doi:10.1093/scan/nsu028 SCAN (2015) 10, 129–135

Stimulus-induced reversal of information flow through a cortical network for animacy perception

Sarah Shultz, Rebecca N. van den Honert, Andrew D. Engell, and Gregory McCarthy
Department of Psychology, Yale University, New Haven, CT 06520-8205, USA

Decades of research have demonstrated that a region of the right fusiform gyrus (FG) and right posterior superior temporal sulcus (pSTS) responds preferentially to static faces and biological motion, respectively. Despite this view, both regions activate in response to both stimulus categories and to a range of other stimuli, such as goal-directed actions, suggesting that these regions respond to characteristics of animate agents more generally. Here we propose a neural model for animacy detection composed of processing streams that are initially differentially sensitive to cues signaling animacy, but that ultimately act in concert to support reasoning about animate agents. We use dynamic causal modeling, a measure of effective connectivity, to demonstrate that the directional flow of information between the FG and pSTS is initially dependent on the characteristics of the animate agent presented, a key prediction of our proposed network for animacy detection.

Keywords: animate; dynamic casual modeling; effective connectivity; fusiform; pSTS

INTRODUCTION

Detecting and reasoning about animate entities are critical skills for survival and successful social interactions. Given the centrality of animacy detection to social life, it is perhaps not surprising that our perceptual system is tuned to efficiently detect characteristics of animate agents (Scholl and Tremoulet, 2000; Simion *et al.*, 2008).

Three well-studied characteristics of animate agents are (i) surface features, such as a human's face or limbs (Carey and Spelke, 1994, 1996; Baron-Cohen, 1995; Guajardo and Woodward, 2004); (ii) biological motion, such as self-propelled motion (Premack, 1990; Leslie, 1994, 1995; Baron-Cohen, 1995) and non-rigid transformation (Gibson *et al.*, 1978); and (iii) goal-directed actions—actions that are purposeful and efficient given the constraints of the surrounding environment (Csibra *et al.*, 1999, 2003; Gergely and Csibra, 2003; Bíró *et al.*, 2007). These animacy cues activate a network of ventral and lateral occipitotemporal cortical regions, most prominently the right fusiform gyrus (FG), right posterior superior temporal sulcus (pSTS) and adjacent temporoparietal junction (TPJ) (Puce *et al.*, 1995, 1996, 1998; McCarthy *et al.*, 1997; Puce and Perrett, 2003). The FG and the pSTS are thought to play a critical role in processing two types of animacy cues: static human-like surface features (such as facial form) and biological motion, respectively (Puce *et al.*, 1995; Kanwisher *et al.*, 1997; Grossman and Blake, 2002; Puce and Perrett, 2003).

Studies of the neural correlates of animacy perception have traditionally focused on localizing the processing of specific animacy cues to particular brain regions. However, it is important to note that in the natural world cues signaling the presence of animacy typically co-occur and are inherently linked: animate agents have a human form, move in biologically plausible ways and engage in meaningful goal-directed behavior. Whenever a face is detected, there is a high probability that it will be accompanied by biological motion and vice versa. Given that both cues co-occur and are critical for detecting and reasoning about animate agents, it is plausible that viewing any one cue would cause activity to propagate through the entire network of regions involved in animacy perception.

Indeed, although the FG and pSTS are thought to play key roles in processing invariant aspects of a face and biological motion, respectively, many studies have reported that *both* regions activate in response to static faces and biological motion (Puce *et al.*, 1995, 1996; Kanwisher *et al.*, 1997; Rossion *et al.*, 2012). The most direct evidence of the functional similarity between the FG and pSTS comes from a recent study that directly compared activation maps from two large data sets, designed to localize face-sensitive and biological motion-sensitive regions of cortex (Engell and McCarthy, 2013). Both datasets activated strikingly similar clusters in bilateral occipitotemporal cortices, including the FG and the pSTS. This suggests that the FG and pSTS might be responsive to both stimulus categories or to properties of animate entities more generally, rather than responding preferentially to faces or biological motion. Indeed, a recent study demonstrated that the FG and pSTS activate in response to goal-directed actions, an aforementioned cue for animacy, even in the absence of faces or biological motion (Shultz and McCarthy, 2012), suggesting that the FG and pSTS may be sensitive to a wide range of animacy cues. Given the substantial overlap in the response profiles of the FG and pSTS, what functionally differentiates the roles of these regions and, more importantly, how might they work together as part of an extended network for detecting animate agents and their underlying intentions?

Here, we propose a model of animacy detection that will provide a framework for answering these questions. The model posits that the FG and pSTS are part of a distributed network of brain regions that act in concert to support the detection of animate agents and their underlying intentions. According to our model, animate agents are detected through three processing streams that are initially differentially sensitive to human-like surface features (faces and body limbs), biological motion and goal-directed actions. Information about animate agents is then shared across streams for further processing (Figure 1). This model builds upon and is consistent with existing models of face processing that posit different pathways in ventral and lateral temporal cortex for the processing of human form and motion (Haxby *et al.*, 2000; Hoffman and Haxby, 2000; Adolphs, 2001; O'Toole *et al.*, 2002; Beauchamp *et al.*, 2003) and more recent models of high-level vision that posit distinct but interacting processing streams for aspects of visual perception (including the recognition of object form), aspects

