

Modulation of Perception and Brain Activity by Predictable Trajectories of Facial Expressions

N. Furl¹, N. J. van Rijsbergen², S. J. Kiebel¹, K. J. Friston¹,
A. Treves² and R. J. Dolan¹

¹Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, WC1N 3BG, UK and ²Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, 34104 Trieste, Italy

People track facial expression dynamics with ease to accurately perceive distinct emotions. Although the superior temporal sulcus (STS) appears to possess mechanisms for perceiving changeable facial attributes such as expressions, the nature of the underlying neural computations is not known. Motivated by novel theoretical accounts, we hypothesized that visual and motor areas represent expressions as anticipated motion trajectories. Using magnetoencephalography, we show predictable transitions between fearful and neutral expressions (compared with scrambled and static presentations) heighten activity in visual cortex as quickly as 165 ms poststimulus onset and later (237 ms) engage fusiform gyrus, STS and premotor areas. Consistent with proposed models of biological motion representation, we suggest that visual areas predictively represent coherent facial trajectories. We show that such representations bias emotion perception of subsequent static faces, suggesting that facial movements elicit predictions that bias perception. Our findings reveal critical processes evoked in the perception of dynamic stimuli such as facial expressions, which can endow perception with temporal continuity.

Keywords: face perception, facial emotion, fear perception, magnetoencephalography, premotor cortex, superior temporal sulcus, visual motion

Introduction

Face perception provides a model for investigating fundamental issues of neural coding. For example, faces in the natural environment are usually dynamic, and facial movements convey critical social signals including gaze direction, speech-related movements, and expressions of emotion and pain. This epitomizes a central challenge for research in biological and engineered visual systems: how can reliable and stable perception result from such dynamic input? In the case of facial movements, which express emotions, such percepts likely arise from representations within a dorsally projecting temporal lobe pathway, including the superior temporal sulcus (STS; Haxby et al. 2000). However, less is known about the neural mechanisms that the STS and associated visual areas employ to derive expression percepts from face dynamics (Calder and Young 2005). Many of the established findings come from studies of static faces, which manifest implied motion, but do not allow the visual system to represent naturalistic movement trajectories, which unfold over time.

One hypothesis afforded by use of dynamic stimuli is that the visual system employs anticipatory representations of sensory trajectories of facial attributes. This is based on theories which propose that perceptual representations (possibly encoded by neuronal interactions with attractor dynamics [Akrami et al.

2008]) depend on prediction of sensory states (Rao and Ballard 1999; Giese and Poggio 2003; Treves 2004; Friston 2005). Many of these models are motivated specifically to explain representation of stimulus dynamics (Giese and Poggio 2003; Jehee et al. 2006; Friston et al. 2008; Kiebel et al. 2008). Moreover, empirical evidence is mounting that the visual system may use such predictive coding at multiple levels (Murray et al. 2002; Bar et al. 2006; Summerfield et al. 2006, 2008; Schweidrzik et al. 2007; Summerfield and Kochlin 2008), beginning even in the retina (Hosoya et al. 2005). For low-level vision, primary visual cortex appears to extrapolate apparent motion trajectories, by “filling-in” trajectories through unseen stimulus positions (Muckli et al. 2005; Larsen et al. 2006; Sterzer et al. 2006). For higher level biological motion stimuli, some have proposed the STS predicts visual patterns (Giese and Poggio 2003; Kilner et al. 2007). More controversially, some suggest that predictions rely partly on representations in motor areas (Jeannerod 2001; Kilner et al. 2007) perhaps transmitted by mirror neurons (van der Gaag et al. 2007). These latter proposals attempt to explain evidence that STS and the motor system respond concurrently to body actions (Saygin et al. 2004; Calvo-Merino et al. 2006; Dayan et al. 2007) and to faces (Buccino et al. 2001; Sato et al. 2004; Montgomery and Haxby 2008). Importantly, predictability affects not just neural activity but also perception: predictable point-light bodily action sequences modulate perception of subsequent stimuli (Verfaillie and Daems 2002; Graf et al. 2007). However, no studies have directly addressed 1) whether predictive mechanisms operate in the STS, 2) whether these predictions contribute to perception of face expressions, and 3) the timing of prediction-related responses in different visual areas, especially with respect to well-known evoked components such as M100 and M170. Moreover, it has not yet been shown whether the motor system is preferentially responsive to predictable facial movements as usually encountered in our natural environment.

We addressed these questions using a combination of magnetoencephalography (MEG) and behavioral measures to examine effects of dynamic facial expressions, which varied in their predictability. Specifically, our participants viewed dynamic stimuli that resembled naturalistic transitions between fearful and neutral expressions. Evoked responses to these were compared with those of unpredictable scrambled transitions. These scrambled stimuli were random, unnatural, and lacked a coherent trajectory, although they were closely matched with the predictable stimuli for emotional image content and the final image presented. We expected that predictable transitions would engage visual areas including the STS, resulting in heightened activity in these areas, relative to scrambled stimuli. Indeed, we found that predictable

expression dynamics evoked very early effects in primary visual cortex (165 ms), followed by heightened activity in bilateral visual cortex, right posterior STS and posterior fusiform gyrus (237 ms). We also observed these effects in bilateral premotor cortex. Although the motor system has been observed in prior biological motion studies (Buccino et al. 2001; Jeannerod 2001; Sato et al. 2004; Saygin et al. 2004; Calvo-Merino et al. 2006; Dayan et al. 2007; Montgomery and Haxby 2008), we show motor activity specifically responsive to facial expression predictability. Additionally, we tested whether the sensory trajectories bias subsequent perception. On each trial, following presentation of the predictable stimuli, participants saw a static face (morphed midway between neutral and fearful) and rated this face for fearfulness. We show behaviorally that fear perception is biased in the direction predicted by the preceding trajectory. Thus, exposure to dynamic stimuli seems to prime the visual system to perceive facial expressions consistent with the cause of the immediately preceding sensory trajectory. Collectively, these results point to representations of expressions that encode sensory trajectories.

Materials and Methods

Participants

We measured MEG-evoked fields in 22 participants (8 females). Informed consent was obtained in accordance with procedures approved by the joint ethics committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology, London.

Design

The paradigm consisted of a series of trials, where participants viewed 2 successive stimuli (S1 and S2), separated by an interstimulus interval. For S1 presentations, we used morphed faces linearly interpolated between fearful and neutral expressions to construct image sequences that were predictable or scrambled. We also selected static S1 images from these morph continua (Fig. 1*a*).

For predictable S1 sequences, participants viewed 6 morphed images presented rapidly in order either from neutral to fearful (fear predictable) or from fearful to neutral (neutral predictable). These S1 sequences (360 ms in duration) appeared as animated natural expressions evolving in time and were predictable in the sense that the images followed a coherent movement trajectory. In contrast, for scrambled stimuli, we altered the fear- and neutral-predictable sequences so that the first 5 images were presented in a random order. Consequently, each scrambled sequence included the same images as a corresponding predictable sequence. The sixth image, the endpoint image, was also the same as in the corresponding predictable sequence. The scrambled sequences were constructed such that they never depicted any coherent expression trajectory and image transitions could not be predicted from the preceding transitions. We will hereafter refer to scrambled sequences which are “fear scrambled” if they are scrambled versions of fear-predictable sequences and “neutral scrambled” if they are scrambled versions of neutral-predictable sequences.

