoda M (2012) 'Social error monitoring in macaque frontal cortex', Nature Neuroscience. Nature Publishing Group, 15(9), pp. 1307–1312. doi: 10.1038/nn.3180. [PubMed: 22864610]

Highlights:

- Facial perception and facial movement activate non-overlapping networks.

- Face-to-face interaction recruits medial prefrontal cortex.

- Expression activates medial and ingestion lateral parts of a shared network.

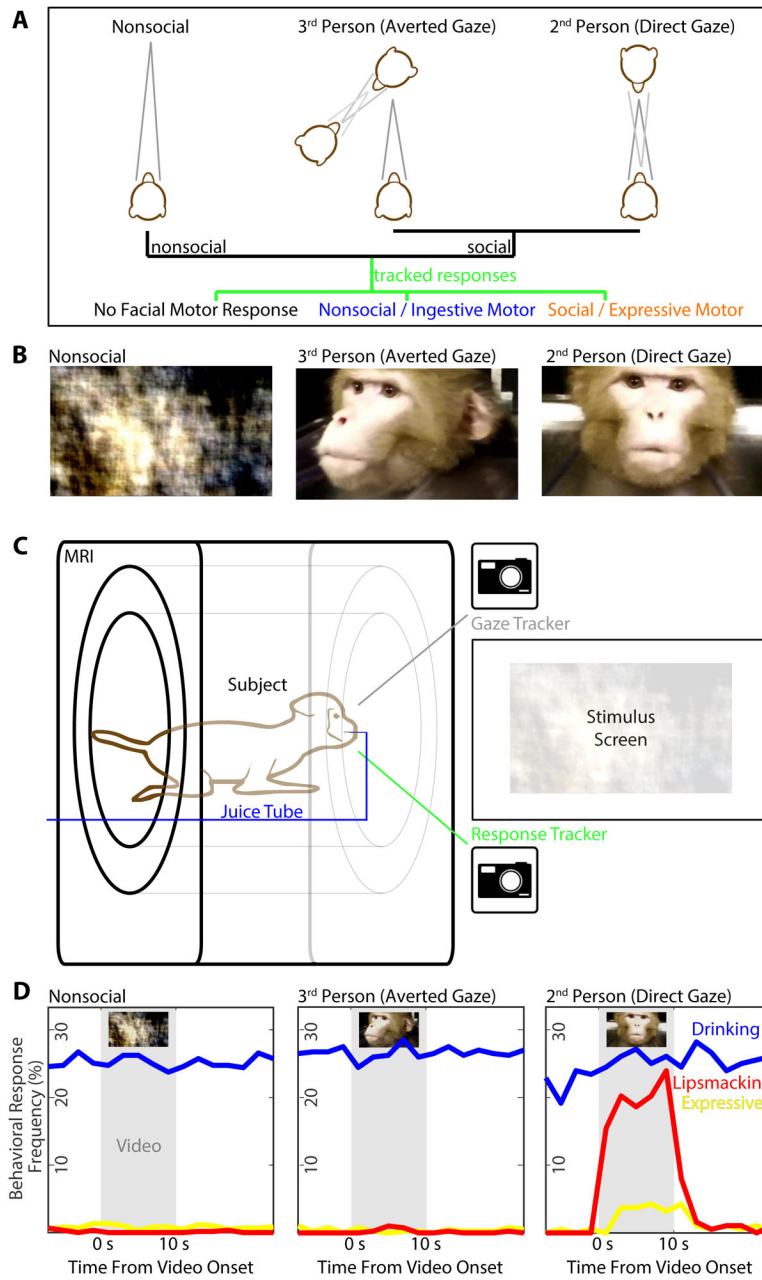- This facial motor network includes homologs of Broca's area.