Received 11 June 2013; Revised 29 January 2014; Accepted 10 February 2014
Advance Access publication 13 March 2014
This work was supported by funding from the National Institutes of Mental Health (MH-005286 to G.M.) and the National Science Foundation (Graduate Research Fellowship to S.S.).
Correspondence should be addressed to Gregory McCarthy, Department of Psychology, Yale University, PO Box 208205, New Haven, CT 06520-8205, USA. E-mail: gregory.mccarthy@yale.edu

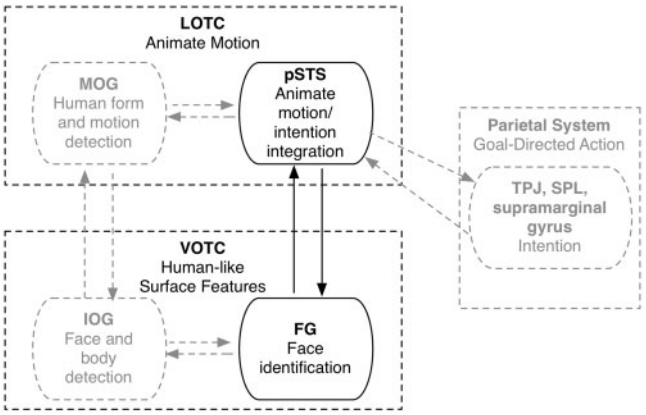

Fig. 1 A proposed neural model for identifying animate agents and their underlying intentions. We propose that three processing streams (the VOTC, LOTC and parietal system) are initially differentially sensitive to three types of cues signaling animacy (human-like surface features, animate motion and goal-directed actions). Nodes of the proposed network and their functions are listed within each processing stream. Arrows indicate direction of information flow between nodes of the proposed network. Solid black lines indicate hypothesized nodes and connections that are tested in this study. Dashed gray lines indicate hypothesized nodes and connections that are not tested in this study.

of spatial vision (including visually guided actions) and multimodal processing (i.e. integration of form and motion information) (Weiner and Grill-Spector, 2011).

The first stream of our proposed model for animacy detection recruits the ventral occipitotemporal cortex (VOTC) and includes localized areas in the anterior occipital sulcus (termed the 'occipital face area') and the FG (termed the 'fusiform body area' and the 'fusiform face area'). These regions have been implicated in detecting face parts (Pitcher *et al.*, 2007), holistic face processing and face-individuation (Haxby *et al.*, 2000), and perception of body parts (Peelen and Downing, 2005; Schwarzlose *et al.*, 2005). We therefore propose that the VOTC stream is specialized for the detection of human-like surface features, such as faces and body forms. The second stream recruits lateral occipitotemporal cortex (LOTC) and includes localized areas in the middle temporal cortex (area MT), the intersection of the anterior occipital and inferior temporal sulci (termed the 'extra-striate body area') and the pSTS. These regions have been implicated in processing motion (Tootell and Taylor, 1995), human body form and motion (Grossman and Blake, 2002; Puce and Perrett, 2003), and more abstract representations of biological motion, such as geometric shapes that move in a contingent and self-propelled manner (Gao *et al.*, 2012). We therefore propose that the LOTC stream is specialized for the detection of biological or animate motion. The third stream recruits the parietal system, including areas such as the TPJ, superior parietal lobule and supramarginal gyrus, which have been implicated in understanding action and intentions (Bonda *et al.*, 1996; Fogassi *et al.*, 2005). We therefore propose that the parietal processing stream plays a role in processing goal-directed actions.

We propose that the VOTC, LOTC and parietal processing streams are interdependent and act in concert to identify animate agents from their form and motion and infer their intentions from their actions, while acknowledging their initial distinct roles in detecting cues that signal animacy. As mentioned previously, natural cues signaling the presence of animacy typically co-occur. As such, any one animacy cue may cause activation of all nodes of the animacy network. Therefore, a key hypothesis of our model is that the directional flow of activation between the VOTC, LOTC and parietal system depends on the characteristics of the particular stimulus presented. Specifically, we predict that viewing static human faces activates the VOTC, which then activates the LOTC and parietal system, whereas viewing biological motion activates the LOTC, which then activates the VOTC and parietal system. The functional connectivity between the FG and pSTS (Turk-Browne *et al.*, 2010) might then be an important conduit between the VOTC and LOTC streams.

Although our hypothesized model is presented here in its entirety, the goal of this study is to test one critical prediction of our model for animacy detection; that the causal flow of information between the FG and pSTS changes as a function of the animacy cue type. This prediction was selected as a first step toward testing our larger model because of the critical role of the FG and the pSTS in both face and biological motion processing and the functional connectivity between these regions. We use dynamic causal modeling (DCM) (Friston *et al.*, 2003), a technique that allows one to draw conclusions about effective connectivity, or the influence that one region exerts on another region, to test this key prediction. Effective connectivity between the FG and the pSTS was measured during two separate functional MRI (fMRI) localizer tasks, a face localizer and a biological motion localizer. In the face localizer, animacy cues were presented to participants in the form of static faces without any animate motion cues. In the biological motion localizer, animacy cues were presented to participants in the form of biological motion cues *without* the presence of a canonical face.