These fear- and neutral-predictable sequences and their scrambled versions conform to a 2×2 factorial design (Fig. 1*b*) with factors “sequence type” (predictable/scrambled) and “expression type” (fearful/neutral). Our principal aim when analyzing MEG responses was to test for a main effect of sequence type. To this end, we measured the averaged MEG responses across trials to S1 faces from 100 ms prestimulus until 500 ms poststimulus (Fig. 1*c*). Besides testing our primary hypotheses concerning predictability of dynamic sequences, we also included static S1 presentations to compare with previous literature. Participants viewed static S1 faces for the same duration as predictable and scrambled S1 sequences (360 ms). We matched the expressions of the static S1 faces to the 3 possible endpoints of the predictable and scrambled sequences: (28%, 45%, or 63% fearfulness).

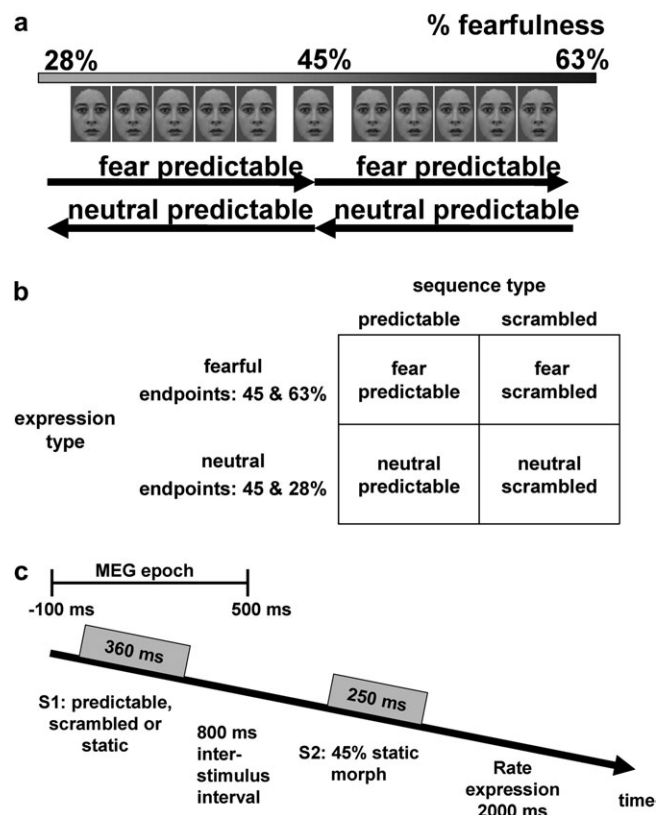


Figure 1. Stimuli and procedures. (*a*) A morph continuum for one face. S1 presentations comprised predictable and scrambled animated sequences constructed using the 6 images between 28%–45% and 45%–63% and static images (28%, 45%, and 63%). (*b*) Factorial design. The factor sequence type controls whether sequences depict a coherent transition between neutral and fearful expressions or a scrambled, unpredictable version of this transition. For the factor expression type, we describe as “fearful” sequences which transition predictably from neutral toward fear and the scrambled versions of these sequences. We describe as “neutral” sequences which transition predictably from fear toward neutral and scrambled versions of these sequences. (*c*) For each trial, S1 presentations were followed by an 800 ms-blank screen and then a static 250-ms target (S2) which participants rated for fearfulness.

Following presentation of S1 (Fig. 1*c*) and an 800-ms interstimulus interval, participants then viewed a brief (250 ms) static target face (S2). S2 faces always expressed 45% fearfulness, and participants rated the fearfulness of S2 faces on a 4-point visual analog scale. Two seconds then elapsed before the onset of the next trial. Our principal aim when analyzing the behavioral data was to test whether perception was biased by predictable sequences, relative to scrambled sequences. A variation of this design was also tested behaviorally in a pilot study in 8 participants using the same stimuli (plus 5 additional identities), which (similar to the findings herein) showed ratings of emotion in targets (compared with the endpoint-matched scrambled sequences) that were biased according to the preceding emotion trajectory direction.

Stimuli

We selected images of 7 facial identities (5 females) from the KDEF database (Lundqvist et al. 1998). For each identity, we selected images depicting fear and neutral expressions. Using landmark-based morphing software (M.J. Gourlay; Georgia Institute of Technology, Atlanta, GA), for each identity, we constructed a morph continuum consisting of 27 equally spaced images between the fearful (100%) and neutral (0%) expression images, retaining the 11 images between 28% and 63% (Fig. 1*a*). These images were converted to grayscale and placed within a gray oval mask (occluding hair, clothing, etc.). Regions of each image not occluded by the mask were equated for luminance mean and range.

Predictable and scrambled sequences of images were constructed using either the 6 morphs between 28% and 45% or those between 45% and 63% (Fig. 1a). For each identity, there were 4 predictable sequences consisting of 360 ms animations in which the 6 images were presented (60 ms each) in order either from neutral to fearful (fear predictable) or from fearful to neutral (neutral predictable). Importantly (Fig. 1a), fear-predictable sequences on the average finish on a more fearful image (endpoints: 45% and 63%) than neutral-predictable sequences (28% and 45%). It was therefore necessary to construct control stimuli that were matched for endpoint. For this purpose, we constructed scrambled sequences for which we randomized the order of all but the endpoint image and presented this scrambled order as a 360-ms animation. We also presented to participants static faces continuously for 360 ms, which express zero image change. These were matched for sequence endpoints (28%, 45%, or 63% fearfulness). Predictable sequences entail consistently small transitions from image to image compared with scrambled sequences, and the inclusion of a static condition reduced this confound by showing that static faces evoke smaller amplitude responses than the predictable sequences despite having zero image change. All results we examined showed this hypothesized reduced response to static faces. Although we tested statistically for responses that were enhanced for static faces or for scrambled faces (relative to the other conditions), we did not detect any such effects. These predictable, scrambled, and static stimuli are all denoted as S1. S2 targets were static 45% morph faces of each of the 7 identities.

Experimental Procedures

The experiment consisted of 6 scanning sessions. Each session contained the same 84 trials, although in a random order, thereby replicating all experimental conditions. Each of the 3 sequence conditions (predictable, scrambled, and static) composed a third of the trials (randomly intermixed), giving 168 trials per condition for each participant. Each trial (Fig. 1b) began with the presentation of a fixation cross for 500 ms, followed by an S1 stimulus that could be predictable, scrambled, or static (360 ms). After a blank screen for 800 ms, an S2 static target face appeared for 250 ms. Following the offset of S2, a blank screen was presented for 2000 ms. Participants rated the fearfulness of S2 on a 4-point scale ("1" was most neutral and "4" was most fearful) using a button box in their right hands. Participants were not told that the image was always 45% but that variation in expression would appear small and to nevertheless try to use the whole scale. They were given about a minute of experience with the stimuli to calibrate their responses after which they typically reported perceiving variation in expression of the targets.