Figure 1: Simulating Interactions During Monkey fMRI.
(**A**) Schematic of visual stimuli (top) an observer (middle) might perceive and respond to (bottom). Stimuli can be non-social (left), social directed away (middle) or towards (right) the observer, who might not respond or generate a social or an (unrelated) non-social motor response. (**B**) Sample frames of three categories of video (left to right: phase-scrambled, socially-expressive dynamic faces directed away or toward subject); for stimulus details, see Methods and Figure S1. (**C**) Schematic of setup with subject, seated inside the MRI scanner, viewing videos while their gaze and facial movements were recorded by infrared video cameras. (**D**) Mean response rates of drinking (blue), lipsmacking (red), and other expressive facial movements (yellow) of ten subjects as a function of stimulus context (left to right) and

time relative to the period of video display (light grey). While drinking behaviors were observed regardless of social context and video presentation, lipsmacks and other expressive behaviors were selectively elicited during 2nd person video presentation.
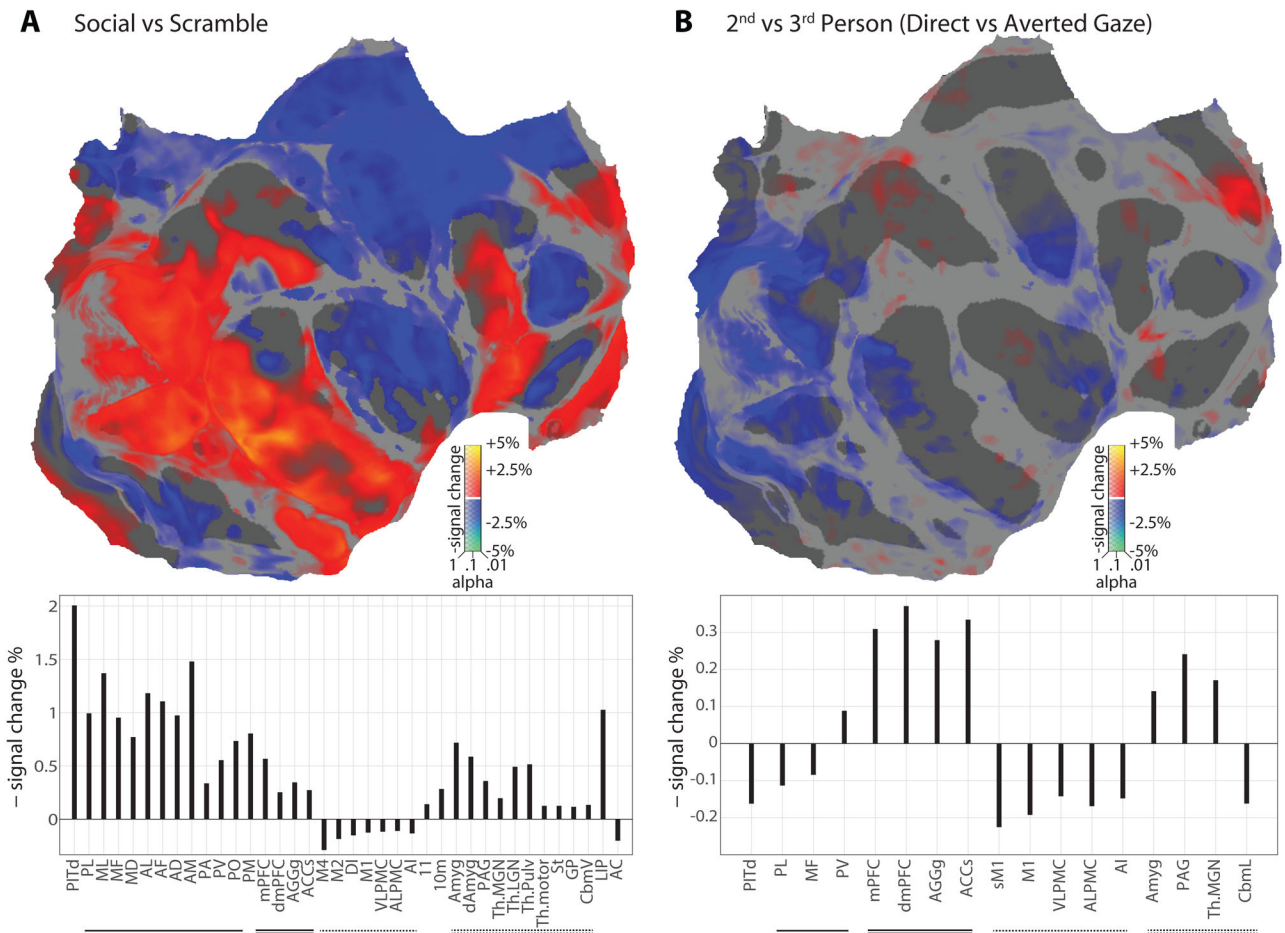
Figure 2: Neural Responses to Social Stimuli and to Interactive Contexts.
