Cerebral Cortex, July 2021;31: 3522–3535

doi: 10.1093/cercor/bhab029
Advance Access Publication Date: 25 February 2021
Original Article

ORIGINAL ARTICLE

Top-Down Attention Guidance Shapes Action Encoding in the pSTS

Daniel A. Stehr[1], Xiaojue Zhou[1], Mariel Tisby[1], Patrick T. Hwu[1], John A. Pyles[2] and Emily D. Grossman[1]

[1]Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697, USA and [2]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Address correspondence to Emily D. Grossman, Department of Cognitive Sciences, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA. Email: grossman@uci.edu.

Abstract

The posterior superior temporal sulcus (pSTS) is a brain region characterized by perceptual representations of human body actions that promote the understanding of observed behavior. Increasingly, action observation is recognized as being strongly shaped by the expectations of the observer (Kilner 2011; Koster-Hale and Saxe 2013; Patel et al. 2019). Therefore, to characterize top-down influences on action observation, we evaluated the statistical structure of multivariate activation patterns from the action observation network (AON) while observers attended to the different dimensions of action vignettes (the action kinematics, goal, or identity of avatars jumping or crouching). Decoding accuracy varied as a function of attention instruction in the right pSTS and left inferior frontal cortex (IFC), with the right pSTS classifying actions most accurately when observers attended to the action kinematics and the left IFC classifying most accurately when observed attended to the actor's goal. Functional connectivity also increased between the right pSTS and right IFC when observers attended to the actions portrayed in the vignettes. Our findings are evidence that the attentive state of the viewer modulates sensory representations in the pSTS, consistent with proposals that the pSTS occupies an interstitial zone mediating top-down context and bottom-up perceptual cues during action observation.

Key words: action observation, action observation network, attention, multivariate pattern analysis, posterior superior temporal sulcus

Introduction

The posterior superior temporal sulcus (pSTS) is linked to the perceptual representations of body actions during action observation. Classically, the pSTS is characterized as providing the key sensory input needed to facilitate the interpretation of goals from motor behavior and ascribe intentions in social interactions (Thompson and Parasuraman 2012; Pyles and Grossman 2013; Lingnau and Downing 2015). This strictly perceptual characterization of the pSTS, however, fails to account for the influence of high-level contextual factors on the neural response. Activation in the pSTS is modulated by recent history of the observed action events (Vangeneugden et al. 2011), whether the viewer is attending to the social dimensions of an event (Tavares et al. 2008), whether the observed action is consistent with the expectation of the viewer (Wyk et al. 2009; Jastorff et al. 2011; Saygin et al. 2012; Maffei et al. 2015; Urgen and Saygin 2020), and whether the action is construed as intentional or incidental (Morris et al. 2008).

Contemporary theories of the action observation network (AON) now emphasize an integrative role of the pSTS rather than one of strict sensory encoding. In these proposals, specific action features, such as body postures and local kinematics, are encoded in the lateral occipitotemporal cortex (LOTC) regions and subsequently bound into action representations in the pSTS (Giese and Poggio 2003). The action representations are further tuned by modulatory signals that reflect top-down influences imposed by cognitively derived internal models (Geng and Vossel 2013; Sokolov et al. 2018). These modulatory influences are constructed in higher levels of the AON (i.e., the inferior frontal cortex; IFC) and are proposed to shape action representations so as to facilitate the behavioral goals of the viewer (Carter and Huettel 2013; Patel et al. 2019). In one special class of these models, predictive coding models, top-down signals bias perceptual encoding in favor of expected actions as determined from prior knowledge of action goals, increasing the efficiency of perceptual encoding of the subsequently observed action (Kilner 2011; Koster-Hale and Saxe 2013; Bach and Schenke 2017).

An important innovation in this new class of theoretical models is the specialized role of the pSTS as the integrator of two information streams: bottom-up sensory encoding of observed actions and top-down cognitively derived context. Unlike strictly representational accounts, integrative models are highly flexible in that they emphasize the encoding of sensory cues dependent on the observer's cognitive state. This integrative role, therefore, provides a new framework by which the local functional heterogeneity of the pSTS can be interpreted (Patel et al. 2019), namely that sensory information may be represented uniquely depending on the attentive goals of the viewer. This stands in contrast to proposals that characterize the pSTS as host to distinct neural populations for low-level perceptual and high-level social cognitive functions, intermixed and distributed throughout LOTC (Hein and Knight 2008; Bahnemann et al. 2009; Deen et al. 2015).