MEG Data Acquisition and Analysis

We scanned participants while testing them with the aforementioned behavioral paradigm and acquired all behavioral data reported here during scanning. We acquired MEG recordings in a magnetically shielded room using a 275-channel CTF system with SQUID-based third-order axial gradiometers (VSM MedTech Ltd., Coquitlam, British Columbia). Neuromagnetic signals were digitized continuously at a sampling rate of 480 Hz. Data were analyzed using SPM5 (Wellcome Trust Centre for Neuroimaging, London; <http://www.fil.ion.ucl.ac.uk/spm/>) and MATLAB (The MathWorks, Natick, MA). The continuous time series for each participant was subjected to a Butterworth band-pass filter at 0.5–50 Hz. Baseline-corrected epochs were extracted from the data beginning 100 ms prior to S1 onset and ending 500 ms post-S1 onset (Fig. 1b). EPOCHED trials for which the signal strength exceeded 3000 femtoesla were discarded. Averaged sensor data were converted to 3-dimensional spatiotemporal volumes by "stacking" 2-dimensional linearly interpolated sensor images in peristimulus time. These 3-dimensional spatiotemporal volumes were submitted to mass univariate general linear models using conventional SPM procedures (Kilner et al. 2005). This enabled us to test for responses in all 3 dimensions (2-dimensional sensor space and peristimulus time). The resultant statistical parametric maps were multiple comparison corrected by applying Gaussian random field theory family-wise error (FWE) correction to small volumes encompassing either occipital or temporal sensors.

When relevant sensor-space effects were identified, we then identified the sources of these effects using source reconstruction as implemented in SPM5. For each participant, we constructed a forward model describing the transformation between dipolar sources distributed over the cortical surface, and the magnetic field distribution measured by the MEG sensors. Sources were modeled using the 7204 vertex template cortical mesh available in SPM5, defined in the standardized space of Talairach and Tournoux and coregistered to the sensor locations via 3 fiducial marker positions (Mattout, Henson, and Friston 2007). The gain matrix of the lead field model was computed using a spherical head model (<http://neuroimage.usc.edu/brainstorm/>), which has been shown to produce satisfactory reconstructions of ventral temporal sources in face perception paradigms (Henson et al. 2007). Source estimates were computed on the ensuing canonical mesh using restricted maximum likelihood estimation to invert the forward model (Mattout, Phillips, et al. 2007; Mattout et al. 2008). This inversion proceeded by modeling covariance components using multiple sparse priors (Friston, Harrison, et al. 2008). The hyperparameters on these multiple sparse priors were estimated using a greedy search (Friston, Chu, et al. 2008). This algorithm was deployed under group constraints (Litvak and Friston 2008), which provides an optimal mixture of empirical sparse priors on sources that is consistent over participants. By factorizing participant-specific and source-specific variation, the reconstructed activity across different participants can be attributed to the same set of empirically determined sources. This yielded source reconstructions for each experimental condition and for each participant. A temporal contrast was used to summarize responses at specific times of interest. This entailed multiplying the data with a (8-ms standard deviation [SD]) Gaussian window, centered on the peristimulus time of interest, and computing the sum of squared activity at each source. Contrasts were smoothed on the canonical mesh using a graph Laplacian (diffusion coefficient of 0.8) and projected to standard anatomical image space for between-participant analysis. To ensure isotropic smoothness, the contrast images were smoothed with a 3-dimensional Gaussian filter (8-mm full-width at half-maximum). The contrasts were analyzed using the same procedures used for the sensor data, namely conventional statistical parametric mapping (with whole-brain random field theory control over FWE at the cluster level).

Results

Behavioral Results

Behavioral analysis of the fear perception of S2 faces proceeded using 21 participants (one participant was excluded from analysis because behavioral results showed extreme outlying scores, >3 SDs). Figure 2 shows fear perception of S2 faces, normalized to the Z-score of the sample responses. As hypothesized, subjects' perception was biased by predictable sequences compared with the scrambled sequences. We tested the contrast (fear predictable–fear scrambled) – (neutral predictable–neutral scrambled), $t(20) = 1.76$, $P = 0.055$. Thus, predictable sequences (when compared against scrambled sequences) biased perception to the expression consistent with the cause of the preceding sensory trajectory.

We also found a main effect of expression type, in which fearful sequences (regardless of predictability) heightened S2 fear perception compared with the neutral sequences ([fear predictable + fear scrambled] – [neutral predictable + neutral scrambled]), $F_{1,20} = 19.43$, mean square error (MSE) = 0.56, $P = 0.001$. Note that we used identical images for the fearful and neutral sequences, the only difference was their sequence endpoints (Fig. 1a,b). Thus, the endpoints prime fear perception of the S2 morphs. We found matching results for the static S1 faces: fearfulness of static S1 faces (28%, 45%, or 63%) strongly enhanced fear perception of S2 faces, reflected by

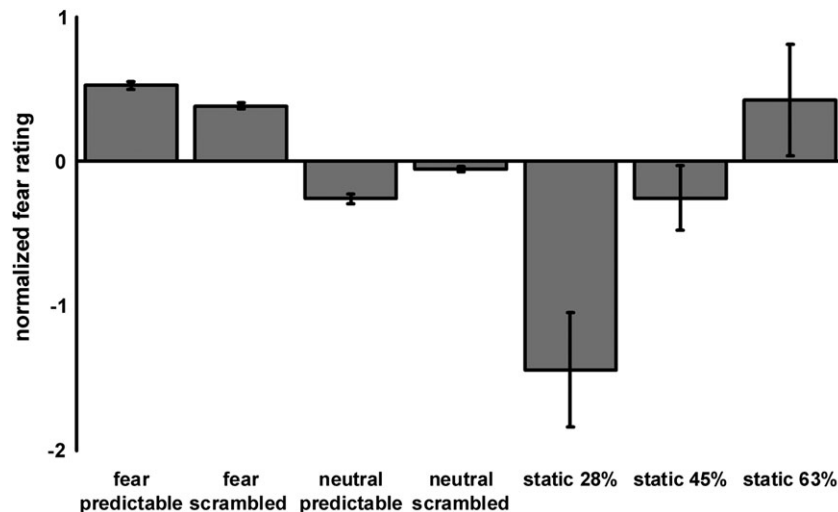


Figure 2. Behavioral results. Z-normalized means and standard errors of fear ratings to S2 faces following fear-and neutral-predictable sequences, scrambled sequences, and static S1 faces expressing 28%, 45%, and 63% fearfulness. Participants' fear perception is biased in the direction predicted by the preceding predictable sequence.

a significant main effect of static S1 fearfulness, $F_{2,40} = 13.93$, $MSE = 0.09$, $P < 0.0001$.

MEG Responses to S1 Sequences

At the between-participant (group) level, we specified general linear models to test (in sensor space and source space) our principal contrast between the 2 types of S1 sequence: predictable versus scrambled (i.e., the main effect of sequence type). In sensor space, we performed *t*-tests for every point (voxel) in the 3-dimensional space defined by the 2-dimensional MEG sensor-space projections and time. For S1 responses, significant voxels showed well-defined spatiotemporal clusters, each with a peak time and sensor-space location. For S2 responses, trends toward predictability-related effects (when scrambled stimuli were used as a control) were not robust and are not detailed here.

