OXFORD UNIVERSITY PRESS Cerebral Cortex

Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus

Journal:	*Cerebral Cortex*
Manuscript ID:	CerCor-2013-01178.R1
Manuscript Type:	Original Articles
Date Submitted by the Author:	n/a
Complete List of Authors:	Furl, Nicholas; MRC Cognition and Brain Sciences Unit, Emotion Group Henson, Rik; MRC, CBU Friston, Karl J; Wellcome Trust Centre for Neuroimaging, Institute of Neurology Calder, Andy; Medical Research Council, Cognition and Brain Sciences
Keywords:	functional magnetic resonance imaging, face perception, biological motion, superior temporal sulcus, dynamic causal modeling

SCHOLARONE™ Manuscripts

For Review

Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus

Running title: Sensitivity to dynamic faces

Nicholas Furl[1], Richard N. Henson[1], Karl J. Friston[2], Andrew J. Calder[1]

[1]MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF United Kingdom

[2]Wellcome Centre for Imaging Neuroscience, University College London, 12 Queen Square, London WC1N 3BG

Corresponding author:

Nicholas Furl, Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF, United Kingdom, nick.furl@mrc-cbu.cam.ac.uk

For Peer Review

ABSTRACT

The superior temporal sulcus (STS) in the human and monkey is sensitive to the motion of complex forms such as facial and bodily actions. We used functional magnetic resonance imaging (fMRI) to explore network-level explanations for how the form and motion information in dynamic facial expressions might be combined in the human STS. Ventral occipitotemporal areas selective for facial form were localized in occipital and fusiform face areas (OFA and FFA), and motion sensitivity was localized in the more dorsal temporal area V5. We then tested various connectivity models that modeled communication between the ventral form and dorsal motion pathways. We show that facial form information modulated transmission of motion information from V5 to the STS, and that this face-selective modulation likely originated in OFA. This finding shows that form-selective motion sensitivity in the STS can be explained in terms of modulation of gain control on information flow in the motion pathway, and provides a substantial constraint for theories of the perception of faces and biological motion.

For Peer Review

KEYWORDS

functional magnetic resonance imaging; face perception; biological motion; superior temporal sulcus; dynamic causal modeling

For Peer Review

Humans and other animals effortlessly recognize facial identities and actions such as emotional expressions even when faces continuously move. Brain representations of dynamic faces may be manifested as greater responses in the superior temporal sulcus (STS) to facial motion than motion of non-face objects (Pitcher et al. 2012), suggesting localized representations that combine information about motion and facial form. This finding relates to a considerable literature on "biological motion", which studies how the complex forms of bodily actions are perceived from only the motion of light points fixed to limb joints, with form-related texture cues removed (Johansson 1973). Perception of such stimuli has been repeatedly associated with the human posterior STS (Vaina et al. 2001; Giese and Poggio 2003; Vaina et al. 2004; Hein and Knight 2008; Jastorff and Orban 2009) with similar results observed in potentially corresponding areas of the macaque STS (Oram and Perrett 1994; Jastorff et al. 2012). The STS has been described as integrating form and motion information (Vaina et al. 2001; Giese and Poggio 2003), containing neurons that code for conjunctions of certain forms and movements (Oram and Perrett 1996). Nevertheless, the mechanisms by which STS neurons come to be sensitive to the motion of some forms but not others remains a matter of speculation (Giese and Poggio 2003).

We propose that network interactions can provide a mechanistic explanation for STS sensitivity to motion that is selective to certain forms, in this case, faces. Specifically, STS responses to dynamic faces could result from communicative interactions between pathways sensitive to motion and facial form. Such interactions can occur when one pathway modulates or "gates" the ability of the other pathway to transmit information to the STS. Using functional magnetic resonance imaging (fMRI), we localized face-selective motion sensitivity in the STS of the human and then used causal connectivity analyses to model how these STS responses are influenced by areas sensitive to motion and areas selective to facial form. We localized ventral occipital and fusiform face areas (OFA, FFA) (Kanwisher et al. 1997),

For Peer Review

which selectively respond to facial form versus other objects (Calder and Young 2005; Calder 2011). We also localized motion sensitivity to faces and non-faces in the more dorsal temporal hMT+/V5 complex (hereafter, V5). Together, these areas provide ventral and dorsal pathways to the STS. The ventral pathway transmits facial form information, via OFA and FFA, and the dorsal pathway transmits motion information, via V5. We then compared combinations of bilinear and non-linear dynamic causal models (Friston et al. 2003) to identify connectivity models that optimally explain how interactions between these form and motion pathways could generate STS responses to dynamic faces. We found that information about facial form, most likely originating in the OFA, gates the transmission of information about motion from V5 to the STS. Thus, integrated facial form and motion information in the STS can arise due to network interactions, where form and motion pathways play distinct roles.

MATERIALS AND METHODS

Participants

fMRI data were collected from 18 healthy, right-handed participants (over 18 years, 13 female) with normal or corrected-to-normal vision. Experimental procedures were approved by the Cambridge Psychology Research Ethics Committee.

Imaging Acquisition

A 3T Siemens Tim Trio MRI scanner with a 32 channel head coil was used for data acquisition. We collected a structural T1-weighted MPRAGE image (1 mm isotropic voxels). Functional data consisted of whole-brain T2*-weighted echo-planar imaging volumes with 32 oblique axial slices that were 3.5 mm thick, in-plane 64 × 64 matrix with resolution of 3 × 3

For Peer Review

mm, TR 2s, TE 30 ms, flip angle 78°. We discarded the first five "dummy" volumes to ensure magnetic equilibration.