We defined two model spaces, one for each localizer task (Figure 2), and used Bayesian model selection (BMS) to select the model with the best balance between accuracy and complexity (Stephan *et al.*, 2010). Each model consisted of regions (the FG and pSTS), directional intrinsic connections between these regions so that the past state of one region may influence the future state of another and direct inputs (experimental manipulations, i.e. faces or biological motion). Direct inputs may influence regions directly or they can serve as a proxy for unmodeled upstream regions, such as primary visual areas, that may influence the FG and pSTS. The model spaces were constrained by the assumption, based on previous studies, that both the FG and the pSTS activate in response to both faces and biological motion (Puce *et al.*, 1995; Kanwisher *et al.*, 1997; Puce and Perrett, 2003). As such, in every model both the FG and the pSTS received information about both cue types. The models varied according to three plausible alternatives for how the information was transmitted to each region. The first possibility, referred to as 'telling', is that one region receives direct input (i.e. is stimulus-driven), and then 'tells' the other region about the stimulus through its intrinsic connection (Figure 2, Models 1–4). The second possibility, referred to as 'knowing and telling', is that both regions receive direct inputs about both cue types and 'tell' each other about this information via intrinsic connections (Figure 2, Models 5–7). The third possibility, referred to as 'knowing', is that both regions receive direct inputs about both cue types, but do not share information (Figure 2, Model 8).

Our proposed model of animacy detection predicts that viewing faces will activate the FG, which will then cause activation in the pSTS. This prediction would be supported by models in which *only* the FG receives information about faces as a direct input. The FG then drives activation in the pSTS by 'telling' the pSTS about faces through its intrinsic connection (Figure 2, Model 1 or 2). In contrast, we predict that the directional flow of information will be *reversed* when animacy cues are presented via biological motion. Specifically, we predict that viewing biological motion will activate the pSTS which will then cause activation in the FG. This prediction would be supported by models in which *only* the pSTS receives information about biological motion as a direct input and then 'tells' the FG about this information through its intrinsic connection (Figure 2, Model 3 or 4).

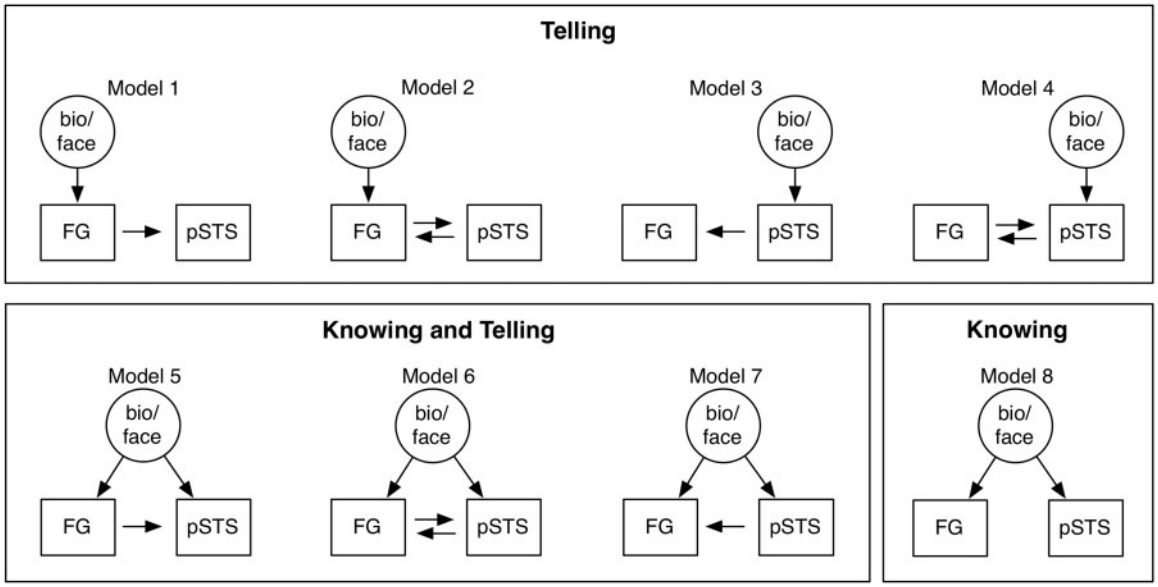

Fig. 2 Model space for the face and biological motion localizer tasks. Inputs are shown as circles and regions are shown as rectangles. Direct inputs are shown as arrows between inputs and regions. Intrinsic connections are shown as arrows between regions. Although faces and biological motion are indicated as inputs, faces are the *only* input modeled in the face task and biological motion is the *only* input modeled in the biological motion task. All models assume that both the FG and the pSTS respond to faces and biological motion, but make different predictions about how the information gets to each region. In 'telling' models, one region receives information about faces or biological motion as a direct input and then 'tells' the other region about this information. In Models 1 and 2 faces or biological motion drives the FG which then influences the pSTS, assuming unidirectional (Model 1) or bidirectional (Model 2) intrinsic connections. In Models 3 and 4 the information flows in the other direction, with faces or biological motion driving the pSTS, which then influences the FG, assuming unidirectional (Model 3) or bidirectional (Model 4) intrinsic connections. In 'knowing and telling' models (Models 5–7), both regions receive information about faces or biological motion as direct inputs and 'tell' each other about this information via unidirectional (Models 5 and 7) or bidirectional (Model 6) intrinsic connections. In 'knowing' models (Model 8), both regions independently receive faces or biological motion as a direct input and do not 'tell' each other about this information.