Predictable S1 sequences evoked the earliest responses (Fig. 3*a*) around 165 ms peaking at right occipital sensors (peak voxel $P < 0.05$, FWE corrected). Figure 3*b* shows the time course of a sensor that was located near the peak voxel and also clearly shows the M100 component (Liu et al. 2002). Note that there was no significant main effect of expression type or an interaction (Fig. 3*c*). Also note that responses to predictable sequences are heightened relative to responses to static S1 faces. As this effect (165 ms) arose prior to the onsets of the endpoint images (shown at 300–360 ms), it can only reflect responses to the first 1 or 2 image transitions. This means that the increased response to the predictable sequences, as compared with the scrambled and static S1 faces, is evidence for computations due to the coherent trajectories of the stimuli.

Having identified this early effect in sensor space, we then localized the anatomic sources causing this effect by performing source reconstructions within a time window of 160–170 ms for every participant in every condition. These reconstructions were analyzed using a general linear model identical to that used for sensor-space analysis. Figure 4*a* shows the results for the main effect of sequence type (predictable > scrambled), and Table 1 shows the anatomic locations of areas that survived a cluster-level FWE correction of $P < 0.05$. Sensitivity to

predictable S1 sequences was observed in right visual cortex, peaking in Brodmann area 18 and extending into area 17. Figure 4*b* shows the mean adjusted responses at the peak voxel, which approximates the pattern of effects observed in sensor space.

We observed another effect in sensor space later in peristimulus time showing similar sensitivity to the predictable sequences (Fig. 5*a*). This manifested as a dipolar field pattern, which was sustained for about 100 ms, showing a right lateral negativity (peaking at 237 ms) and a left medial positivity (peaking at 230 ms). Peak voxels were $P < 0.05$ FWE corrected. As before, the predictable sequences heightened responses compared with the static S1 faces, as well as to the scrambled sequences. As this dipolar topography bears some similarity to that of the well-studied M170 component (Liu et al. 2002), we illustrate in Figure 5*b* the relationship of this effect to the M170 by selecting lateral temporal sensors from the right and left hemisphere which express both this predictability-sensitive response and the M170 (see also Supplementary Fig. 1).

We performed source reconstructions within a window of 232–242 ms for every participant in every condition and then submitted these reconstructions to the identical general linear model used for sensor-space analysis. Figure 6*a* shows the statistical parametric map for the contrast predictable > scrambled, and Table 1 reports peaks for clusters which survived a FWE cluster-level correction of $P < 0.05$. We found a large right occipital response, subsuming Brodmann areas 17, 18, and 19 and extending ventrally into right posterior fusiform gyrus. There was also a smaller cluster in left occipital cortex, area 18. We observed another cluster in right posterior STS, near the temporal-parietal junction. Figure 6*b* shows the patterns of adjusted response means at the peak voxel in the right fusiform gyrus and posterior STS, which approximate the pattern of effects observed in sensor space. We also observed sensitivity to predictable dynamics in bilateral premotor areas. These were located in dorsal midprecentral gyrus, primarily in Brodmann area 6 in both hemispheres, but extending ventrally into Brodmann area 44 in the right hemisphere.

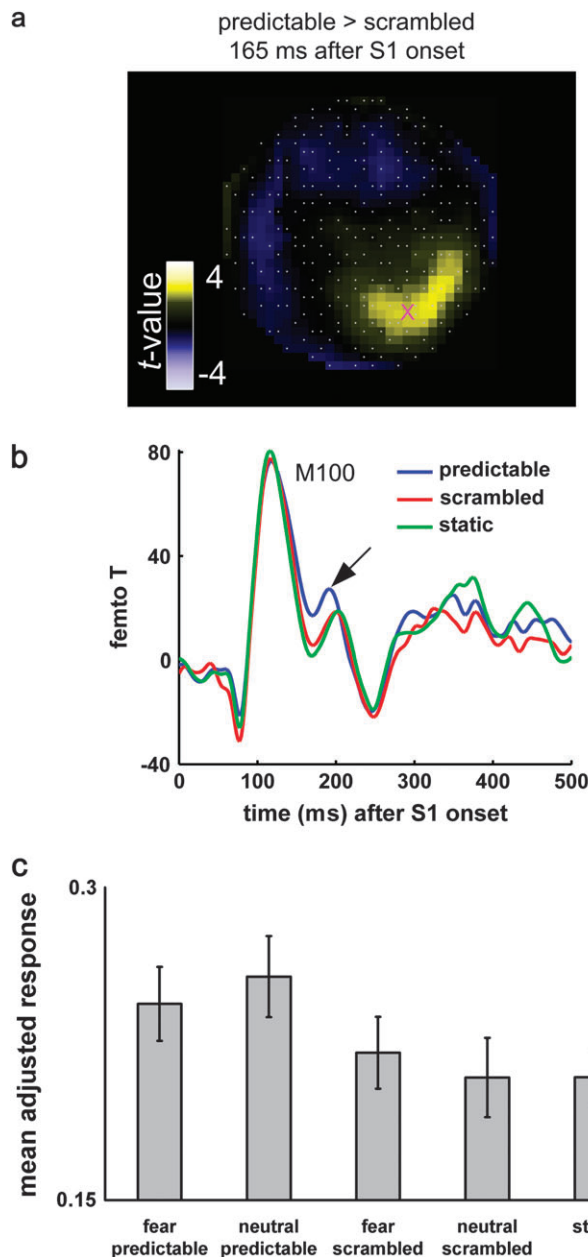


Figure 3. Early occipital effects in sensor space. (a) Statistical parametric map of the t -statistic in sensor space at 165 ms for the contrast predictable > scrambled, showing a cluster peaking at occipital sensors. (b) Time course of response at a sensor (denoted by magenta cross in [a]) near the peak occipital effect. The M100 deflection is labeled, and the arrow indicates the effect of predictable dynamics. (c) Mean adjusted responses at the occipital peak showing activation height over conditions, at 165 ms including 90% confidence intervals (based on between-participant variability). Predictable S1 sequences produce greater activation than scrambled and static.