Experimental Design

The experiment used a block design with two runs (229 scans per run), which were collected as the localizer for another experiment (Furl et al. 2013b). Note that the DCM analyses reported in Furl et al. (2013b) used independent data (from separate runs using different stimuli) to address a different phenomenon than considered here. All blocks were 11 s, comprised eight 1375 ms presentations of greyscale stimuli and were followed by a 1 s inter-block fixation interval. Participants fixated on a grey dot in the center of the display, overlaying the image. On a random one third of stimulus presentations this dot turned red and they pressed a key. Participants viewed six types of blocks, each presented six times. Dynamic face blocks contained dynamic facial expressions taken from the Amsterdam Dynamic Facial Expression Set (ADFES) (Van der Schalk et al. 2011). Four male and four female identities changed among neutral and either disgust, fearful, happy, or sad expressions. Identities and expressions appeared in a pseudo-random order, with each of the four expressions appearing twice in each dynamic face block. Dynamic object blocks included eight dynamic objects (Fox et al. 2009). For comparison, we also included dynamic and static patterns. We used a conventional low-level motion localizer, commonly used to localize and study motion sensitive areas hMT+/V5 and KO (Van Oostende et al. 1997). This ensured that our results are directly comparable to previous studies of low-level motion-sensitivity and verifies that the V5 voxels we identify using faces are a subset of hMT+/V5 voxels, as conventionally defined. These dynamic pattern blocks consisted of random-dot pattern videos with motion-defined oriented gratings. The stimuli depicted 50% randomly-luminous pixels which could move at one frame per second horizontally, vertically or

For Peer Review

diagonally left or right. Oriented gratings were defined by moving the dots within four strips of pixels in the opposite direction to the rest of the display, but at the same rate (Van Oostende et al. 1997). The remaining three block types — static face, object and pattern blocks — consisted of the final frames of the corresponding dynamic blocks.

Pre-processing and Analysis

We performed pre-processing and analysis using SPM8, DCM10 (Wellcome Trust Centre for Neuroimaging, London http://www.fil.ion.ucl.ac.uk/spm/) and MATLAB (The MathWorks, Natick, MA). Data were motion and slice-time corrected, spatially normalized to an EPI template in MNI space, smoothed to 8 mm full-width half-maximum and analyzed using the general linear model. At the first (within-participant) level, general linear models used proportionately-scaled data, an AR(1) autocorrelation model, a high-pass filter of 128 s and regressors constructed by convolving the onset times and durations for the different experimental blocks with a canonical hemodynamic response function.

At the first level, we localized face-selective ROIs in the right OFA and FFA by contrasting the average response to dynamic and static faces versus the average response to dynamic and static objects and random-dot patterns. We also identified an ROI showing motion sensitivity to faces in the vicinity of area hMT+/V5 (V5) by contrasting dynamic versus static faces. We further localized an area in the STS by computing the interaction effect in which motion sensitivity was larger for faces than for non-faces using the contrast (dynamic faces > static faces) > (dynamic objects/patterns > static objects/patterns). Lastly, we contrasted faces, objects and patterns versus fixation to localize the peak visual response, which was located in right Brodmann area 18 (BA18). For BA 18, we located the peak response to faces, objects and patterns in the whole sample of 18 participants (MNI: 16 -90 -4) and then identified subject-specific peaks within 8 mm of the group peak. Eleven of the 18

For Peer Review

participants evidenced significant responses (at $p < 0.01$ uncorrected) in all five ROIs in the right hemisphere and further analyses focused on these ROIs – given the right hemispheric dominance in face perception (Kanwisher et al. 1997). Note that our selection of ROIs for subsequent DCM analyses is slightly more conservative than standard approaches. This is because we chose subject-specific maxima that were within a specified distance of peaks in an orthogonal contrast (at the group level) c.f., (Friston et al 1997). In other words, they were selected using orthogonal (independent) criteria, rendering a correction for multiple comparisons redundant.

For connectivity analysis, we employed dynamic causal modeling (DCM; Friston et al. 2003) to test hypotheses about connectivity mechanisms that potentially could give rise to the selective facial motion sensitivity that we observed in the STS. DCM models ROI time series data by estimating coupling: the extent to which neural activity (hidden variables) in each brain area influences dynamics in connected brain areas. DCM parameters include exogenous inputs, endogenous connections, bilinear and non-linear modulatory connections. Exogenous inputs are estimates of the perturbation of the neuronal states by stimulus presentations, in this case, faces, objects and random-dot patterns. Endogenous connections reflect directed coupling among areas, averaged over experimental conditions. Connections with bilinear modulation show changes in coupling induced by an experimental factor. Connections in our models could be bilinearly modulated by the presence of motion or by facial form. Nonlinear modulations reflect changes in coupling induced by another ROI. Note that non-linear modulations can be used to explain bilinear effects. For example, bilinear modulation of faces versus non-faces might arise on a connection because it is non-linearly modulated by a face-selective area. We used nonlinear parameters to examine how areas in one pathway (e.g., facial form pathway) affect information flow in the other pathway (e.g., motion pathway). With DCM, we varied the presence or absence of endogenous, bilinear or

For Peer Review

nonlinear parameters and performed Bayesian model comparisons to identify the optimal model architecture. We first compared a bilinear model space, where we identified the model that best explained how bilinear influences of motion and facial form explain STS responses. We then performed a second model comparison, using nonlinear models, to identify the brain areas whose activity could optimally account for the motion and facial form modulation we observed in the optimal bilinear model (See below).

Before model comparison, we formulated a 'base model' that accounted for; (a) the fact that the entire network is driven by face and non-face stimuli (objects and patterns), and (b) that OFA and FFA respond preferentially to faces, while V5 responds preferentially to motion. We drove the network by face, object and pattern stimulation by including an input area (BA18) that responded to these three stimuli and is in a position to propagate neuronal signals throughout the network. This BA18 area corresponds to, low-level visual cortex, which is known to respond to visual stimuli generally and to feedforward its responses to higher visual areas. Consistent with this role for BA18, we accounted for face-selectivity by adding (bilinear) modulation by faces to the connection from BA18 to OFA. Similarly, we accounted for motion-sensitivity by adding modulation by motion to the connection from BA18 to V5. Model comparison then proceeded by varying other properties of this base model.