Integrative and predictive coding models are both influential in understanding brain systems that underlie action observation and moreover are supported by univariate mapping studies showing that the behavioral goals of the observer alter activation maps distributed along the superior temporal sulcus. What is currently lacking, however, is direct evidence that the behavioral goals of the observer alter action representations that are constructed during action observation. One proposed mechanism by which this may occur is the sharpening of neural tuning to the attended actions, akin to the attention-mediated gain increases observed in early visual cortex during feature-based attention tasks (Treue and Martínez Trujillo 1999; Saenz et al. 2002; Kok et al. 2012). Feature-based attention gain is a mechanism consistent with all classes of top-down integrative models and has been observed widely throughout sensory systems (Maunsell and Treue 2006). Alternatively, observer goals have the potential to alter behavior without restructuring action representations directly. This could be achieved through the introduction of bias in the decision-making process, which would manifest in later stages of cortical processing while leaving action representations unadulterated (i.e., Summerfield and Egner 2009).

A further consideration is the level of abstraction of the top-down influences that may shape perceptual representations. This is particularly important during action observation, in which a specific goal can be achieved through various combinations of an individual's actions, while specific actions may not be diagnostic of an individual's current goals or intentions (Thompson et al. 2019). Thus, expectations of upcoming actions could include anticipated kinematic events, action outcomes, or perhaps even abstracted representations of action goals (Kilner 2011).

In this study, we investigate how feature-based attention modifies the statistical structure of action representations embedded within the spatial activation patterns elicited during action observation. We test the hypothesis that directing attention to kinematic aspects of an action vignette sharpens the tuning of these representations and compare it to when attention is directed to features not associated with action recognition (namely, the identity of the actor, which is associated with the LOTC and inferior temporal cortex but not the pSTS; Lingnau and Downing 2015). We also evaluate the efficacy by which directing attention to observer goals (rather than specific actions) facilitates the decoding of action representations. We evaluate this hypothesis in three regions of the AON: the pSTS, the form and motion-selective LOTC (Oosterhof et al. 2010; Wurm and Lingnau 2015), and the IFC (Ogawa and Inui 2011; Wurm and Lingnau 2015). In a second analysis, we compare connectivity strength within the AON as a function of observer attention state to evaluate if the changes in information are likewise associated with selective strengthening of information through key pathways. Our results are consistent with models of the pSTS as dynamically restructuring action representations depending on the viewer's attentive state, with actions most strongly differentiated when observers attend to the kinematic content. These results are consistent with top-down and predictive coding models that emphasize the role of prior knowledge in shaping action representations.

Methods

Participants

Twenty-five healthy adults (8 male, 17 female) ranging in age from 21 to 42 years old (mean = 24.7, SD = 3.6) from the UC Irvine campus and surrounding community enrolled in and completed the study. Participants gave written informed consent. All experimental procedures were approved by the University of California Irvine Institutional Review Board. All participants had normal or corrected-to-normal vision. One participant was excluded from the analysis due to excessive motion during scanning.

MR Image Acquisition

Participants were scanned at the Facility for Imaging and Brain Research at the University of California, Irvine, on a 3 Tesla Siemens Prisma MRI scanner (Siemens Medical Solutions) equipped with a 32-channel receive-only phased array head coil. High-resolution anatomical images were collected using a single T1-weighted magnetization–prepared rapid acquisition gradient echo (MPRAGE) sequence (176 sagittal slices; 1 mm isovoxel resolution; field of view = 256 mm; TR = 2000 ms; TE = 1.99 ms; TI = 900 ms: flip angle = 9 degrees; GRAPPA acceleration factor = 2; bandwidth = 240 Hz/Px).

Two types of functional scans were acquired across two sessions, both using a T2*-weighted gradient recalled echoplanar imaging multiband pulse sequence (cmrrmbep2dbold) from the University of Minnesota Center for Magnetic Resonance Research (CMRR). Session 1 consisted of localizer scans designed to identity regions of interest (ROIs) within the AON (69 slices coplanar with the AC/PC; in-plane resolution = 2 2 mm; 106 106 matrix size; 2 mm slice thickness, no gap; interleaved acquisition; field of view = 212 mm; phase partial Fourier scheme of 6/8; TR = 2000 ms; TE = 30 ms; flip angle = 79 degrees; bandwidth = 1814 Hz/Px; echo spacing = 0.66 ms; excite pulse duration = 8200 microseconds; multiband factor = 3; phase-encoding direction = PA; fat saturation on; advanced shim mode on). Session 2 comprised the main experiment and therefore incorporated rapid event–related scans that were designed to sample the hemodynamic response more rapidly (68 slices coplanar with the AC/PC; in-plane resolution = 2 2 mm; 106 106 matrix size; 2 mm slice thickness, no gap; interleaved acquisition; field of view = 212 mm; phase partial Fourier scheme of 6/8; TR = 1500 ms; TE = 30 ms; flip angle = 79 degrees;