Discussion

We measured evoked MEG responses to dynamic sequences of facial expressions, which varied in the predictability of movement trajectories. Predictable sequences comprised naturalistic and coherent transitions between fearful and neutral facial expressions. In contrast, scrambled sequences were formed of the same images but subtended unnatural facial motion. Our findings demonstrate neuronal representations of coherent, predictable motion trajectories, which first arose

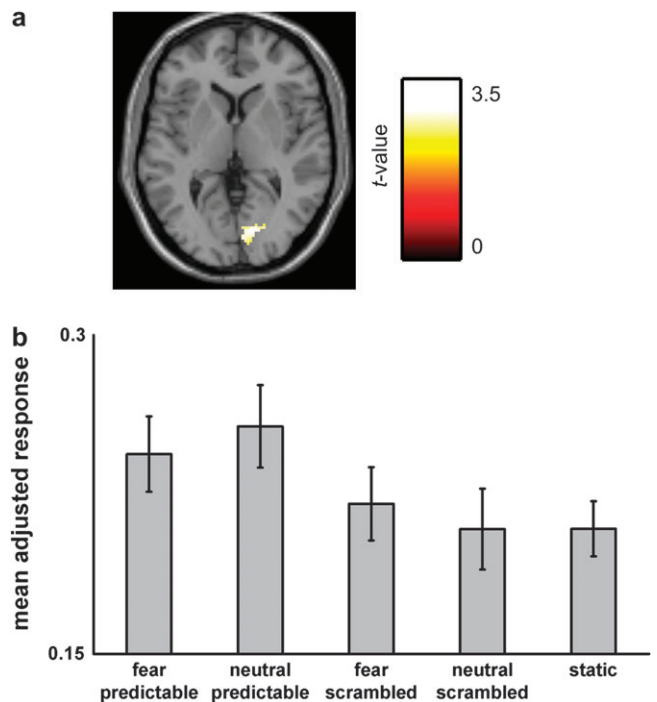


Figure 4. Occipital effects around 165 ms in source space. (a) Statistical parametric map of the t -statistic in source space (Montreal Neurological Institute coordinate: $z = 4$) for the contrast predictable > scrambled, thresholded at $P < 0.005$ uncorrected and showing sensitivity to predictable S1 sequences in right visual cortex, and Brodmann areas 17 and 18. (b) Mean adjusted responses at peak voxel in right occipital cortex including 90% confidence intervals (based on between-participant variability).

Table 1

Anatomical sources sensitive to predictable dynamics

Area	Talairach (x, y, z)	P value, uncorrected	P value FWE corrected
160–170 ms, predictable > scrambled			
Right medial occipital	12, -78, 4	$P < 0.001$	$P < 0.001$
232–242 ms, predictable > scrambled			
Right medial occipital	6, -82, 2	$P < 0.001$	$P < 0.001$
Right posterior fusiform	26, -84, 16	$P = 0.003$	
Left medial occipital	-14, -88, 40	$P = 0.002$	$P < 0.001$
Right STS	50, -66, 36	$P = 0.001$	$P < 0.001$
Left precentral gyrus	-42, -14, 32	$P = 0.001$	$P < 0.001$
Right precentral gyrus	58, 8, 16	$P = 0.003$	$P < 0.001$

(165 ms) in low-level occipital areas and later (237 ms) engaged the posterior fusiform gyrus and STS, areas known to contribute to face perception and biological motion perception. Also at this later time, the presence of predictable movement heightened activity in premotor areas. Effects were robust and reproducible in sensor and source space. Importantly, we observed effects of the predictable trajectory on behavior. Participants' fear perception of subsequent (S2) static target faces was biased toward the expression causing the preceding predictable sequence (S1), underscoring a role for the representation of dynamics in the perception of facial expressions.

Representations of Facial Expression Trajectory

We found, as hypothesized, that the right posterior STS shows heightened responses to predictable sensory trajectories, compared with scrambled presentations with no coherent

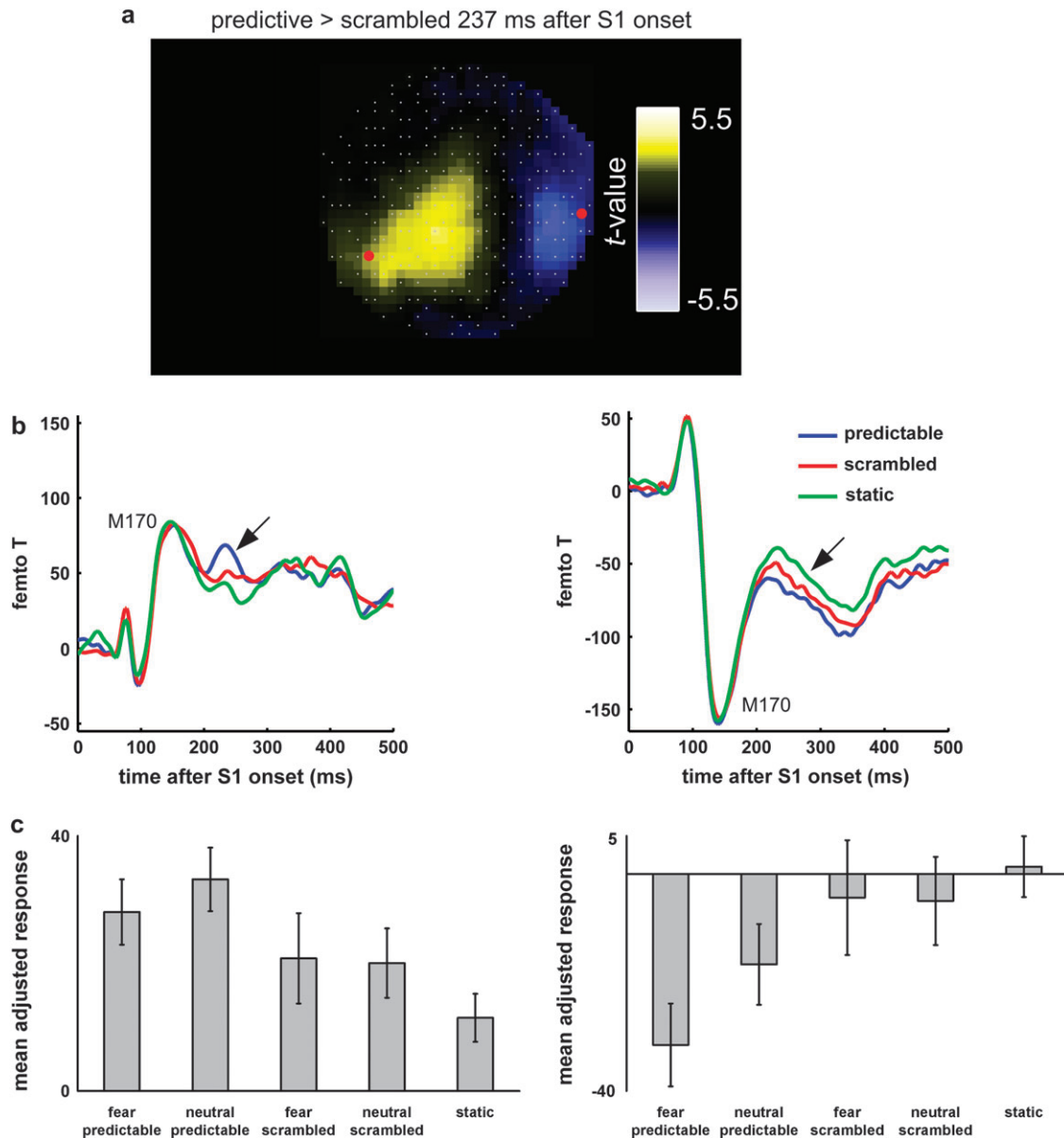


Figure 5. Sensor space effects at 237 ms. (a) Statistical parametric map of the t -statistic in sensor space at 237 ms for the contrast predictable > scrambled, showing peaks over left medial and right lateral temporal sensors. (b) Time courses of response at sensors in left and right hemispheres shown in the red circles in (a). The M170 deflections are labeled, and the arrows indicate sensitivity to predictable dynamics. (c) Mean adjusted responses at lateral temporal voxels in left and right hemisphere showing pattern of effects in sensor space at 237 ms including 90% confidence intervals. Predictable S1 sequences produce greater activation than scrambled and static.

trajectory. This is consistent with studies showing the STS is responsive to stimuli depicting facial and bodily actions, including head pose (Andrews and Ewbank 2004), yawning (Schürmann et al. 2005), eye gaze, and facial expressions of emotion (Haxby et al. 2000; Calder and Young 2005; Furl et al. 2007) and pain (Simon et al. 2006). Posterior STS expresses activity common to slowly evolving mouth and hand movements, whereas responses in mid-STS are selective for mouth movements (Thompson et al. 2007) and responds more to biological point-light displays of body actions that are coherent than when they are scrambled (Grossman and Blake 2002). Our findings go beyond a demonstration that STS represents facial expressions or bodily actions and suggests further that this representation entails recognition of coherent facial trajectories.