MATERIALS AND METHODS

Stimuli and experimental design

This study used two separate fMRI localizer tasks, a face localizer and a biological motion localizer. The localizer experiments were collected over several years as part of several independent studies. In the face localizer, participants viewed 4–6 blocks of either static faces (male and female) or static scenes (indoor and outdoor). Blocks lasted 16 or 24 s ($n = 37$ and $n = 3$, respectively, mean = 16.6 s) and were interleaved with a 10, 12 or 16 s rest interval ($n = 9$, 16 and 15, respectively, mean = 13.1 s). In the biological motion localizer, participants saw 4–6 blocks of either 'point-light' movies of biological motion or 'point-light' movies of non-biological motion. Blocks lasted 12 or 32 s ($n = 19$ and 20, respectively, mean = 22.3 s) and were interleaved with a 12 or 16 s rest interval ($n = 19$ and 20, respectively, mean = 14.1 s). Independent samples *t*-tests revealed that the face and biological motion localizer task differed in both block length [$t(77) = -3.4$, $P < 0.01$] and rest period length [$t(77) = -2.0$, $P = 0.05$]. This same dataset was also used in a recent large-scale study to create functional probabilistic atlases of face and biological motion perception (for additional details, see Engell and McCarthy, 2013).

fMRI image acquisition

124 adults (66 females, 58 males, mean age 23 years) participated in the face localizer. 106 adults (59 females, 47 males, mean age 24 years) participated in the biological motion localizer. All subjects gave written, informed consent and the Yale Human Investigations Committee approved the protocol. Brain images were acquired at the Magnetic Resonance Research Center at Yale University using a 3.0 T TIM Trio Siemens scanner. Echo-planar images (EPIs) with near whole-brain coverage were acquired from all participants. The precise acquisition parameters varied slightly across participants due to the data being acquired as part of several independent experiments. All study acquisition parameters used the same echo time = 25 ms, flip angle = 90° and matrix size = 64^2. Thirty-four, 36 or 37 slices were acquired with slice thickness = 3.5 or 4 mm, field of view = 224 or 240 and a repetition time = 2 s. Two sets of structural images were collected across all studies to facilitate registration of the EPIs: co-planar T1-Flash images and high-resolution T1-MPRAGE (Magnetization Prepared Rapid Gradient Echo) images.

fMRI preprocessing and analysis

Preprocessing was performed using the FMRIB Software Library (FSL, http://www.fmrib.ox.ac.uk/fsl/). All images were skull-stripped using FSL's brain extraction tool. The first two volumes (4 s) of each functional dataset were discarded to allow for T1-equilibration. Data were temporally realigned to correct for interleaved slice acquisition and spatially realigned to correct for head motion using FSL's linear realignment tool (MCFLIRT). Images were spatially smoothed with a 5 mm full-width-half-maximum isotropic Gaussian kernel. Each time series was high-pass-filtered (0.01 Hz cutoff) to eliminate low-frequency drift. Functional images were registered to co-planar images, which were then registered to high-resolution anatomical images, and normalized to the Montreal Neurological Institute's (MNI) MNI152 template.

Defining regions of interest

An important step in using DCM for fMRI is to select key brain regions involved in the process of interest (Stephan *et al.*, 2010). For the purposes of this study, we selected face-sensitive regions of the right FG and right pSTS and biological motion-sensitive regions of the right FG and right pSTS. We focused on the right, as opposed to left, FG and pSTS given that neuroimaging studies consistently find a right hemisphere bias for face and biological motion processing (Sergent *et al.*, 1992; Kanwisher *et al.*, 1997; McCarthy *et al.*, 1997; Puce and Perrett, 2003; Rossion *et al.*, 2012). These regions were selected in a two-stage process, using both anatomical and functional criteria.

First, anatomical regions of interest (aROIs) were created for the right FG and the right pSTS. The aROIs were defined anatomically on the cortical surface of a standard brain. The FG aROI extended laterally from the collateral to the inferior temporal sulcus and was bounded in the anterior–posterior dimension at MNI *y*-coordinates of approximately −30 and −60. The pSTS aROI included the posterior segment of the superior temporal sulcus ($y \leq -32$) and its descending and ascending limbs. These aROIs served to constrain our search for face-sensitive and biological motion-sensitive ROIs.

Next, we used probabilistic atlases for face and biological motion perception to further constrain our ROIs using functional criteria (Engell and McCarthy, 2013). The probabilistic atlases [created in a previous study (Engell and McCarthy, 2013) using this very same dataset] represent, at each voxel, the percentage of participants who showed a category-sensitive response (defined by a *z*-score of ≥1.65 for the localizer contrast) to either faces or biological motion. The peak voxel of each probabilistic atlas from within the aROIs was identified. Finally, a small sphere (radius 3 mm) was then centered around this peak voxel, resulting in four anatomically and functionally defined ROIs: (i) face-sensitive FG, (ii) face-sensitive pSTS, (iii) biological motion-sensitive FG and (iv) biological motion-sensitive pSTS (Figure 3).

Extracting individual time-series data from ROIs

DCM is a technique that allows one to compare hypotheses about the mechanisms that underlie the regional responses detected in conventional general linear model (GLM) analyses. An important minimal requirement for inclusion of any time-series in a DCM analysis is that it must show an experimental effect (Stephan *et al.*, 2010). As such, we evaluated the response of each participant to faces and biological motion to ensure that only participants exhibiting category-specific responses within our ROIs were included in the analysis.

We adopted an automated approach (described below) to evaluate participant inclusion. This approach was admittedly strict, resulting in excluding more than half of our original sample. It is likely that many of the excluded participants did indeed show category-specific responses to faces and biological motion within the FG and pSTS but that these activations failed to overlap with our strictly defined ROIs. It is possible that manual identification of ROIs in individual subjects, using more subjective and flexible criteria, would have permitted the retention of more subjects in our sample. However, given the large initial sample size, we chose to adopt an automated approach to voxel section that is more rigorous and replicable relative to a more subjective approach.