(**A**, top) Digitally flattened map of monkey cortex, in which the occipital pole is toward the left (cut along the calcarine), the temporal pole is at bottom right, the anterior pole is to the right, and the medial wall is in the upper right (cut across the mid-cingulate and posterior to the ascending limb). Areas with increased/decreased activity during social video versus scrambled video are shown in hot/cool colors respectively, scaled according to percent signal change; significance (corrected for multiple comparisons by estimating the false discovery rate) is indicated by opacity. Note that in IRON-fMRI, negative signal change corresponds to increased activity. Significantly modulated ROI are labeled. (**A**, bottom) Bar graph of significant ROI responses to social versus scrambled video, including both cortical and subcortical regions. Upward bars indicate decreased signal, and hence increased blood flow and brain activity. Activity changes were calculated by weighted least squares estimation across all subjects and brain hemispheres (20 hemispheres total); significance was not corrected for multiple comparisons. Face patches are underline, medial decision-associated are doubly underlined, facial motor regions are dash-lined, and subcortical regions are double dash-lined. Key regions activated by social stimuli include form-selective visual cortex, the temporal and frontal face patch systems, and medial frontal regions. (**B**, top) Flatmap of increased/decreased activity during 2nd person social video (direct gaze) versus 3rd person social video (averted gaze). Significantly affected ROI are labeled. (**B**, bottom)