Interestingly, we observed a hierarchical timing of predictability-related activity. Predictability effects emerged first in early visual cortex after only an image transition or 2 and then later spread to higher visual areas. This hierarchical response pattern is particularly notable as it is predicted by extant theories. For example, Giese and Poggio (2003) propose that earlier visual areas code information over relatively short time scales, even as near-instantaneous “snapshots.” Higher areas (STS and perhaps premotor cortex) then integrate these lower level representations over longer time scales and respond only if snapshots transition “as predicted.” This approach therefore hypothesizes a hierarchical organization in which higher areas (which are sensitive to predictable information) require longer time scales for response than lower level areas. Bayesian models constitute another

a predictable > scrambled
232-242 ms after S1 onset

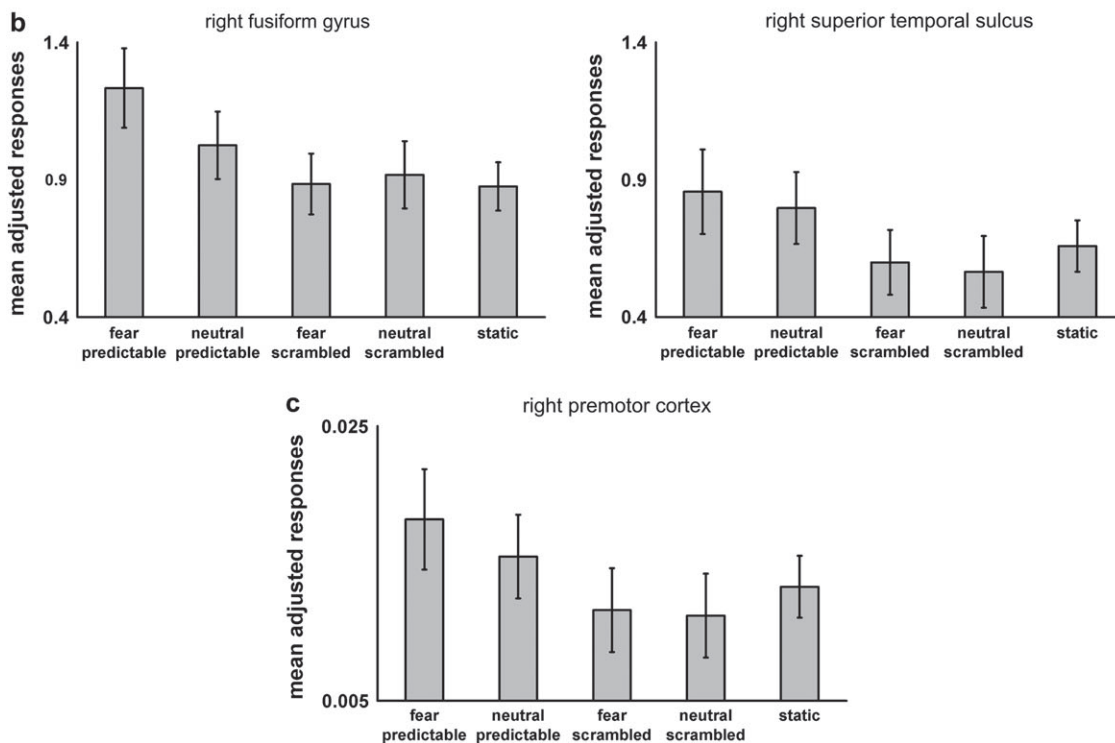
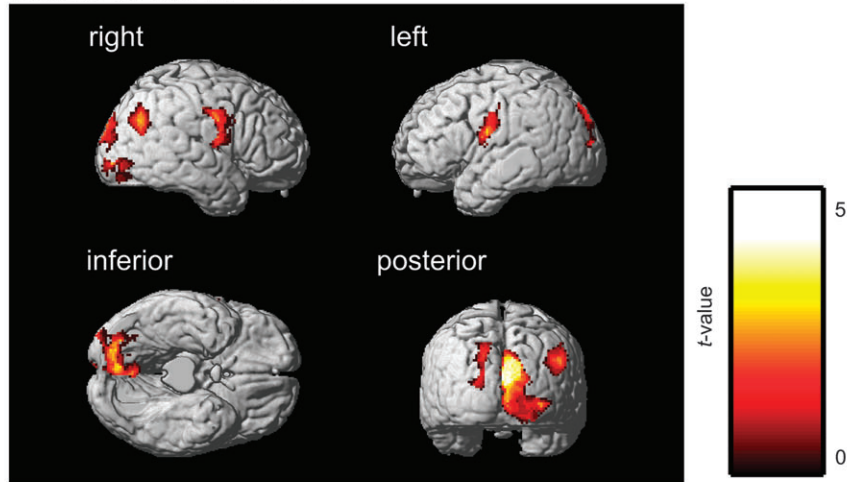


Figure 6. Source space effects around 237 ms. (a) Statistical parametric map of the t -statistic in source space for the contrast predictable > scrambled, thresholded at $P < 0.005$ uncorrected and showing effects in bilateral occipital cortex, right STS, right fusiform gyrus, and bilateral premotor areas. (b) Mean adjusted responses of all conditions at peak voxels in right fusiform, STS, and right premotor cortex including 90% confidence intervals.

important class of predictive hierarchies (Friston 2005). Recent applications of Bayesian models to dynamic stimulus representation have led to a convergent prediction: higher levels may be sensitive to progressively longer temporal scales (Kiebel et al. 2008).

Empirically, a recent functional magnetic resonance imaging study (Hasson et al. 2008) provides evidence for a similar hierarchical response patterns to dynamic visual information as we observed. They examined responses to dynamic sequences from movies and reported that the sensitivity to temporal structure in STS responses spans longer time periods than that of lower level visual areas. These authors similarly propose a visual

hierarchy of temporal receptive fields, which accumulates information over progressively longer temporal windows, with STS accumulating over longer time periods than lower level areas. Our results are therefore predicted by these models and are consistent with the results of Hasson et al. (2008). The early occipital response to predictable dynamics (according to this view) reflects accumulations over a shorter time interval and thereby responds sooner than the fusiform/STS, which accumulates information over longer intervals. From this perspective, activity around 237 ms might reflect a first response to the ongoing integration of information at a temporal scale relevant for the recognition of facial expressions.