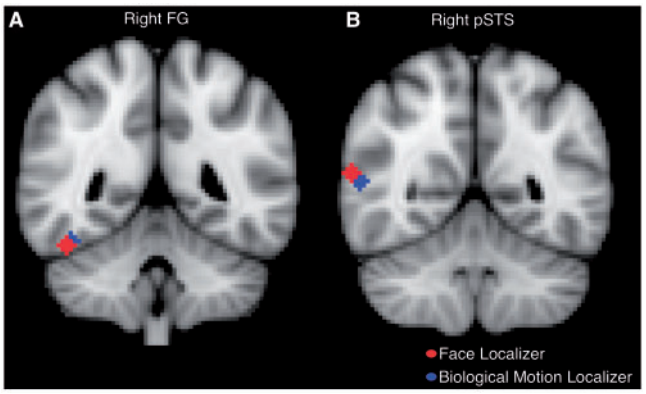

Fig. 3 Spheres defining ROIs for the face and biological motion localizer tasks. **(A)** FG ROIs for the face localizer (in red) and the biological motion localizer (in blue). **(B)** pSTS ROIs for the face localizer (in red) and the biological motion localizer (in blue).

First, whole-brain voxelwise regression analyses were performed using FSL's FEAT program. Each condition (faces and scenes or biological and non-biological motion) was modeled with a boxcar function convolved with a single-gamma hemodynamic response function to create regressors-of-interest for analysis within a GLM. Voxels exhibiting category-selectivity for faces or biological motion were identified using linear contrasts of the model coefficients (i.e. faces > scenes, biological > non-biological). The statistic associated with each contrast was converted to a *z*-score.

Only participants with voxels exhibiting face or biological motion sensitivity (defined by a *z*-score of ≥1.65, uncorrected) in both the FG and pSTS ROIs were selected for inclusion in the final analysis. Forty-four of the 124 adults who participated in the face localizer had face-sensitive voxels within both the FG and pSTS ROIs. Forty-six of the 106 adults who participated in the biological motion localizer had biological motion-sensitive voxels within both the FG and pSTS ROIs. Finally, we visually inspected the activation maps of each individual subject who met inclusion for either the face or biological motion task. Participants were further excluded if their active voxels were determined to be part of a cluster centered outside the bounds of the anatomically defined pSTS and FG (e.g. a small subset of subjects had clusters of activation in occipital cortex that encroached upon the pSTS ROI but had no cluster of activation centered in the anatomically defined pSTS). On this basis, an additional four subjects were excluded from the face task and an additional seven subjects were excluded from the biological motion task. Thus, a total of 40 participants contributed data to the face task and 39 participants for the biological motion task. Time-series data for each participant were extracted by averaging the time-series of all voxels with a *z*-score of ≥1.65 within each ROI.

DCM and BMS

DCM analyses were carried out using DCM 10 in SPM8 (http://www.fil.ion.ucl.ac.uk/spm). BMS was used to select the optimal model in each model space (see Supplementary Data for validation of the model space). Competing models were compared via the calculation of negative free energy (Friston *et al.*, 2007). Negative free energy serves as an approximation of model evidence that balances both accuracy and complexity. We used a fixed effects approach to compare the negative free energy of each model under the assumption that the best fitting model structure would be constant across subjects (Stephan *et al.*, 2010). This procedure yields a posterior probability (the probability that the model generated the data) for each model. A model with a posterior probability of at least 95% is considered to have strong evidence of being the most likely model (Penny *et al.*, 2004).

RESULTS

Face localizer

As predicted, a model in which the FG received information about faces as a direct input and then 'told' the pSTS about this information received the most evidence [where evidence represents $P(y|m)$, the probability of dataset y, given model m]. That is, the model was chosen as the most probable among all possible models. Specifically, Model 2, which specified faces as a direct input to the FG and bidirectional intrinsic connections between the FG and pSTS received the most evidence. It obtained a posterior probability of 100% (Figure 4), a clear winner over all other models. The log-evidence difference, or log Bayes factor, between this best model (Model 2) and the second-best model (Model 1) was 12.49, which is regarded as very strong evidence in favor of Model 2 (Kass and Raftery, 1995). Note that Model 1 is the other model consistent with our prediction.