bandwidth = 2144 Hz/Px; echo spacing = 0.57 ms; excite pulse duration = 8200 microseconds; multiband factor = 4; phase-encoding direction = PA; fat saturation on; advanced shim mode on). At the beginning of each session, an additional pair of EPI images with phase-encoding directions of opposite polarity in the anterior-to-posterior plane was acquired to correct for susceptibility distortions in each participant's functional data.

Session 1: Functional Localizers

In the first session, all participants underwent three functional localizer scans (two repetitions each) to identify the pSTS, middle temporal complex (hMT+), and extrastriate body area (EBA). Stimuli were displayed on a BOLDScreen32 LCD monitor controlled by MATLAB (Mathworks, Inc.) and the Psychophysics Toolbox extensions (Brainard 1997) on a Windows desktop. Subjects viewed the animations through a mirror mounted on the head coil and directed at a screen positioned at the head end of the scanner.

Posterior superior temporal sulcus. To localize areas of the brain that respond selectively to biological motion, participants were shown 12 alternating blocks of intact and scrambled point-light biological motion (Grossman et al. 2010). Animations depicted an actor with 12 lights attached to the joints performing 25 unique actions, such as walking, jogging, throwing, kicking, etc. Scrambled animations were produced by randomizing the starting position of the point-light dots within a region approximating the target figure and then leaving their motion vectors intact. Animations had a duration of 1 s and were separated by a 1-s fixation intertrial interval (ITI). Participants performed a 1-back task on each animation, indicating by button press whether the current animation was the same or different action as the one immediately prior. The pSTS was identified using a group random-effects GLM that contrasted intact versus scrambled trials, thresholded using a false discovery rate (FDR; Genovese et al. 2002) of $q < 0.005$.

Lateral occipitotemporal cortex. The LOTC was identified jointly using two localizers, one targeting hMT+ and the other targeting the EBA. Although separable in individual subjects (Weiner and Grill-Spector 2011, 2013), the hMT+ and EBA in group analyses jointly occupy the ascending limb of the posterior inferotemporal sulcus (pITS; Downing et al. 2007). The LOTC was therefore identified as the union of the hMT+ and EBA (described below), constrained to the dorsal extent by the inferior temporal gyrus.

To isolate the motion-selective hMT+, participants passively viewed alternating 12-s blocks of optic flow dot motion and stationary dot patterns (Huk et al. 2002). Optic flow was constructed with 500 black dots randomly dispersed within a circular aperture, alternating between expansion and contraction. In the stationary interval, dots remained frozen in position for 12 s. The motion-selective responses on the pITS were thresholded using FDR, $q < 0.005$.

To isolate the body-selective EBA, participants viewed images of headless bodies, cars, and limbs (hands and feet) (Stigliani et al. 2015). Each image was superimposed on top of a 10.5-degree phase-scrambled background generated from a randomly selected image to minimize low-level differences across categories. Images were presented in 12 blocks, with 9 images shown per block. Body and limb selective brain regions were identified as those with higher brain response when viewing bodies and limbs versus images of cars, FDR, $q < 0.005$.

Session 2: Action Observation

Action vignettes spanning 3 s (see Fig. 1A) were generated in Poser Pro 11 (2015) and depicted one of two avatars (a boy or a man) performing the same sequence of actions in which the avatar walked toward a bookshelf, indicated intent to reach one of two boxes, and then either crouched down or jumped up to reach the box. The vignette ended after the execution of the action and prior to the avatar making contact with the box. Each vignette was constructed such that it was visualized from eight unique viewpoints that spanned an 80-degree viewing range on each side (left and right, profile to rear views). Two different avatars (a man and a boy) were selected such that participants could discriminate identity based on body type without directed attention to facial features, and so that attention would be directed to articulating body features in all three attention conditions (including the attention to identity condition). The man and boy differed in height and therefore gait stride and traversed the same distance as they approached the target objects. In order to synchronize the timing of key events in the action vignette (arrival to the shelf, initiation of the action), the boy walked with a slightly faster gait relative to the man.