Bar graph of significant ROI responses to 2nd-person versus 3rd-person video perspectives. Key regions activated by 2nd person context were largely restricted to medial frontal cortex near the caudalmost anterior cingulate.
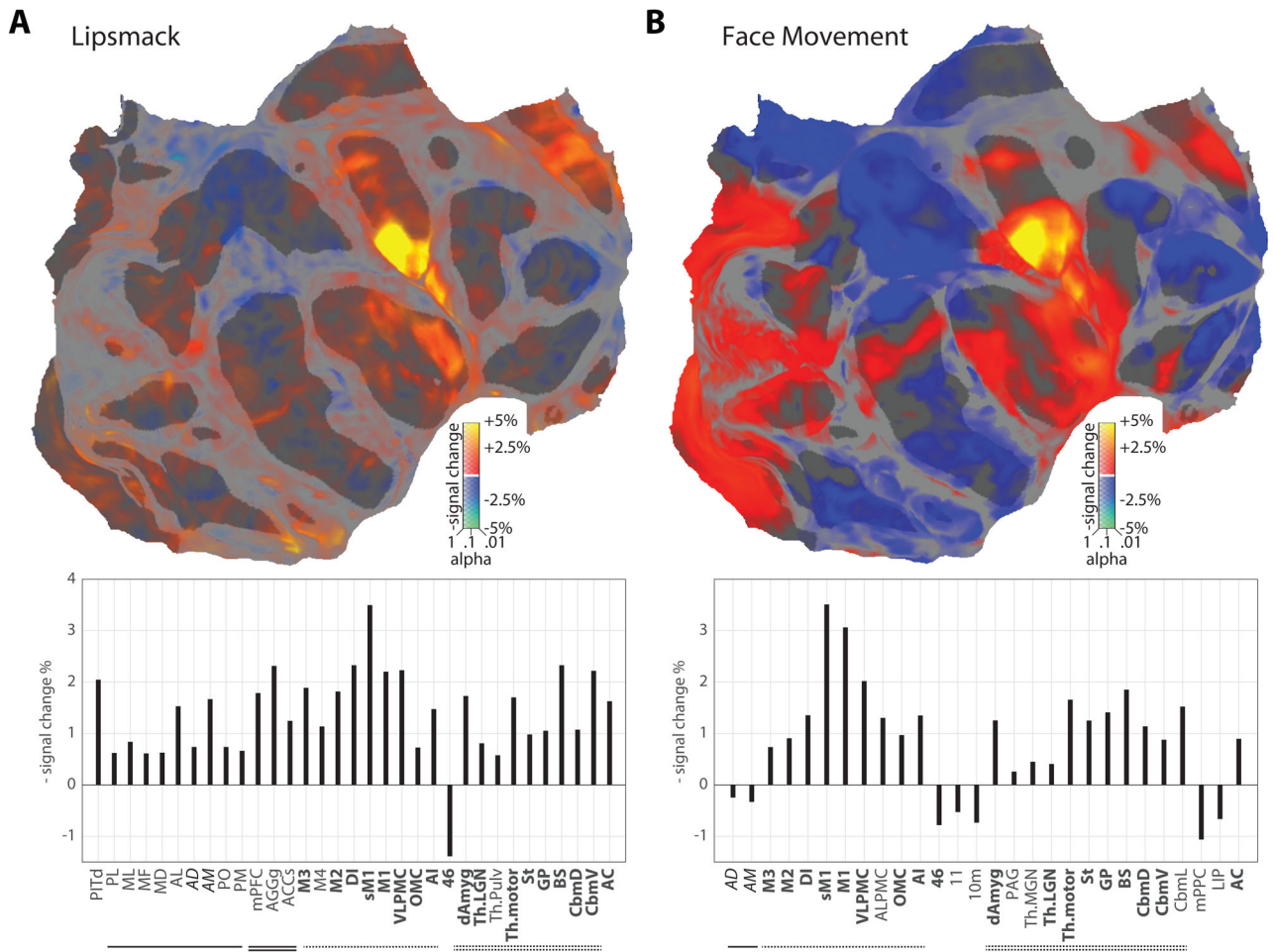
Figure 3: Neural Responses Associated with Communicative Facial Movement.
(**A**, top) Digitally flattened map of monkey cortex, in which the occipital pole is toward the left (cut along the calcarine), the temporal pole is at bottom right, the anterior pole is to the right, and the medial wall is in the upper right (cut across the mid-cingulate and posterior to the ascending limb). Areas with increased/decreased activity correlated with the presence of subject-produced 'lipsmacks' are shown in hot/cool colors respectively, scaled according to percent signal change; significance (corrected for multiple comparisons by estimating the false discovery rate) is indicated by opacity. Image-based motion estimates have been removed by nuisance regression (see Methods). Significantly modulated ROI are labeled. (**a**, bottom) Bar graph of significant ROI activity correlated with the presence of subject-produced 'lipsmacks', including both cortical and subcortical regions. Upward bars indicate decreased signal, and hence increased blood flow and brain activity. Activity changes were calculated by weighted least squares estimation across all subjects and brain hemispheres (20 hemispheres total); significance was not corrected for multiple comparisons. Face patches are underlined, medial decision-associated are doubly underlined, facial motor regions are dash-lined, and subcortical regions are doubledash-lined. Key regions activated include the lateral frontal cortex, supplementary motor cortex, and anterior cingulate motor cortex. Pronounced activity was also recorded in the brainstem facial nuclei, the motor