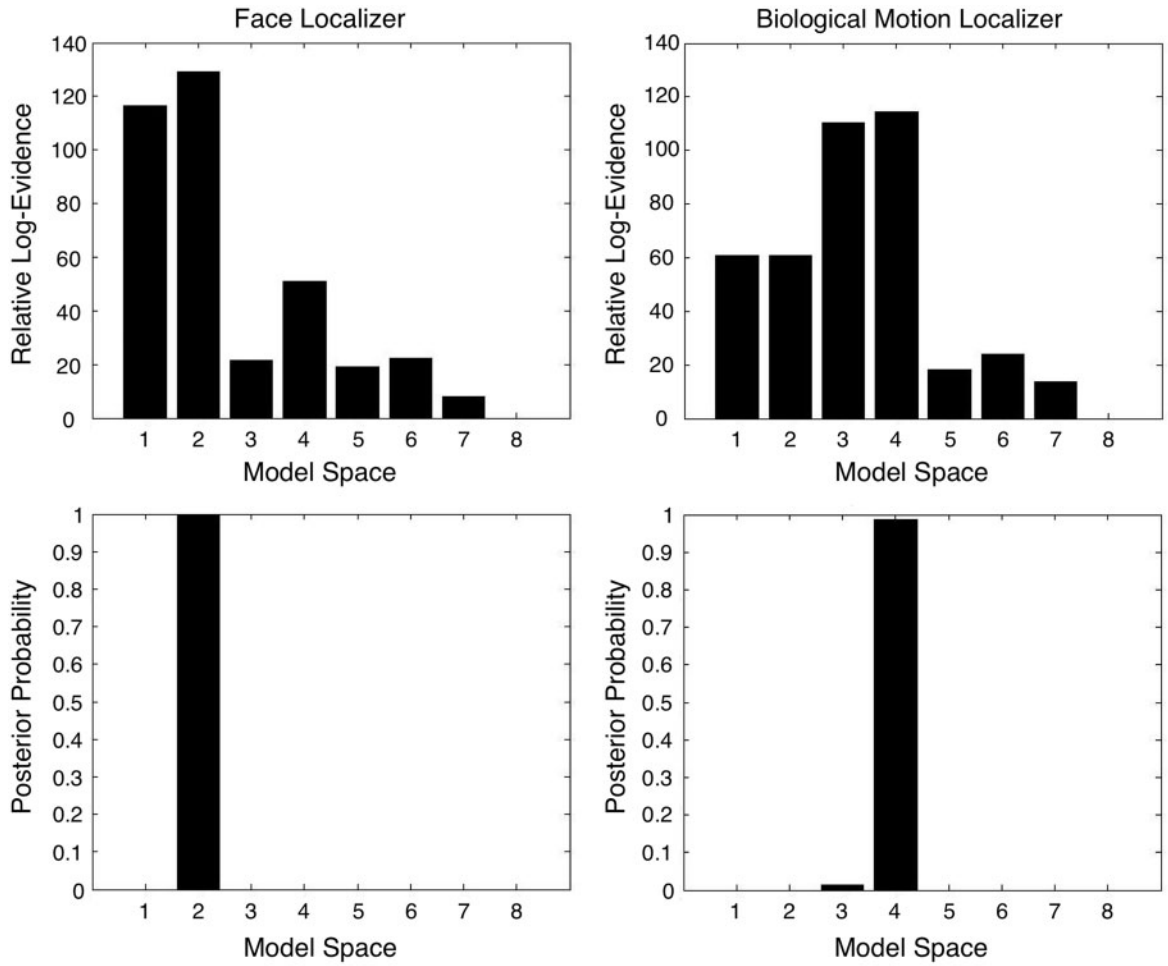

Fig. 4 BMS results for the face localizer (left panel) and biological motion localizer (right panel). For the face localizer, Model 2 received the highest relative-log-evidence and a posterior probability of 100%. For the biological motion localizer, Model 4 received the highest relative-log-evidence and a posterior probability of 99%. One can compare specific pairs of models by judging whether the difference in their log-evidences (i.e., the relative-log-evidence) is greater than conventional thresholds. A difference of 3 = strong support (5 = very strong support) for the model with higher evidence.

Parameters for the winning model were also examined using Bayesian Parameter Averaging (BPA) (Stephan *et al.*, 2010). Across participants, the strength of the direct input of faces to the FG was 0.02 Hz; the strength of the feedforward intrinsic connection between the FG and pSTS was 0.56 Hz; and the strength of the feedback connection from the pSTS to the FG was 0.38 Hz (Figure 5). The posterior probabilities of the parameter estimates were 1, 0.94 and 1, indicating that these values are significantly different from zero.

Biological motion localizer

As predicted, a model in which the pSTS received information about biological motion as a direct input and then 'told' the FG about this information was selected as being most probable. Specifically, Model 4, which specified biological motion as a direct input to the pSTS and bidirectional intrinsic connections between the pSTS and FG received the most evidence. It obtained a posterior probability of 99% (Figure 4), a clear winner over all other models. The log-evidence difference, or log Bayes factor, between this best model (Model 4) and the second-best model (Model 3) was 4.31, which is regarded as strong evidence in favor of Model 4 (Kass and Raftery, 1995). Note that Model 3 is the other model consistent with our prediction.

Parameters for the winning model were also examined using BPA (Stephan *et al.*, 2010). Across participants, the strength of the direct input of biological motion to the pSTS was 0.02 Hz; the strength of the feedforward intrinsic connection between the pSTS and FG was 0.52 Hz; and the strength of the feedback connection from the FG to the pSTS was 0.45 Hz (Figure 5). The posterior probability for each parameter estimate was 1, indicating that all parameter estimates were significantly different from 0.

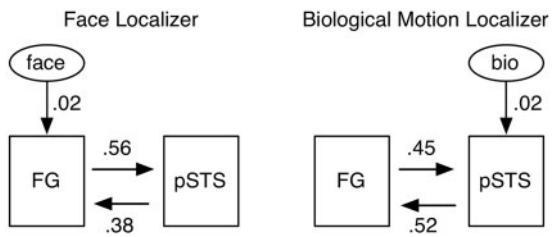

Fig. 5 Models receiving the highest relative-log-evidence for the face localizer (left) and the biological motion localizer (right). The estimated strengths (in Hz) of each intrinsic connection are shown next to the corresponding arrow.

DISCUSSION

Here, we used DCM to investigate the causal flow of information between the FG and the pSTS when viewing faces and biological motion. The 'winning' models suggest that when animacy cues are presented in the form of static faces, activation in the FG drives activation in the pSTS. In contrast, when animacy cues are presented in the form of

biological motion the directional flow of information is reversed—activation in the pSTS drives activation in the FG. These results support our hypothesis that the causal flow of information between the FG and the pSTS depends on the characteristics of the stimulus presented.