We compared individual models (Table 1) and model families (Table 2) using their relative log-evidences and posterior probabilities – assuming all participants used the same connectivity architecture (Penny et al. 2004, Penny et al. 2007, Stephan et al. 2010). The main focus of our model comparisons was to determine whether motion sensitivity that is selective to facial form in the STS could be explained by network interactions between motion and facial form pathways. We considered two alternative mechanisms for this interaction. First, the connection between a face-selective area (OFA and/or FFA) and the

For Peer Review

STS could be modulated by motion. Second, the connection between a motion-sensitive area (V5) and STS could be modulated by facial form. We first cast these hypotheses in the form of bilinear models, and performed a model comparison using 16 models (cells in Table 1). These 16 models were divided into four model families, corresponding to the mechanisms that could produce the form by motion interaction in STS (See columns in Table 1). These four families tested (1) face modulation of the motion pathway to STS from V5, (2) motion modulation of the face pathway to STS from OFA, (3) motion modulation of the face pathway to STS from OFA and (4) motion modulation of both face pathways from OFA and FFA. These four families were crossed with two other variants of model, which tested incidental hypotheses, as shown in the rows of Table 1. First, the bilinear models could be either "full connectivity", with all possible endogenous connections, or the connectivity could be sparse. The sparse models were motivated by a previous study of magnetoencephalographic induced responses that showed no endogenous connectivity between FFA and the STS and only feedforward connections (Furl et al. 2013a). Second, the bilinear models either possessed modulation by faces on only the connection from BA18 to OFA ("OFA only" rows in Table 1) or possessed modulation on the connections from BA18 to both OFA and FFA ("OFA/FFA" rows in Table 1).

Our bilinear model comparison revealed that facial form information modulated the connection from V5 to STS (See Results for more information). However, this result does not identify the mechanism that causes this modulation. To do this, we used non-linear models in which face-selective areas can directly influence the connection from V5 to STS. Here, we could test whether the face-selective responses in OFA, FFA or both influenced the motion information propagating to STS from V5. Non-linear influences from these face-selective areas could account for the bilinear modulation of faces that we observed. Note that, in principle, it would be preferable to test all our hypotheses in one non-linear model space. In

For Peer Review

this case, we would have compared non-linear models where face-selective areas influence the connection from V5 to STS against non-linear models where the motion-sensitive area V5 influences connections from the face-selective areas. However, the multiplicative nature of nonlinear terms (Stephan et al. 2008) results in mathematically symmetrical nonlinear DCMs, preventing this model comparison in practice. We therefore first tested bilinear models which showed that faces modulated the connection from V5 to STS and then we tested non-linear models to identify a possible face-selective area responsible for the bilinear modulation of faces.

RESULTS

ROI specification

We located ROIs in individual participants. We used the contrast of faces, objects and patterns versus fixation to identify BA18; the contrast of dynamic and static faces versus dynamic and static objects and patterns to identify the conventional face-selective areas OFA and FFA; the contrast of dynamic versus static faces to identify the motion-sensitive area V5; and the contrast (dynamic faces > static faces) > (dynamic objects/patterns > static objects/patterns) to identify face-specific motion sensitivity in the STS. For display purposes, Fig. 1 illustrates the results of this contrast in the STS at the group level, using the 11 participants who showed every ROI (peak voxel MNI: 56 -24 -8). This STS area was observed at $P < 0.005$ uncorrected where it also met the $P < 0.0001$ threshold for family-wise error correction at the cluster level (Brett et al. 2003).

Group-level ROI analyses

For Peer Review

Fig. 2 shows the response patterns in our ROIs at the group level using ANOVAs with motion (dynamic or static) and category (face, object or pattern) as factors, followed by post-hoc tests (Tukey honest significant difference corrected $P < 0.05$). Some of the ANOVA effects duplicate the contrasts used to define the ROIs including the main effect of category in face-selective ROIs and the mean effect of motion in motion-sensitive ROIs. We include these tests here for completeness and to illustrate the quantitative patterns of means within the voxels identified in the ROIs However, our main conclusions from the ROI analyses are drawn from orthogonal ANOVA effects to preclude biased inferences. These include effects of motion in face-selective ROIs and effects of category in motion-sensitive ROIs.

BA18 (Fig. 2a) showed robust responses in every condition, with enhanced responses to dynamic patterns, resulting in a motion × category interaction ($F(1,50) = 9.30$, $P < 0.001$) and a significant pairwise difference between dynamic and static patterns (there were no other significant pairwise effects). V5 (Fig. 2b) showed robust responses to all dynamic stimuli, with no positive responses to any static stimulus, and significant differences between dynamic versus static versions of all three categories of stimuli, resulting in our hypothesized main effect of motion $F(1,50) = 304.65$, $P < 0.001$. Because motion sensitivity was numerically smaller for faces than for objects and patterns, there was a motion × category interaction ($F(1,50 = 9.52$, $P = 0.009$). The STS also showed a motion × category interaction ($F(1,50) = 18.72$, $P < 0.001$), but because of a different response pattern than for V5 and BA18. In the STS, pairwise tests showed significant motion sensitivity only for faces, but not for objects or random dot patterns. Neither ventral area showed any motion × category interaction (OFA: $P = 0.077$; FFA: $P = 0.264$), although we detected main effects of motion (OFA: $F(1,50) = 17.73$, $P < 0.001$; FFA: $F(1,50) = 16.51$, $P < 0.001$) in addition to the main effect of category (OFA: $F(1,50) = 91.06$, $P < 0.001$; FFA: $F(1,50) = 108.79$, $P < 0.001$). Closer inspection using pairwise tests showed that the main effect for OFA was driven by

For Peer Review

motion sensitivity for patterns but no significant motion sensitivity for faces or objects. For the FFA, no category showed significant motion sensitivity when tested alone. In summary, only the STS showed motion sensitivity that was selective for faces. V5 showed motion sensitivity to faces as well as objects and patterns, while BA18, the OFA and FFA showed no evidence for motion sensitivity to faces.