Before beginning the experiment in the scanner, all participants were familiarized with the action vignettes and practiced the task under all three attention conditions: attending to the actor's identity, the action category, or the proximate goal of the action. To prevent motor response preparation while viewing the action vignette, stimulus–response mappings were obscured until the response interval, during which the labels for the three binary dimensions of the action (identity: boy/man; action: crouch/jump; goal: low box/high box) were randomly assigned to the left and right sides of fixation on each trial. Participants reported the correct label by pressing the button corresponding to the side of the screen correctly displaying the value of the feature they attended. Classification was always conducted on the trials with the action labels jumping versus crouching.

Trials were separated by a 3-, 4.5-, or 6-s ITI, pseudorandomized within each run such that, in total, each trial lasted 10.5, 12, or 13.5 s. The onset of each trial event was synchronized with the onset of volume acquisition to ensure synchronization with the event-related acquisition. Each run of the experiment contained eight conditions per attention task from a fully crossed design comprising two avatars (boy and man), two actions (crouching and jumping) and two viewpoints (leftward and rightward walking). The three attention tasks (attend to avatar identity, action category, and action goal) were randomly interleaved within each run, resulting in a total of 24 trials per run or approximately 5 min of scan time. The experiment was organized into eight runs for a total of 192 trials.

Imaging Analysis

Preprocessing

Preprocessing of functional imaging data was conducted using BrainVoyager QX v20.6 (Goebel et al. 2006). All functional images were slice-time corrected, motion-corrected within and between runs, linearly detrended, and temporally high-pass filtered (cutoff frequency 0.01 Hz). Session 2 scans were additionally corrected for susceptibility-induced magnetic field distortions using the field map method (Jezzard and Balaban 1995), implemented in BrainVoyager's COPE v1.0 plugin. All functional images were coregistered to each individual's T1-weighted image.