These findings are consistent with our proposed model for animacy detection and with existing models of high-level vision, providing additional support for the notion that animacy perception may involve separate but interacting processing streams (Haxby *et al.*, 2000; Hoffman and Haxby, 2000; O'Toole *et al.*, 2002; Beauchamp *et al.*, 2003; Weiner and Grill-Spector, 2011). However, although previous studies have focused primarily on the functional and anatomical distinctions between processing streams, the present results provide the first demonstration of interactions between streams for processing features of animate agents, revealing a potential explanation for why both the FG and pSTS respond to both faces and biological motion. Our results demonstrate that the FG and pSTS are, at least during the earliest stages of processing, differentially sensitive to faces and biological motion—when faces are presented the FG drives activation in the pSTS, whereas when biological motion is presented the pSTS drives activation in the FG. Further, the winning models specify bidirectional intrinsic connections between the FG and the pSTS. Thus, although FG activation in response to faces may initially be driven by receiving faces as a direct input, its activity then both influences and becomes influenced by feedback from the pSTS. Similarly, although pSTS activation in response to biological motion may initially be driven by receiving information about biological motion as a direct input, its activity then influences and becomes influenced by feedback from the FG. These results reveal a temporal dimension to category selectivity: although both regions are initially differentially sensitive to animacy cues, this selectivity is transient as both regions ultimately play a role in processing faces and biological motion.

The notion that functional specialization can be limited by time as well as space has been suggested by studies examining intracranial event-related potential correlates of face processing. These studies demonstrate that faces evoke a characteristic N200 response at discrete loci in the FG, followed by spectral perturbations in the gamma band at these same recording sites (Puce *et al.*, 1999; Engell and McCarthy, 2010, 2011). Importantly, the N200 and the gamma response may reflect separate components of the face processing system. Specifically, the N200 initially responded to faces in a largely obligatory and invariant manner, perhaps reflecting structural encoding of faces, whereas the gamma response seemed to reflect elaborative processing of faces. Thus, although initial operations at these cortical sites may be induced by the structure of a face, other processes may influence later operations at these same sites (Puce *et al.*, 1999; Engell and McCarthy, 2010, 2011).

These data demonstrate that our understanding of the functions of particular brain regions may depend on *when* brain responses are measured. Our results suggest that although FG and pSTS activation is *initially* driven by particular stimulus types (faces and biological motion, respectively), later responses in the FG both influence and become influenced by ongoing processes in the pSTS, and vice versa. The functional role of each region during these later stages of processing is an important area for future research. Rather than focusing exclusively on the initial narrow selectivity of these areas, it will be crucial to consider the larger functional role that they play in the service of an extended network involved in identifying animate agents and their intentions. Achieving a truly mechanistic understanding of the role of each region in this network requires the examination of how the functional role of each region may change as they influence or become influenced by activation in other regions.

Although the current results support a critical hypothesis of our proposed network for animacy detection, many predictions from our model remain untested. First, an important animacy cue, goal-directed actions, was not investigated in this study. Our model posits that the parietal system activates in response to goal-directed actions, which then causes activation in the VOTC and LOTC processing streams (Figure 1). Future studies measuring effective connectivity are required to test this claim.

Second, although this study tested only the pSTS and FG as key nodes for transmitting information between LOTC and VOTC processing streams, it is possible that earlier nodes in our proposed network, such as the middle occipital gyrus (MOG) and inferior occipital gyrus (IOG), may also transmit information between processing streams (Figure 1). If this were the case, the LOTC and VOTC streams may actually receive information about faces and biological motion, respectively, 'prior' to the transmission of this information via intrinsic connection between the FG and pSTS. However, this possibility seems unlikely given that models specifying faces/biological motion as direct inputs to *both* the FG and the pSTS (Figure 2, Models 5–8) were not favored over models specifying faces as a direct input to the FG and biological motion as a direct input to the pSTS, suggesting that the FG and pSTS serve as key nodes for sharing this information across processing streams (but see Fairhall and Ishai, 2007). Nonetheless, this is an important question that requires further investigation by comparing models that include the FG, pSTS, MOG and IOG. Although future studies are clearly needed to test our proposed network in its entirety, we would like to emphasize that the inclusion of additional regions in future DCM studies should not substantially alter the present finding of a reversal of information flow between the FG and pSTS when animacy cues are presented via faces or biological motion, as estimates of effective connectivity from a full network (including all ROIs) are often very consistent with estimates based on a subset of the regions within the full network (Friston, 2011).

Third, this study examined the causal flow of information between processing streams in response to single animacy cues presented in isolation. As mentioned, these cues typically co-occur in the natural world. As such, it would be important to examine how more veridical stimuli containing multiple animacy cues (e.g. a moving face) might affect causal relationships between brain regions in our proposed network. Our model predicts that if both cues were presented in a single stimulus, the VOTC and LOTC processing streams would receive information about the stimulus simultaneously, through direct inputs to both streams. Finally, it is interesting to consider whether manipulating the presence or task-relevance of particular stimulus characteristics might modulate the strength of intrinsic connections between regions. Although it was not the focus of this study, we noted that the strengths of the feedforward intrinsic connections (FG → pSTS for the face localizer and pSTS → FG for the biological motion localizer) were stronger than the corresponding feedback connections. A stimulus containing information about both faces and biological motion may equalize the strengths of the bidirectional intrinsic connections, whereas focusing attention on a particular stimulus characteristic might strengthen the intrinsic connection through which that information is transmitted.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

REFERENCES

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*(2), 231–9.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience*, *15*(7), 991–1001.