Connectivity models

Our ROI analysis confirmed the presence of dorsal temporal motion sensitivity in V5, facial motion sensitivity in the STS, and ventral temporal face selectivity in the OFA and FFA. We used connectivity modeling to test how interactions between the dorsal motion-sensitive and the ventral face-selective pathways could give rise to motion sensitivity that is selective to faces in the STS. We first compared bilinear models to test whether STS responses might be explained by a network, either in which faces modulate dorsal motion-sensitive pathway connections from V5 to STS, or in which motion modulates the ventral face-selective pathway connections from the OFA and/or FFA to the STS. This space of bilinear models further explored as secondary hypotheses whether (a) endogenous connectivity is full or sparse and (b) face selectivity in the ventral pathway arises from modulation by faces on only forward connections to the OFA, or if forward connections to the FFA are modulated by faces as well ("OFA only" and "OFA/FFA" rows in Table 1). Of the 16 models we tested, we found a high posterior probability (near 1.0) favoring a model where faces modulate the dorsal motion-sensitive connections from V5 to the STS. For our secondary hypotheses, we found (a) full (rather than sparse) endogenous connectivity and (b) face modulation on connections from BA18 to the OFA only (and not also to the FFA). These properties of the optimal model were confirmed using model family comparisons (Table 2).

For Peer Review

Having established that faces modulate the dorsal motion-sensitive connection from V5 to the STS, we assumed that this face modulation arose from the activity in a face-selective area in the ventral pathway. We therefore used three additional non-linear models to test whether face modulation on the dorsal motion-sensitive connections from V5 to the STS was more likely to arise from face-selective responses in OFA or FFA or both. We found a near perfect posterior probability favoring the model where the OFA, but not the FFA (nor both), modulates the connection from V5 to the STS.

DISCUSSION

We show that motion sensitivity to facial form in the STS was best explained by a DCM where transmission of motion information from V5 to the STS is gated or modulated by information about facial form. Face-selective responses in the OFA most likely implemented this gating. This model provides a network-based account for the emergence of face-selective motion sensitivity in the STS and, perhaps, could also explain the integration of motion and form information when viewing biological motion.

Responses to biological motion constitute a type of form-selective motion sensitivity, in the sense that they respond only to conjunctions of motion with specific forms. Consequently, studies in this area often characterize perception of biological motion as resulting from a mixture of contributions of form and motion representations (Thompson and Baccus 2012), which may be transmitted by separate occipitotemporal pathways (Giese and Poggio 2003) and may converge on the STS, where the form and motion information is combined (Oram and Perrett 1996; Vaina et al. 2001; Thompson et al. 2005; Lange and Lappe 2006). Not surprisingly, the dominant theoretical frameworks from the face perception literature are similarly structured, with distinct pathways representing facial form and movements. Low-level facial feature information might be processed in the OFA and then

For Peer Review

fed-forward into dorsal and ventral pathways (Haxby et al. 2000). Information about static form or invariant facial features is considered to be represented in ventral areas like the OFA and FFA (O'Toole et al. 2002; Calder and Young 2005; Calder 2011; Haxby and Gobbini 2011), which are selective for facial form (Kanwisher et al. 1997). More dorsal areas, such as the STS (Haxby et al. 2000; Haxby and Gobbini 2011) and V5 (O'Toole et al. 2002), however, are more sensitive to facial motion than OFA and FFA (Schultz and Pilz 2009; Trautmann et al. 2009; Pitcher et al. 2011; Foley et al. 2012; Grosbras et al. 2012; Schultz et al. 2013). These dorsal areas may employ motion-based representations to recognize the changeable aspects of faces (Haxby and Gobbini, 2011; Foley et al., 2012). While our results suggest that the STS is driven by facial motion information, they further show that STS responses are not dependent on a single, motion-based pathway, but instead are the result of non-linear interactions between motion and form pathways.

A previous study using connectivity analyses (Foley et al., 2012) showed that responses in inferior occipital gyrus and STS were more correlated for dynamic than for static faces. Indeed, a model like this could plausibly explain the form by motion interaction that we observed in the STS. In this case, the STS would receive signals from OFA that are already form-dependent (because OFA is face-selective) and the addition of motion modulation on the OFA to STS connection would introduce an interaction of form and motion in the STS. However, our bilinear model space tested a family of models with this property (Table 1, column 2, Table 2, row 2) and it was suboptimal, compared to another means of introducing a form by motion interaction in STS. The more likely model family showed that facial form modulated the motion-sensitive responses conveyed to STS from V5 (Table 1, column 1, Table 2, row 1). We then showed that this facial form modulation could occur when OFA activity (which is selective to facial form) non-linearly modulates the flow of motion information from V5 to STS. In other words, the OFA acted as a modulatory gain

For Peer Review

control on the "driving signal" in the motion pathway, rather than simply conveying the motion information itself (Foley et al., 2012). These non-linear interactions also go beyond previous work because they predict hypothetical neural mechanisms (Stephan et al. 2008), where a neural population in the OFA might introduce short term synaptic plasticity in its target (the STS) by altering its receptivity to other neural populations that drive it (V5). Our results therefore provide neural-level hypotheses to be explored in the non-human primate, which has well-characterized visual areas sensitive to faces (Tsao et al. 2006) as well as motion (Dubner and Zeki 1971; Desimone and Ungerleider 1986; Nelisson et al. 2006), including biological motion (Oram and Perrett 1994; 1996; Nelisson et al. 2011).