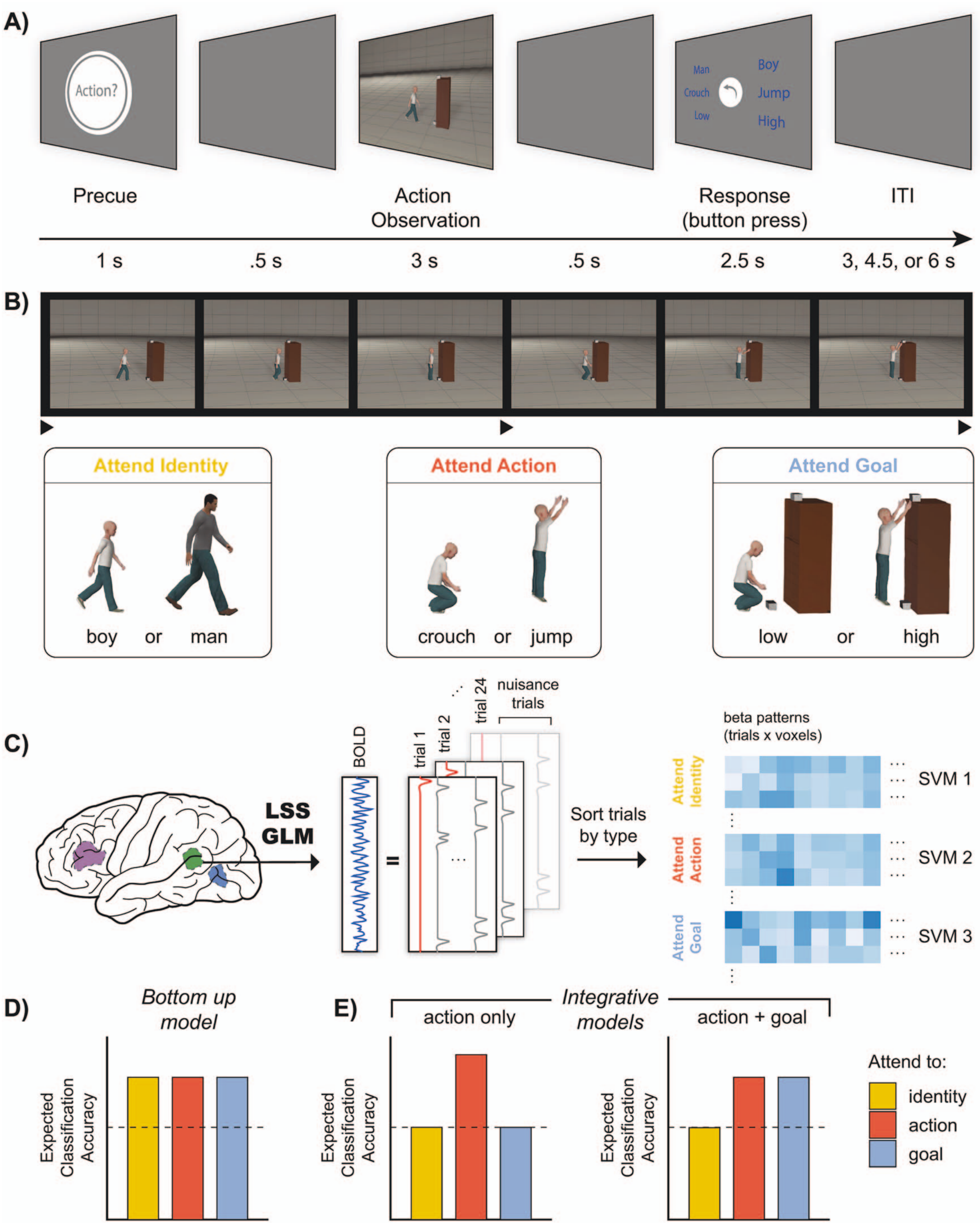

Figure 1. (*A*) Timing of trial events in the rapid event–related design. Precue specified the dimension to be attended (action, goal, or identity). The response cue psuedorandomized judgment–response mappings to prevent the participant from engaging in motor planning during the action observation interval. (*B*) Filmstrip view of stills from the action vignette showing detailed views of an avatar jumping with the intention to reach the box on top of the bookshelf. (*C*) Regions of interest were identified using independent localizer scans (pSTS and LOTC) and surface-based atlas segmentation (IFC). The BOLD response to each event was estimated by iteratively fitting a linear model that included a separate regressor for each trial and confound regressors for all other trials grouped by type (LSS approach). The resulting matrix of beta coefficients was separated by trial type (attend to action, goal, or identity) and used to train and test three separate SVM classifiers per participant. (*D*) Expected patterns of MVPA results for action classification based on models proposing distinct sensory representations of actions unaltered by the attentive state of the viewer, and (*E*) integrative models that are tuned to action kinematics (left) and more broadly to actions and the corresponding goals (right).

Session 1: Regions of Interest

Functional data in session 1 were aligned to a template pilot subject using cortex-based alignment (Frost and Goebel 2012). Sulcal curvature was constructed on white matter surfaces derived from FreeSurfer's recon-all algorithm (http://surfer.nmr.mgh.harvard.edu/), imported into BrainVoyager using custom library functions (https://github.com/tarrlab/Freesurfer-to-BrainVoyager). Regions of interest were identified on the cortical surface and then projected back into native volumetric coordinates by searching along the vertex normal 1 mm towards the white matter and 3 mm towards the gray matter.

The IFC (comprising BA44, BA45A, BA45B, BA47, and the inferior frontal sulcus) was identified anatomically in each individuals' native anatomical images using FreeSurfer's cortical surface atlas mapping algorithms in conjunction with the 400 atom resolution Schaefer atlas (Schaefer et al. 2018). This atlas emphasizes homogeneity of functional systems within the parcels, coupled with high-resolution "atomic" parcellation in approximately equisized units, and therefore higher precision in identifying ROI boundaries.

Session 2: Action Observation under Attentional Instructions

The time series from each voxel in the ROIs was first *z*-scored across time, and trial-by-trial patterns of estimated blood oxygen level–dependent (BOLD) activation were derived using the least squares separate (LSS) general linear model approach (Mumford et al. 2012; Turner et al. 2012). The LSS procedure uses a separate GLM to estimate the pattern of activity for each trial where the model for the ith trial is