thalamus and the striatum. (**B**, top) Flatmap of increased/decreased activity correlated with general computer-scored face movement. Image-based motion estimates have been removed by nuisance regression (see Methods). Significantly affected ROI are labeled. (**B**, bottom) Bar graph of significant ROI activity correlated with general computer-scored face movement. As in (A), key regions of activation include the lateral frontal cortex, supplementary motor cortex, and anterior cingulate motor cortex, as well as the brainstem facial nuclei, the motor thalamus and the striatum.

A Motor activity in select ROI across diverse movement conditions

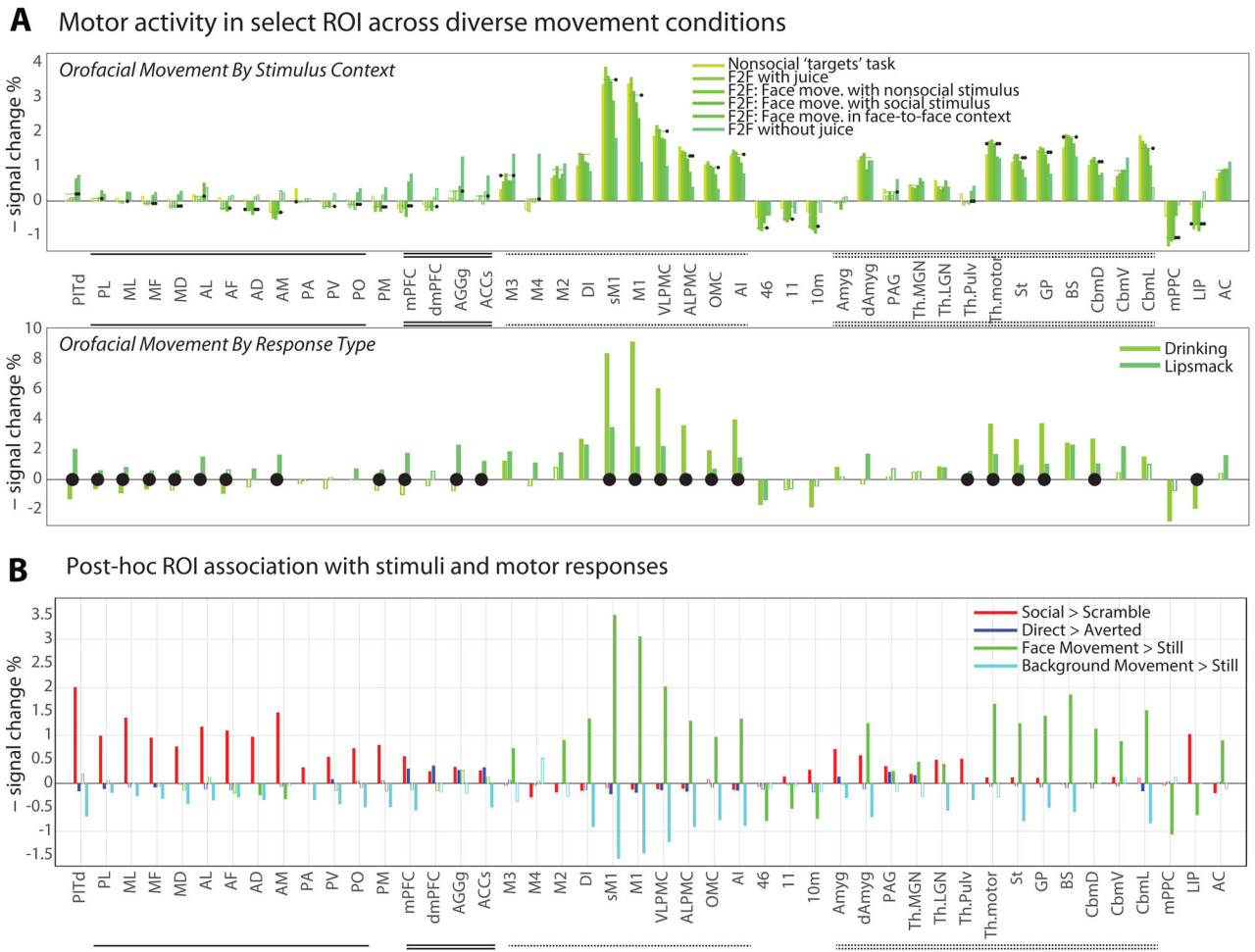**B** Post-hoc ROI association with stimuli and motor responses

Figure 4: Functional Selectivity of Regions of Interest to Types of Facial Movement, to 2nd Person Context, and to Social Perception.