Our study focused on explaining STS motion-sensitive responses to faces versus objects. However, some areas in the STS are well-known to be generally sensitive to biological forms. Our results suggest a mechanism that might generalize to integration of motion and form in cases of biological motion, although this requires confirmation using speech movements, grasping actions or point light displays. We can claim that our STS area is not involved simply in representing low-level motion or motion-defined shape features, because it did not show sensitivity to random-dot patterns with motion-defined contours. We can also claim that our STS area did not show sensitivity to motion that depicts complex forms, as it was not sensitive to object motion (Beauchamp et al. 2002; 2003; Pitcher et al. 2012). However, we do not know how sensitive our STS area is to non-face body movements. There is evidence that different areas in the STS show sensitivity to specific body parts (Wheaton et al. 2004; Thompson et al. 2007; Grosbras et al. 2012). However, motion sensitivity to different stimuli may overlap as well. The posterior STS responds in common to a variety of different types of movements when they are compared to scrambled movements without form cues (Santi et al. 2003; Thompson et al. 2007; Grosbras et al. 2012). And similar areas in the posterior STS are associated with point-light body actions as

For Peer Review

well as faces (Hein and Knight 2008). The interaction of facial form and motion we observed, however, showed its peak effect in a more anterior area of STS than that commonly observed for point-light displays of bodily actions. Thus, any overlap between the STS area we observed and motion sensitivity to other types of complex stimuli such as bodies still needs to be established.

Our results suggest that access of facial motion to the STS is dependent on an occipital area that is selective to facial form, the OFA. It remains to be seen whether other form-selective areas perform similar gating on motion information in other stimulus domains. For example, the extrastriate or fusiform body areas might gate connections between V5 and the STS during body perception. Hein and Knight (2008) hypothesized that STS responses to actions associated with theory of mind inferences or audiovisual speech movements might be dependent, respectively, on responses in medial and inferior prefrontal areas. It remains unclear whether these areas might have a driving (like V5) or a gating/modulatory (like the OFA) relationship with STS responses. Inferior frontal cortex, in particular, has been implicated in perception of facial and other types of biological motion (Saygin et al. 2004; Wheaton et al. 2004; Casile et al. 2010; Furl et al. 2010; van Kemenade et al. 2012). Indeed, inferior frontal involvement has been characterized as a top-down process involving motor representations coded by mirror neuron responses (Caggiano et al. 2011; Kilner 2011; Nelissen et al. 2011). We did not observe reliable inferior frontal responses in our individual participants useful for modeling using our current data. However, connectivity analyses like DCM may provide a powerful technique for measuring top-down influences on STS responses to dynamic visual stimuli.

In summary, we present a connectivity model of fMRI data that explains, in terms of network dynamics, the origin of motion sensitivity that is selective to facial form in the STS. We demonstrate how responses in the STS can depend on interactions between information

For Peer Review

flow in a dorsal motion-sensitive pathway and a ventral facial form-selective pathway. The presence of information about facial form enhanced the ability of the motion-sensitive area V5 to influence responses in the STS. This gain control modulation likely originated in the OFA. Our model of network interactions provides a plausible mechanistic explanation for how form and motion information are integrated when viewing biological motion. This new perspective on network-level causes of brain responses to dynamic stimuli opens several future research avenues.

For Peer Review

ACKNOWLEDGEMENTS

We are grateful to Christopher Fox for supplying the dynamic object stimuli. This work was supported by funding from the United Kingdom Economic and Social Research Council [RES-062-23-2925] to NF. AJC (MC_US_A060_5PQ50) and RNH (MC_US_A060_0046) are supported by the Medical Research Council. The authors declare no competing financial interests.

For Peer Review

REFERENCES

Beauchamp MS, Lee KE, Haxby JV, Martin A. 2003. FMRI responses to video and point-light displays of moving humans and manipulable objects. J Cogn Neurosci 15:991-1001.

Beauchamp MS, Lee KE, Haxby JV, Martin A. 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34:149-159.

Brett M, Penny WD, Kiebel SJ. 2003. Introduction to random field theory. In: Frackowiak RSJ, Friston KJ, Frith C, Dolan R, KJ, Price CJ, Zeki S, Ashburner J, Penny WD, editors Human Brain Function, 2nd ed. San Diego (CA): Academic.

Caggiano V, Fogassi L, Rizzolatti G, Pomper JK, Thier P, Giese MA, Casile A. 2011. View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. Curr Biol 21:144-148.

Calder AJ. 2011. Does facial identity and facial expression recognition involve separate visual routes? In: Calder AJ, Rhodes G, Johnson M, Haxby JV, editors. The Oxford Handbook of Face Perception. Oxford (UK): Oxford University Press.

Calder AJ, Young AW. 2005. Understanding recognition of facial identity and facial expression. Nat Rev Neurosci 6: 641-651.

For Peer Review

Casile A, Dayan E, Caggiano V, Hendler T, Flash T, Giese MA. 2010. Neuronal encoding of human kinematic invariants during action observation. Cereb Cortex 20:1647-1655.

Desimone R, Ungerleider LG. 1986. Multiple visual areas in the caudal superior temporal sulcus of the macaque. J Comp Neurol 8: 164-189.

Dubner R, Zeki SM. 1971. Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res 35: 528-532.

Foley E, Rippon G, Thai NJ, Longe O, Senior C. 2012. Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. J Cogn Neurosci 24:507–520.

Fox CJ, Iaria G, Barton JJ. 2009. Defining the face processing network: optimization of the functional localizer in fMRI. Hum Brain Mapp 30:1637-1651.

Friston KJ. 1997. Testing for anatomically specified regional effects. Hum Brain Mapp 5:133-136.

Friston, KJ., Harrison, L. & Penny, W. 2003. Dynamic Causal Modelling. Neuroimage 19:1273-1302.

For Peer Review

Furl N, Coppola R, Averbeck BB, Weinberger DR. 2013a. Cross-Frequency Power Coupling Between Hierarchically Organized Face-Selective Areas. Cereb Cortex. (Epub ahead of print) doi: 10.1093/cercor/bht097.

Furl N, Henson RN, Friston KJ, Calder AJ. 2013b. Top-down control of visual responses to fear by the amygdala. J Neurosci 33:17435-43.

Furl N, van Rijsbergen NJ, Kiebel SJ, Friston KJ, Treves A, Dolan RJ. 2010. Modulation of perception and brain activity by predictable trajectories of facial expressions. Cereb Cortex 20:694-703.