$$Y = X_{\mathrm{LSSi}}\beta_{\mathrm{LSSi}} + \varepsilon i$$

such that the design matrix for the ith trial, X_{LSSi}, contains one regressor of interest modeling the stimulus-evoked BOLD response to the ith trial and several other nuisance regressors modeling responses to the remaining trials grouped by trial type. Stimulus-evoked BOLD responses to each event were modeled as 4.5-s boxcar functions that included both the precue and the action vignette, which were separated by a brief (0.5 s) fixed interstimulus interval. This conservative approach was implemented to account for the variability in temporal integration windows across the regions of interest, which are known to be quite extended in the pSTS as compared with the LOTC (Hasson et al. 2008). The boxcar functions were then convolved with a canonical double-gamma hemodynamic response function (HRF) (Friston et al. 1998; Glover 1999). In order to account for variability in the latency of the HRF across the brain and across subjects (Steffener et al. 2010), we optimized the time-to-response-peak parameter of the two-gamma function (5 possible values between 5 and 7 s in steps of 0.5 s), with the modeled HRF that produced the highest coefficient of determination (R^2) for all trials within the voxel selected for downstream analysis. Our LSS design matrix contained six nuisance regressors, one for each action condition (crouching and jumping) crossed with each of the three attention tasks (attend to identity, action, and goal) and additional nuisance regressors capturing the average signal and first derivative measured from the white matter and ventricles over time. Following beta extraction, trials with extreme movement near the peak response (three or more consecutive timepoints of framewise displacement above 2 mm) were censored from later analysis. Also, variance in the beta series accounted by repetitions of actions was removed.

Multivariate Pattern Analysis

Trial betas were separated by attention task (attend to identity, action, and goal) for each participant, and mean centered within runs to remove spurious correlations between the estimated activity levels of different trial types across runs (Lee and Kable 2018). The resulting normalized betas were then averaged within runs to a single activation estimate per action class. The matrix of *n* activation estimates by *k* voxels and 1x*n* class labels was then used to train three separate support vector machine (SVM) classifiers, one per attention task, implemented in the e1072 package in R (Meyer et al. 2018). The SVM consisted of a linear kernel and a cost value of 1. Classification was performed within subjects using 8-fold leave-one-run out cross validation. Within each fold, two predictions were made from the held-out test set, one from each action class. The final classification accuracy for each subject was computed as the mean accuracy across all 8-folds.

To examine task-related differences in multivariate pattern analysis (MVPA) classification accuracy, we constructed a linear mixed-effects model (LMM) using the lme4 package implemented in R (Bates et al. 2019). The LMM predicted classification accuracy based on the fixed effects of attention task, ROI, and their two-way interaction, with participants as random effects. *P* values were obtained using likelihood ratio tests comparing each model to reduced models lacking the variable (or interaction) in question.

Because classification scores are often discrete and may not be normally distributed, statistical significance was further quantified using a randomization procedure at the group level (Stelzer et al. 2013). In these tests, condition labels were exhaustively permuted within the individual participants, and classifiers were trained and tested using the same procedures as above. This procedure yields the expected distribution of classification accuracy (for each participant) under the null hypothesis. Significance ($P < 0.05$, one-tailed) was ascertained from group-level null distributions, constructed using a bootstrap procedure in which a single sample was drawn from each participant's null distribution (iterated 1000 times and sampled with replacement). This procedure was implemented for the planned comparisons of attention conditions (permutations on task labels): attending to action versus identity, and goal versus identity. In a second set of permutations, this procedure was implemented to evaluate whether classification accuracy deviated significantly from chance in each of the attention conditions (permutations of action labels; chance = 50% classification accuracy).

Functional Connectivity

Functional connectivity was computed as the Pearson correlation of the beta series between two ROIs, separately for each attention task. Beta series connectivity is based on the assumption that if two brain regions are functionally interacting, then the amount of activity captured by beta estimates should correlate across trials (Rissman et al. 2004). Beta series correlations were calculated from ROI-averaged time series in which volumes with FDR greater than .4 mm were excluded. As for the MVPA analysis, trialwise betas were estimated using LSS GLMs with nuisance regressors including the global signal measured from the white matter and ventricles and the Volterra expansion of all six rigid body motion realignment parameters (Fristen et al. 1996). Pearson correlations were computed between each 64 beta time series, Fisher r-z transformed. Paired, one-tailed