Bíró, S., Csibra, G., Gergely, G. (2007). The role of behavioral cues in understanding goal-directed actions in infancy. *Progress in Brain Research*, *164*, 303–22.
Bonda, E., Petrides, M., Ostry, D., Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *The Journal of Neuroscience*, *16*(11), 3737–44.
Carey, S., Spelke, E. (1994). Domain-specific knowledge and conceptual change. In: Hirschfeld, L., Gelman, S., editors. *Mapping the Mind: Domain Specificity in Cognition and Culture* New York: Cambridge University Press, pp. 169–200.
Carey, S., Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, *63*(4), 515–33.
Csibra, G., Bíró, S., Koós, O., Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111–33.
Csibra, G., Gergely, G., Bíró, S., Koós, O., Brockbank, M. (1999). Goal attribution without agency cues: the perception of "pure reason" in infancy. *Cognition*, *72*(3), 237–67.
Engell, A.D., McCarthy, G. (2010). Selective attention modulates face-specific induced gamma oscillations recorded from ventral occipitotemporal cortex. *The Journal of Neuroscience*, *30*(26), 8780–6.
Engell, A.D., McCarthy, G. (2011). The relationship of gamma oscillations and face-specific ERPs recorded subdurally from occipitotemporal cortex. *Cerebral Cortex*, *21*(5), 1213–21.
Engell, A.D., McCarthy, G. (2013). Probabilistic atlases for face and biological motion perception: an analysis of their reliability and overlap. *Neuroimage*, *74*, 140–51.
Fairhall, S.L., Ishai, A. (2007). Effective connectivity within the distributed cortical network for face perception. *Cerebral Cortex*, *17*(10), 2400–6.
Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, *308*(5722), 662–7.
Friston, K. (2011). Dynamic causal modeling and granger causality comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Brain*, *58*, 303–5.
Friston, K.J., Harrison, L., Penny, W. (2003). Dynamic causal modeling. *Neuroimage*, *19*, 1273–302.
Friston, K.J., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage*, *34*(1), 220–34.
Gao, T., Scholl, B.J., McCarthy, G. (2012). Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *The Journal of Neuroscience*, *32*(41), 14276–80.
Gergely, G., Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–92.
Gibson, E.J., Owsley, C.J., Johnston, J. (1978). Perception of invariants by five-month-old infants: differentiation of two types of motion. *Developmental Psychology*, *14*(4), 407–15.
Grossman, E., Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, *35*(6), 1167–75.
Guajardo, J.J., Woodward, A.L. (2004). Is agency skin deep? Surface attributes influence infants' sensitivity to goal-directed action. *Infancy*, *6*(3), 361–84.
Haxby, J., Hoffman, E., Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–33.
Hoffman, E.A., Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–4.
Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–11.
Kass, R.E., Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–95.
Leslie, A. (1994). ToMM, ToBy, and agency: core architecture and domain specificity. In: Hirscfeld, L., Gelman, S., editors. *Mapping the Mind: Domain Specificity in Cognition and Culture* New York: Cambridge University Press, pp. 119–48.
Leslie, A. (1995). A theory of agency. In: Sperber, D., Premack, D., Premack, A., editors. *Causal Cognition: A Multidisciplinary Debate* Oxford: Clarendon Press, pp. 121–41.
McCarthy, G., Puce, A., Gore, J.C., Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–10.
O'Toole, A.J., Roark, D.A., Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*(6), 261–6.
Peelen, M.V., Downing, P.E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, *93*(1), 603–8.
Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J. (2004). Comparing dynamic causal models. *Neuroimage*, *22*, 1157–72.
Pitcher, D., Walsh, V., Yovel, G., Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, *17*(18), 1568–73.
Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, *36*(1), 1–16.
Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, *16*(16), 5205–15.
Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *The Journal of Neuroscience*, *18*(6), 2188–99.
Puce, A., Allison, T., Gore, J.C., McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology*, *74*(3), 1192–9.
Puce, A., Allison, T., McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cerebral Cortex*, *9*(5), 445–58.
Puce, A., Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, *358*(1431), 435–45.
Rossion, B., Hanseeuw, B., Dricot, L. (2012). Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, *79*(2), 138–57.
Scholl, B., Tremoulet, P. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309.
Schwarzlose, R.F., Baker, C.I., Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience*, *25*(47), 11055–9.
Sergent, J., Ohta, S., Macdonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain*, *115*(1), 15–36.
Shultz, S., McCarthy, G. (2012). Goal-directed actions activate the face-sensitive posterior superior temporal sulcus and fusiform gyrus in the absence of human-like perceptual cues. *Cerebral Cortex*, *22*(5), 1098–106.
Simion, F., Regolin, L., Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 809–13.
Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E.M., Daunizeau, J., Friston, K.J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage*, *49*, 3099–109.
Tootell, R.B., Taylor, J.B. (1995). Anatomical evidence for MT and additional cortical visual areas in humans. *Cerebral Cortex*, *5*(1), 39–55.
Turk-Browne, N.B., Norman-Haignere, S.V., McCarthy, G. (2010). Face-specific resting functional connectivity between the fusiform gyrus and posterior superior temporal sulcus. *Frontiers in Human Neuroscience*, *4*, 176.
Weiner, K.S., Grill-Spector, K. (2011). Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97.