Giese MA, Poggio T. 2003. Neural mechanisms for the recognition of biological movements. Nat Rev Neurosci 4:179-192.

Grosbras MH, Beaton S, Eickhoff SB. 2012. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. Hum Brain Mapp 33:431-454.

Haxby JV, Gobbini MI. 2011. Distributed neural systems for face perception. In: Calder AJ, Rhodes G, Johnson M, Haxby JV, editors. The Oxford Handbook of Face Perception. Oxford: Oxford University Press.

Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. Trends Cogn Sci 4:223-233.

For Peer Review

Hein G, Knight RT. 2008. Superior temporal sulcus--It's my area: or is it? J Cogn Neurosci 20:2125-2136.

Jastorff J, Orban GA. 2009. Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. J Neurosci 29:15-29.

Jastorff J, Popivanov ID, Vogels R, Vanduffel W, Orban GA. 2012. Integration of shape and motion cues in biological motion processing in the monkey STS. Neuroimage 60:911-21.

Johansson G. 1973. Visual perception of biological motion and a model for its analysis. Percept Psychophys 14:195-204.

Kanwisher, N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302-4311.

Kilner JM. 2011. More than one pathway to action understanding. Trends Cogn Sci 15:353-357.

Lange J, Lappe M. 2006. A model of biological motion perception from configural form cues. J Neurosci 26:2894-2906.

Nelissen K, Borra E, Gerbella M, Rozzi S, Luppino G, Vanduffel W, Rizzolatti G, Orban GA. 2011) Action observation circuits in the macaque monkey cortex. J Neurosci 31:3743-3756.

For Peer Review

Nelissen K, Vanduffel W, Orban GA. 2006. Charting the lower superior temporal region, a new motion-sensitive region in monkey superior temporal sulcus. J Neurosci 26: 5929-5947.

Oram MW, Perrett DI. 1994. Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. J Cogn Neurosci 6:99-116.

Oram MW, Perrett DI. 1996. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. J Neurophysiol 76:109-129.

O'Toole AJ, Roark DA, Abdi H. 2002. Recognizing moving faces: a psychological and neural synthesis. Trends Cogn Sci 6:261-266.

Penny WD, Stephan KE, Mechelli A, Friston KJ. 2004. Comparing dynamic causal models. Neuroimage 22:1157-1172.

Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP. 2010. Comparing families of dynamic causal models. PLoS Comput Biol 8:e1000709.

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage 56:2356-2363.

Santi A, Servos P, Vatikiotis-Bateson E, Kuratate T, Munhall K. 2003. Perceiving biological motion: dissociating visible speech from walking. J Cogn Neurosci 15:88-809.

For Peer Review

Saygin AP, Wilson SM, Hagler DJ Jr, Bates E, Sereno MI. 2004. Point-light biological motion perception activates human premotor cortex. J Neurosci 24:6181-6188.

Schultz J, Brockhaus M, Bülthoff HH, Pilz KS. 2013. What the human brain likes about facial motion. Cereb Cortex 23:1167-78.

Schultz J, Pilz KS. 2009. Natural facial motion enhances cortical responses to faces. Exp Brain Res 194:465– 475.

Stephan KE, Kasper L, Harrison LM, Daunizeau J, den Ouden HE, Breakspear M, Friston KJ. 2008. Nonlinear dynamic causal models for fMRI. Neuroimage 42:649-662.

Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ. 2010. Ten simple rules for dynamic causal modeling. Neuroimage 49:3099-109.

Thompson JC, Baccus W. 2012. Form and motion make independent contributions to the response to biological motion in occipitotemporal cortex. Neuroimage 59:625-634.

Thompson JC, Clarke M, Stewart T, Puce A. 2005. Configural processing of biological motion in human superior temporal sulcus. J Neurosci 25:9059-9066.

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 37:966-73.

For Peer Review

Trautmann SA, Fehr T, Herrmann M. 2009. Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. Brain Res 1284:100 –115.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS. 2006. A cortical region consisting entirely of face-selective cells. Science 311:670-674.

Vaina LM, Gross CG. 2004. Perceptual deficits in patients with impaired recognition of biological motion after temporal lobe lesions. Proc Natl Acad Sci USA 101:16947-16951.

Vaina LM, Solomon J, Chowdhury S, Sinha P, Belliveau JW. 2001. Functional neuroanatomy of biological motion perception in humans. Proc Natl Acad Sci USA 98:11656-11661.

Van der Schalk J, Hawk ST, Fischer AH, Doosje BJ. 2011. Moving faces, looking places: The Amsterdam Dynamic Facial Expressions Set (ADFES). Emotion 11: 907-920.

van Kemenade BM, Muggleton N, Walsh V, Saygin AP. 2012. Effects of TMS over premotor and superior temporal cortices on biological motion perception. J Cogn Neurosci 24:896-904.

Van Oostende S, Sunaert S, Van Hecke P, Marchal G, Orban GA. 1997. The kinetic occipital (KO) region in man: an fMRI study. Cereb Cortex 7, 690-701.

For Peer Review

Wheaton KJ, Thompson JC, Syngeniotis A, Abbott DF, Puce A. 2004. Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. Neuroimage 22:277-288.

For Peer Review

Table 1. Bilinear model evidences and posterior probabilities

		faces modulate V5 to STS	motion modulates OFA to STS	motion modulates FFA to STS	motion modulates OFA/FFA to STS
full	OFA only	**285.20 (1)**[a]	231.47 (0)	222.64 (0)	87.92 (0)
	OFA/FFA	74.15 (0)	81.65 (0)	92.11 (0)	55.79 (0)
sparse	OFA only	12.71 (0)	9.19 (0)	9.49 (0)	0 (0)
	OFA/FFA	12.62 (0)	9.13 (0)	9.43 (0)	0.09 (0)

[a] We compared 16 bilinear DCMs on the basis of their model evidences (with posterior probabilities shown in parentheses). The highest evidence model is shown in bold.