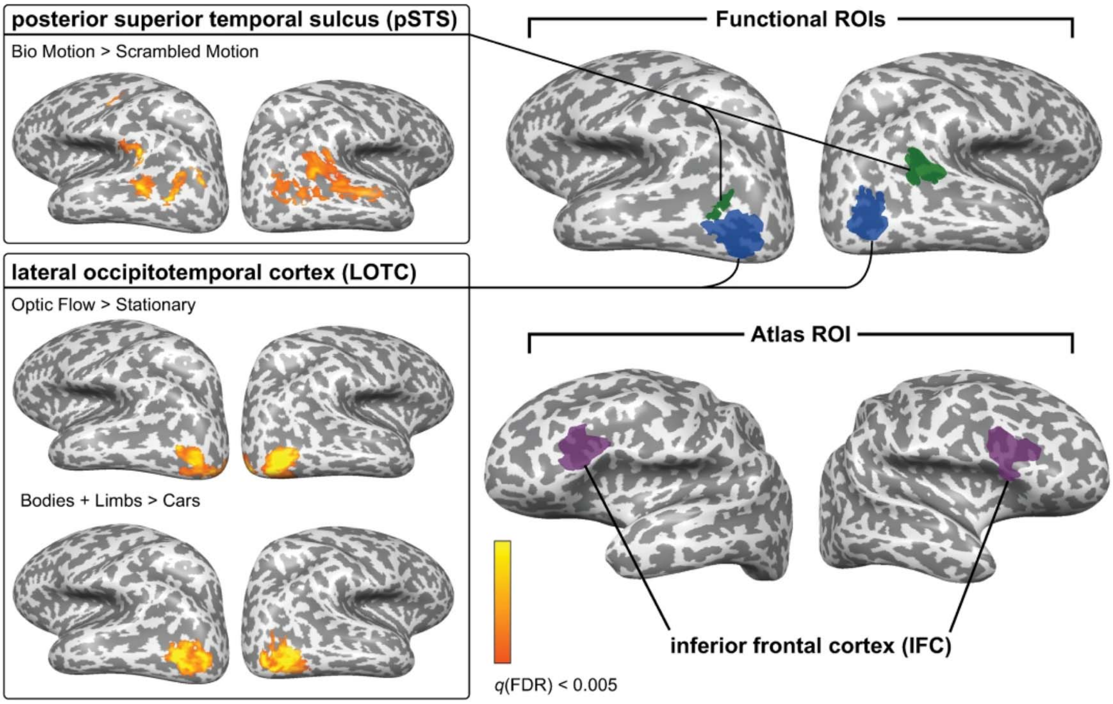

Figure 2. Identification of regions of interest. Left: Group activation maps from the three independent functional localizer scans, displayed on inflated cortical surface meshes of a pilot subject. Right: The regions of interest, including the atlas-derived IFC, projected onto a single subject cortical surface.

repeated measures *t*-tests for the planned contrasts of action > identity and goal > identity were conducted on the transformed correlations.

Results

Functional Localizer Analysis

Results from the independent localizer scans are shown in Figure 2. The biological motion localizer identified large bilateral regions of the pSTS, notably of larger extent in the right hemisphere, whereas the hMT+ and EBA localizers jointly revealed large bilateral coactivation in ventral temporal cortex and LOTC. The spatial overlap between hMT+ and EBA is consistent with reports of functionally distinct neural populations that colocalize to the inferior occipital sulcus when identified in group-based localizers (Downing et al. 2007).

Multivariate Pattern Analysis

To test the hypothesis that the attentional state of the participant sharpens the population tuning of the multivariate informational content during action observation, we evaluated the cross-validated accuracy of action classification (labels: jumping and crouching) from ROI activation patterns (Fig. 3). An LMM with mean classification accuracy as the dependent variable yielded significant main effects of ROI ($\chi^2(5) = 72.83$, $P < 0.001$) and task ($\chi^2(2) = 9.69$, $P < 0.007$), and a significant ROI × task interaction ($\chi^2(10) = 21.81$, $P = 0.016$). Thus, attentional demands influenced the decodability of actions in a subset of ROIs.

To better break down the task x ROI interaction, six within-ROI LMMs were constructed evaluating the influence of the attention instruction on classification accuracies. Planned contrasts compared mean classification accuracies during action and goal attention conditions to the identity attention condition, which served as a baseline in which attention was directed to invariant features (identity) rather than to the dynamic kinematic features. Parameter estimates for each model are presented in Supplemental Table S1. Action decoding in the right pSTS was significantly more accurate when participants attended to the action kinematics versus the identity of the avatar ($b = 0.172$, $SE = 0.062$, $t(48) = 2.786$, $P = 0.008$, uncorrected), consistent with the sharpening hypothesis of the action-tuned neural populations. Permutation tests confirmed that the increase in classification accuracy in the attend-to-action condition exceeded levels expected by chance, with differences greater than observed in 95% of null models (Supplemental Fig. S1).