In addition to STS, evoked fields sensitive to predictability of S1 dynamics also showed sources localized to posterior fusiform gyrus. This area may relate to nearby face-selective areas such as fusiform and occipital face areas, which show robust face-selective activations (Kanwisher and Yovel 2007). Prior accounts claim that ventral temporal areas comprise a pathway (distinct from the dorsal pathway including the STS) that represents invariant facial information, supporting identity perception (Haxby et al. 2000). We therefore did not hypothesize fusiform areas would be sensitive to predictable dynamics, although similar findings have been reported (Grossman and Blake 2002). We note, however, that this area is rather posterior, compared with the classic “fusiform face area,” and so may correspond more closely to the more posterior, face-selective “occipital face area.”

Interestingly, we observed premotor activity concomitant with activity in temporal lobe visual areas. This is consistent with several studies of biological motion that report motor system activity including Brodmann areas 6 and 44 (Buccino et al. 2001; Jeannerod 2001; Sato et al. 2004; Saygin et al. 2004; Calvo-Merino et al. 2006; Dayan et al. 2007; Montgomery and Haxby 2008). Similarly, we also found greater activity in Brodmann area 6 (extending into 44 in the right hemisphere) when stimulus dynamics conveyed a predictable action sequence relative to stimuli that were unnatural. We therefore also show premotor cortex responses to biological motion and further extend these findings by showing such responses are sensitive to predictable facial expression trajectories.

We note that “simulation” (Keysers and Gazzola 2006; Gallese 2007; Hurley 2008) and “common coding” (Hommel et al. 2002) theories posit that representations of one’s own motor acts are recruited when representing the actions of others. Kilner et al. (2007), for example, hypothesize that simulated motor acts provide predictions to the visual system. Although we cannot conclusively demonstrate these motor simulations on the basis of our data, our results suggest that paradigms using dynamic facial expressions may provide a context for exploring whether motor simulation plays any role in visual action prediction.

Relationships to M100 and M170 Components

There has been much interest in 2 robust MEG responses to faces: the M100 and the M170 (Liu et al. 2002). In particular, the M170 has been shown to be face selective and may relate to midfusiform gyrus activation (Furl et al. 2007). We observed an occipital effect at 165 ms (at the same time as the M170). At this time, however, there was no predictability effect on the M170 peaks (Fig. 5), which are situated distant to the significant 165 ms occipital effect (Supplementary Fig. 1). Therefore, we cannot easily conclude that the bipolar deflections typically associated with the M170 (Liu et al. 2002; Furl et al. 2007) show sensitivity to predictable sequences. However, we observed a late-onset sustained response, with similar (but reduced) field topography to the M170 (Supplementary Fig. 1, 237 ms). Similar post-M170 sustained responses have been previously observed in response to facial expressions and associated with STS activity (Furl et al. 2007). Perhaps predictability effects are associated with this sustained response, rather than the M170 itself.

Predictability Effects on Behavior

We used behavioral measures to demonstrate that perception is biased by anticipatory representations. We examined the influences of exposure to face expression trajectories on fear perception of subsequent (S2) static faces. Although S2 faces always depicted the same mixture (45%) of fear and neutral expressions, participants’ fear perception shifted in the direction of the expression predicted by the preceding trajectory.

We designed the behavioral paradigm to reduce or eliminate potential confounding explanations for these effects, such as capture by apparent motion and repulsive aftereffects, where static S1 facial expressions bias expression perception of S2 faces away from the S1 expression (Webster et al. 2004). We consequently used a long 800-ms interstimulus interval and relatively short duration S1 faces (360 ms) and eliminated the oft-used preexposure period (Webster et al. 2004): Experimental parameters intended to attenuate aftereffects. As aftereffects seem to depend also on the size of the expression difference between the S1 and S2 stimuli, we chose S1 stimuli that were close to S2 face expression (Fig. 1*a*).

As known perceptual biases such as aftereffects do not explain our results, it is more likely that expression perception relies on a representation that hierarchically encodes the predicted motion trajectory, which sensitizes the visual system to detect expressions that are consistent with this trajectory. The closely related representation momentum effect (where memory for a sequence endpoint is biased in the direction of the sequence; Freyd and Finke 1984; Hubbard and Bharucha 1988; Thornton and Hubbard 2002; Hubbard 2008) has also been shown for facial expressions (Yoshikawa and Sato 2008). These representational momentum effects reflect memory for the last stimulus, after it has already been perceived. Our finding, however, gives direct evidence that trajectories distort the instantaneous perception of a face. Similar effects have been reported for low-level visual trajectories (Ramachandran and Anstis 1983) and such effects can be modeled using continuous attractors in neuronal networks (Treves 2004). In addition to this predictive bias, we also observed that static S1 faces produced large expression priming effects on perception of S2 faces (Fig. 2). This finding is probably not surprising because static sequences are the most predictable sequence.

Conclusion

We show that low levels of the visual system detect predictable structure in dynamic face expressions as quickly as 165 ms and that higher level regions associated with face perception respond to sensory trajectories within the next 100 ms. Predictions based on these representations may sensitize the visual system to detect subsequent stimuli that are consistent with the cause of the preceding sensory trajectory. These findings raise important questions concerning neural function at different levels of the visual system, particularly with respect to mechanisms in the STS. These neural mechanisms speak directly to how we employ our visual experience to make sense of the continual changes involved in even the simplest everyday social interactions.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Funding

Wellcome Trust Program (to R.J.D.); Human Frontier Science Program (RGP0047/2004-C to A.T. and R.J.D.).

Notes

Thanks to Bruno Averbeck, Emrah Duzel, James Kilner, and Jeremie Mattout for their useful comments and advice. *Conflict of Interest:* None declared.

Address correspondence to Nicholas Furl, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London, WC1N 3BG, UK. Email: n.furl@fil.ion.ucl.ac.uk.