For Peer Review

Table 2. Bilinear family model evidences and posterior probabilities

faces modulate V5 to STS	2.89 (1)
motion modulates OFA to STS	0 (0)
motion modulates FFA to STS	0 (0)
motion modulates OFA/FFA to STS	0 (0)
full	2.89 (1)
sparse	0 (0)
OFA only	2.89 (1)
OFA/FFA	0 (0)

[a] We compared evidences (with posterior probabilities shown in parentheses) aggregated over "families" of bilinear DCMs that shared specific features of interest. The first four rows compare four families that could each differently explain the face-specific motion sensitivity in the STS. The fifth and sixth rows compare families with full versus sparse endogenous connectivity. The seventh and eighth rows compare a family using modulation of faces on the connection from BA18 to OFA versus a family using modulation on connections from BA18 to both OFA and FFA.

For Peer Review

FIGURE LEGENDS

Figure 1. Group-level whole-brain analysis. (a) Results of contrast (dynamic faces > static faces) > (dynamic non-faces > static non-faces). Voxels showing significant effects at $P < 0.005$ (uncorrected) are projected on an inflated cortical surface of the right hemisphere in MNI space. STS = superior temporal sulcus.

Figure 2. Group-level region of interest (ROI) analysis. (a) Mean responses in Brodmann area 18 (BA18) to faces, objects and random-dot patterns; (b) Mean responses in V5; (c) Mean responses in the superior temporal sulcus (STS); (d) Mean responses in the occipital face area (OFA); (e) Mean responses in the fusiform face area (FFA). Graph titles describe contrast used to define ROI.

Figure 3. Optimal dynamic causal models (a) the optimal bilinear model generates motion sensitivity that is selective to facial form in the superior temporal sulcus (STS) when faces modulate connections from the motion-sensitive V5 to STS. Bilinear modulations indicated by black arrows, endogenous connections indicated in light grey. The optimal model had full endogenous connectivity. (b) The optimal nonlinear model shows that the face-selective occipital face area (OFA) is the most likely origin of face modulation on the connections from V5 to STS. Bilinear and nonlinear modulations indicated by black arrows, endogenous connections indicated in light grey. FFA = fusiform face area, BA18 = Brodmann area 18.

For Peer Review

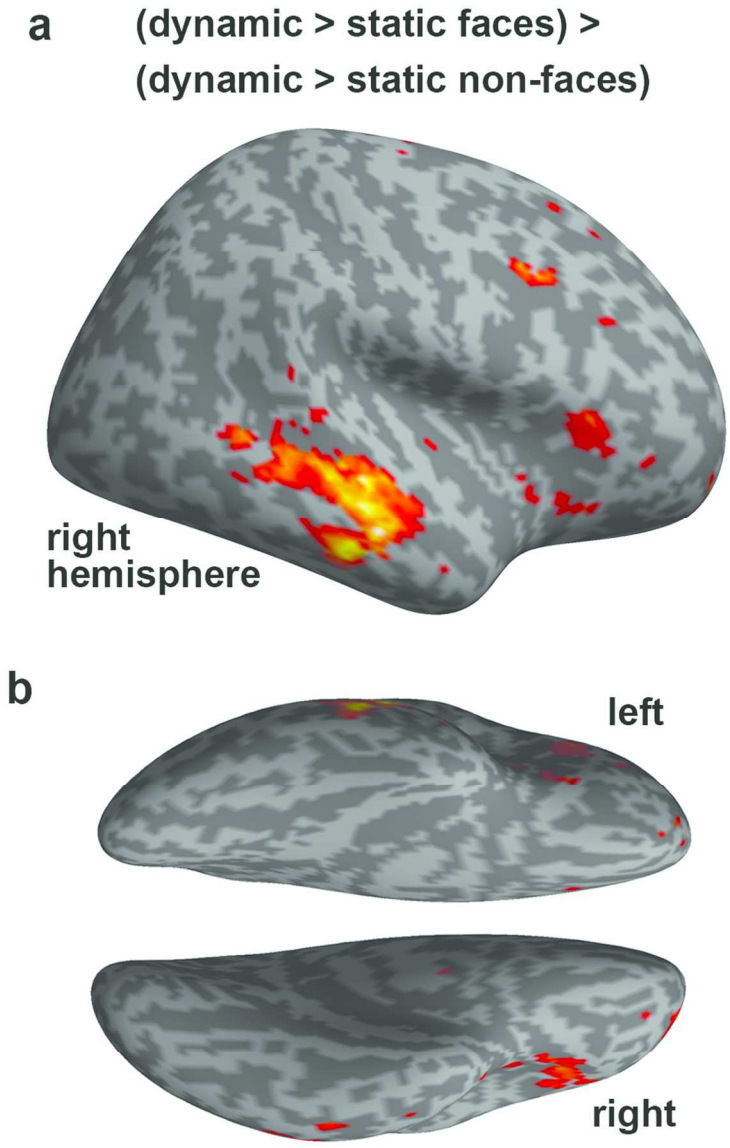

Figure 1. Group-level whole-brain analysis. (a) Results of contrast (dynamic faces > static faces) > (dynamic non-faces > static non-faces). Voxels showing significant effects at P < 0.005 (uncorrected) are projected on an inflated cortical surface of the right hemisphere in MNI space. Blue circle indicates the peak voxel. STS = superior temporal sulcus.
85x132mm (300 x 300 DPI)