(**A**) Neural correlates of produced social versus nonsocial movements were compared in two ways. Above, ROI correlation with computer-scored face movement are plotted under conditions ranging from minimal (left) to maximal (right) sociality; black dots indicated whether the subscore estimates were significantly different from the overall facial movement estimate (i.e. bar plot in Fig. 3B; bootstrap p<0.05 where indicated, see Methods). Below, for a subset of subjects (n=4), neural correlates of experimenter-scored drinking and lipsmack movements are contrasted; black dots indicate significance differences between drinking and lipsmack scores (bootstrap p<0.05 where indicated). These data show that lateral frontal motor and premotor cortex were recruited for both social communicative and nonsocial ingestive movements. Nonetheless, the data show relatively greater recruitment of medial structures during social and of lateral structures during nonsocial movements. (**B**) Activity change across all ROI reflected the presence of dynamic faces (red), 'interactive' context (blue), produced facial movement (green), and background movement (cyan); responses are significant *except* where bars are marked a white line (WLS across 20 hemispheres, no multiple comparisons correction across areas/conditions). For response reliability across subjects, see Figures S3 & S4.
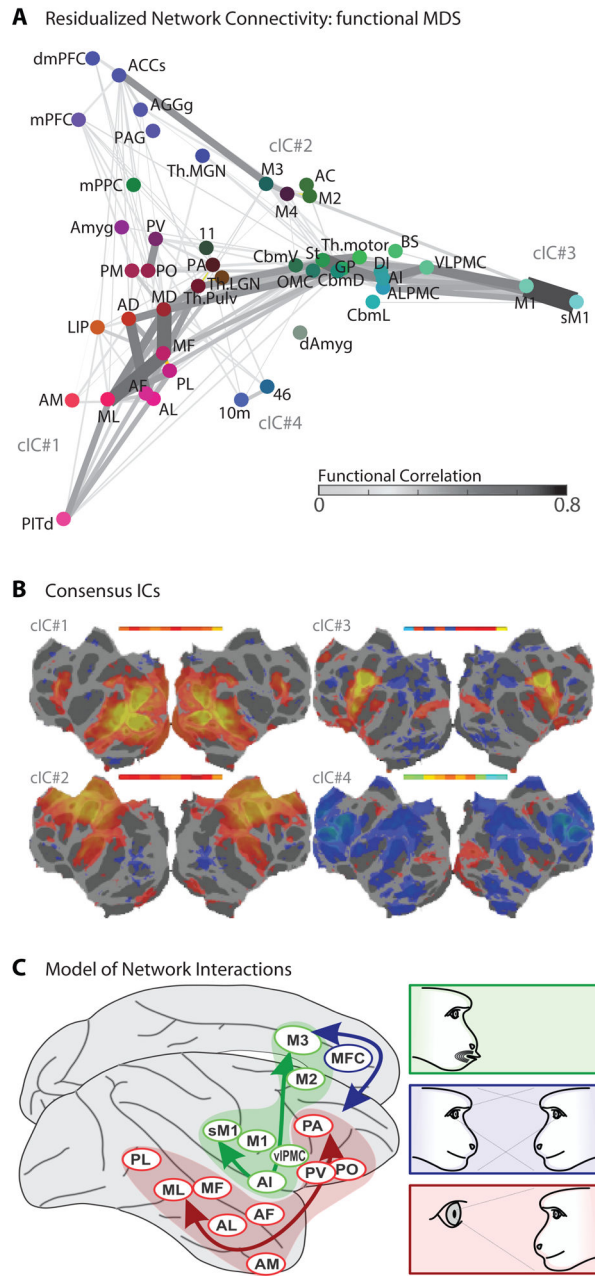
Figure 5: Network Connectivity During Simulated Social Interaction.
(**A**) A multidimensional scaling of ROI association with experimental variables (see Methods), linked to illustrate functional connectivity (computed from residuals in this experimental data after accounting for the experimental variables, see Methods). ROI node color indicates relatively stronger sensitivity to social percepts (red), social context (blue), or produced facial movement (green). (**B**) In a separate analysis, individual subjects' minimally-processed fMRI data was subjected to ICA, and common factors were identified across subjects. Networks corresponding to the face patch system (cIC#1) and to a general motor control network (cIC#2) were universally observed, while a face-specific motor control network including the brainstem facial nucleus (cIC#3) and a putative motor

inhibition network (cIC#4) were frequently identifiable. The maximal correlation of individual ICs with the common factor is illustrated in inset above the factor map. (**C**) Schematic of three independent cortical networks for social communication found in this study: face perception (red), 2nd person context (blue), and facial expression (green).