References

- Akrami A, Liu Y, Treves A, Jagadeesh B. 2008. Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli. *Cereb Cortex*. 19:760-776.
- Andrews TJ, Ewbank MP. 2004. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage*. 23:905-913.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, et al. 2006. Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA*. 103:449-454.
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci*. 13:400-404.
- Calder AJ, Young AW. 2005. Understanding the recognition of facial identity and expression. *Nat Rev Neurosci*. 6:641-651.
- Calvo-Merino B, Grèzes J, Glaser D, Passingham RE, Haggard P. 2006. Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr Biol*. 16:1905-1910.
- Dayan E, Casile A, Levit-Binnun N, Giese MA, Hendler T, Flash T. 2007. Neural representations of kinematic laws of motion: evidence for action-perception coupling. *Proc Natl Acad Sci USA*. 104:20582-20587.
- Freyd JJ, Finke RA. 1984. Representational momentum. *J Exp Psychol Learn Mem Cogn*. 10:126-132.
- Friston K. 2005. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*. 360:815-836.
- Friston K, Harrison L, Daunizeau J, Keibel S, Phillips C, Trujillo-Barreto N, Henson R, Flandin G, Mattout J. 2008. Multiple sparse priors for the M/EEG inverse problem. *Neuroimage*. 39:1104-1120.
- Friston KJ, Chu C, Mourão-Miranda J, Hulme O, Rees G, Penny W, Ashburner J. 2008. Bayesian decoding of brain images. *Neuroimage*. doi: 10.1016/j.neuroimage.2007.08.013.
- Friston KJ, Trujillo-Barreto N, Daunizeau J. 2008. DEM: a variational treatment of dynamic systems. *Neuroimage*. 41:849-885.
- Furl N, van Rijsbergen NJ, Treves A, Friston KJ, Dolan RJ. 2007. Experience-dependent coding of facial expression in superior temporal sulcus. *Proc Natl Acad Sci USA*. 104:13485-13489.
- Gallese V. 2007. Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philos Trans R Soc Lond B Biol Sci*. 362:659-669.
- Giese MA, Poggio T. 2003. Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci*. 4:179-192.
- Graf M, Reitzner B, Corves C, Casile A, Giese M, Prinz W. 2007. Predicting point-light actions in real-time. *Neuroimage*. 36(Suppl 2):T22-T32.
- Grossman ED, Blake R. 2002. Brain areas active during visual perception of biological motion. *Neuron*. 35:1167-1175.
- Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. 2008. A hierarchy of temporal receptive windows in human cortex. *J Neurosci*. 28:2539-2550.
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn Sci*. 4:223-233.
- Henson RN, Mattout J, Singh KD, Barnes GR, Hillebrand A, Friston K. 2007. Population-level inferences for distributed MEG source localization under multiple constraints: application to face-evoked fields. *Neuroimage*. 38:422-438.
- Hommel B, Müsseler J, Aschersleben G, Prinz W. 2002. The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav Brain Sci*. 24:849-878.

- Hosoya T, Baccus SA, Meister M. 2005. Dynamic predictive coding by the retina. *Nature*. 436:71-77.
- Hubbard TL. 2008. Representational momentum and related displacements in spatial memory: a review of the findings. *Psychon Bull Rev*. 12:822-851.
- Hubbard TL, Bharucha JJ. 1988. Judged displacement in apparent vertical and horizontal motion. *Percept Psychophys*. 44:211-221.
- Hurley S. 2008. The shared circuits model (SCM): how control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behav Brain Sci*. 31:1-22.
- Jeannerod M. 2001. Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*. 14:S109.
- Jehee JF, Rothkopf C, Beck JM, Ballard DH. 2006. Learning receptive fields using predictive feedback. *J Physiol Paris*. 100:125-132.
- Kanwisher N, Yovel G. 2007. The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci*. 361:2109-2128.
- Keysers C, Gazzola V. 2006. Towards a unifying neural theory of social cognition. *Prog Brain Res*. 156:379-401.
- Kiebel SJ, Daunizeau J, Friston KJ. 2008. A hierarchy of time-scales and the brain. *PLoS Comput Biol*. 4:e1000209.
- Kilner J, Friston KJ, Frith CD. 2007. Predictive coding: an account of the mirror neuron system. *Cogn Process*. 8:159-166.
- Kilner JM, Kiebel SJ, Friston KJ. 2005. Applications of random field theory to electrophysiology. *Neurosci Lett*. 374:174-178.
- Larsen A, Madsen KH, Lund TE, Bendeson C. 2006. Images of illusory motion in primary visual cortex. *J Cogn Neurosci*. 18:1174-1180.
- Litvak V, Friston K. 2008. Electromagnetic source reconstruction for group studies. *Neuroimage*. 42:1490-1498.
- Liu J, Harris A, Kanwisher N. 2002. Stages of processing in face perception: an MEG study. *Nat Neurosci*. 5:910-916.
- Lundqvist D, Flykt A, Öhman A. 1998. The Karolinska Directed Emotional Faces—KDEF [CD-ROM]. Stockholm, Sweden: Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.
- Mattout J, Henson RN, Friston KJ. 2007. Canonical source reconstruction for MEG. *Comput Intell Neurosci*. doi: 10.1155/2007/67613.
- Mattout J, Phillips C, Daunizeau J, Friston KJ. 2007. Bayesian inversion of EEG models. In: Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD, editors. *Statistical parametric mapping: the analysis of functional brain analysis*. London: Academic Press. p. 367-376.
- Mattout J, Phillips C, Penny W, Rugg MD, Friston KJ. 2008. MEG source localization under multiple constraints: an extended Bayesian framework. *Neuroimage*. 30:753-767.
- Montgomery KJ, Haxby JV. 2008. Mirror neuron system differentially activated by facial expressions and social hand gestures: a functional magnetic resonance imaging study. *J Cogn Neurosci*. 20:1866-1877.
- Muckli L, Kohler A, Kriegeskorte N, Singer W. 2005. Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biol*. 3:e265-e275.
- Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL. 2002. Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci USA*. 99:15164-15169.
- Ramachandran VS, Anstis SM. 1983. Exploration of the motion path in human visual perception. *Vision Res*. 23:83-85.
- Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*. 2:79-87.
- Sato W, Kochiyama T, Yoshikawa S, Naito E, Matsumura M. 2004. Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Brain Res Cogn Brain Res*. 20:81-91.
- Saygin A, Wilson SM, Hagler DJ, Jr, Bates E, Sereno MI. 2004. Point-light biological motion perception activates human premotor cortex. *J Neurosci*. 24:6181-6188.
- Schürmann M, Hesse MD, Stephan K, Saarela M, Zilles K, Hari R, Fink GR. 2005. Yearning to yawn: the neural basis of contagious yawning. *Neuroimage*. 24:1260-1264.
- Schweidrzik CM, Alink A, Kohler A, Singer W, Muckli L. 2007. A spatio-temporal interaction on the apparent motion trace. *Vision Res*. 47:3424-3433.
- Simon D, Craig KD, Miltner WHR, Rainville P. 2006. Brain responses to dynamic facial expressions of pain. *Pain*. 126:309-318.

- Sterzer P, Haynes JD, Rees G. 2006. Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. *Neuroimage*. 32:1308-1316.
- Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J. 2006. Predictive codes for forthcoming perception in frontal cortex. *Science*. 314:1311-1314.
- Summerfield C, Koechlin E. 2008. A neural representation of prior information during perceptual inference. *Neuron*. 59:336-347.
- Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci*. 11:1004-1006.
- Thompson JC, Hardee JE, Panayiotou A, Crewther D, Puce A. 2007. Common and distinct brain activation to viewing dynamic sequences of face and hand movements. *Neuroimage*. 37:966-973.
- Thornton IM, Hubbard TL. 2002. *Representational momentum: new findings, new directions*. Hove (UK): Psychology Press.
- Treves A. 2004. Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation. *Hippocampus*. 14:539-556.
- van der Gaag C, Minderer RB, Keysers C. 2007. Facial expressions: what the mirror neuron system can and cannot tell us. *Soc Neurosci*. 2:179-222.
- Verfaillie K, Daems A. 2002. Representing and anticipating human actions in vision. *Vis Cogn*. 9:217-232.
- Webster MA, Kaping D, Mizokami Y, Duhamel P. 2004. Adaptation to natural facial categories. *Nature*. 428:557-561.
- Yoshikawa S, Sato W. 2008. Dynamic facial expressions of emotion induce representational momentum. *Cogn Affect Behav Neurosci*. 8:25-31.