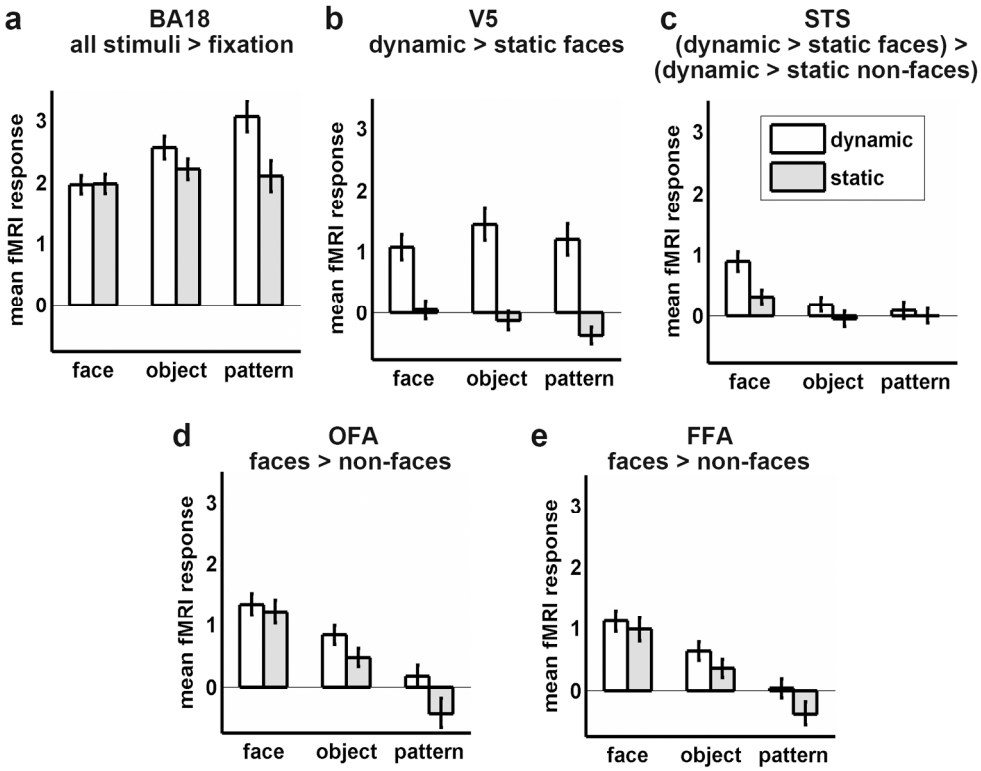

Figure 2. Group-level region of interest (ROI) analysis. (a) Mean responses in Brodmann area 18 (BA18) to faces, objects and random-dot patterns; (b) Mean responses in V5; (c) Mean responses in the superior temporal sulcus (STS); (d) Mean responses in the occipital face area (OFA); (e) Mean responses in the fusiform face area (FFA). Graph titles describe contrast used to define ROI.
176x199mm (300 x 300 DPI)

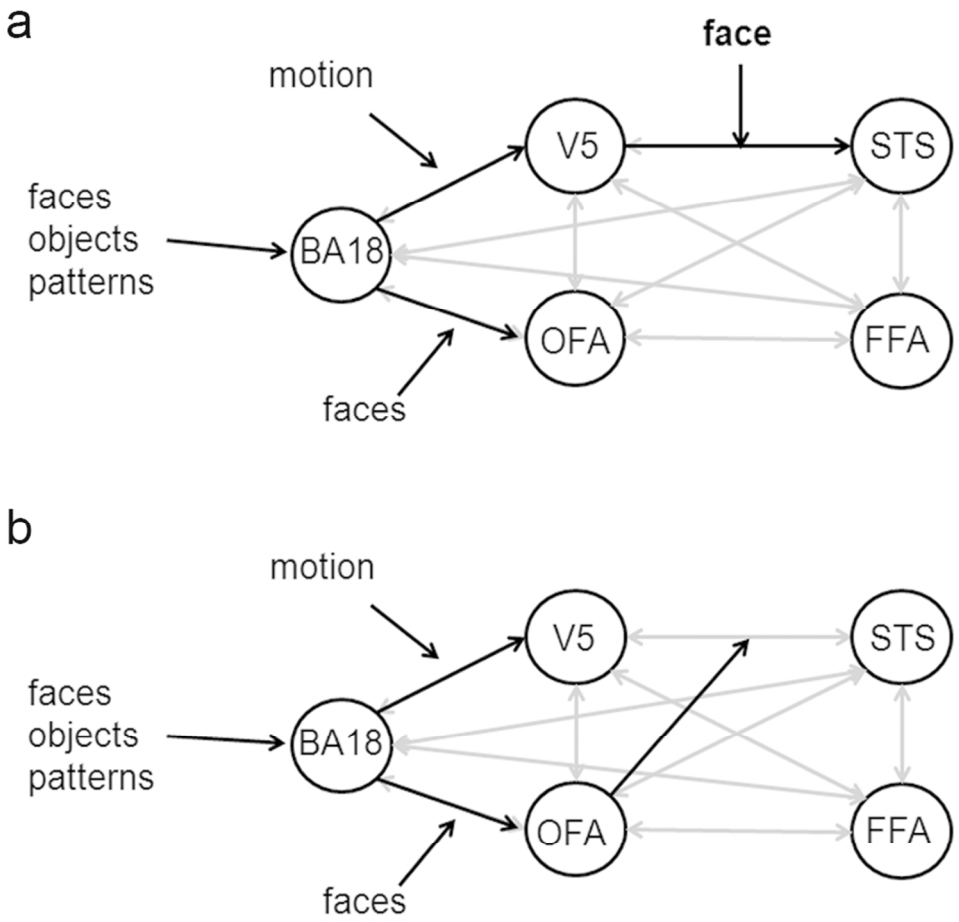

Figure 3. Optimal dynamic causal models (a) the optimal bilinear model generates motion sensitivity that is selective to facial form in the superior temporal sulcus (STS) when faces modulate connections from the motion-sensitive V5 to STS. Bilinear modulations indicated by black arrows, endogenous connections indicated in light grey. The optimal model had full endogenous connectivity. (b) The optimal nonlinear model shows that the face-selective occipital face area (OFA) is the most likely origin of face modulation on the connections from V5 to STS. Bilinear and nonlinear modulations indicated by black arrows, endogenous connections indicated in light grey. FFA = fusiform face area, BA18 = Brodmann area 18.
85x81mm (300 x 300 DPI)