Action decoding in the right pSTS did not, however, differ significantly between trials when the participant attended to the goal of the action versus the identity of the actor ($b = -0.022$, $SE = 0.060$, $t(48) = -0.370$, $P = 0.713$). In contrast, action decoding in the left IFC increased when participants attended to goal as compared with attending to identity, reaching marginal significance in the LME model ($b = 0.13$, $SE = 0.06$, $t(72) = 2.11$, $P = 0.04$, uncorrected) and exceeding 95% of null models in the nonparametric permutation tests (Supplemental Fig. S2).

No other ROIs revealed significant task-related differences in action decoding.

In all conditions, the trials were labeled according to the portrayed action and the goal of the actor, which, in these vignettes, were strictly confounded (i.e., the actor always gazed upward prior to jumping up and gazed downward prior to crouching down). Thus, classification of the two actions in the attend-to-action and attend-to-goal conditions reflects the same stimulus events. We therefore attribute the observed variations in

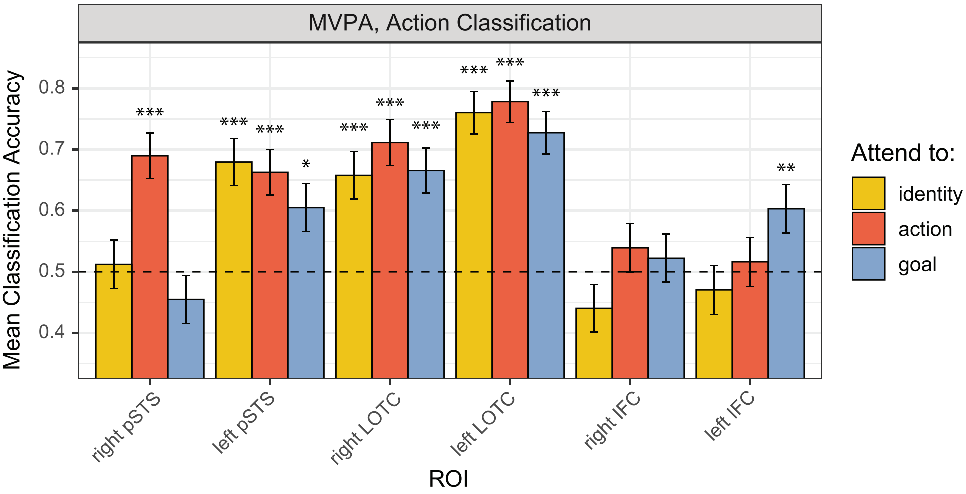

Figure 3. MVPA classification accuracies from decoding action class (jumping and crouching) by task demand. Error bars indicate SEM. Asterisks indicate statistical significance derived from nonparametric permutation tests that evaluated the classification accuracies within each ROI and attention condition expected from chance ($^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$). Dashed line indicates chance level binary classification accuracy (50%).

classification performance across conditions to reflect changes in the cortical state of the observer, rather than to any perceptual differences between the conditions.

To evaluate the specificity of directed attention on decoding accuracy for observed action, we computed classification accuracy for the same trials labeled by actor identity (man vs. boy), which were crossed evenly with the two actions and three attention conditions. Six within-ROI LMMs for each of the ROIs within the AON found no evidence that classification accuracy when coding for identity was modulated by the attentive state of the observer (all $Ps > 0.34$; Supplementary Fig. S3; Supplementary Table S2).

Univariate Analysis

It could be argued that variations in multivariate decodability of actions as a function of top-down instruction may reflect differences in univariate activation levels across tasks, rather than sharpened neural tuning per se. We therefore compared the univariate responses in the ROIs as a function of task (Fig. 4).

Statistical analysis of the average stimulus-evoked responses revealed a main effect of ROI ($\chi^2(5)=496.96$, $P<0.001$), but no main effect of task instruction ($\chi^2(2)=1.26$, $P=0.533$), nor an interaction between task and ROI ($\chi^2(10)=496.96$, $P=0.992$). Thus, the more diagnostic activation patterns in the pSTS when attending to action kinematics versus actor identity cannot be attributed to attentionally driven increases in the average BOLD response within the region. That multivariate classification accuracy is independent of BOLD activation levels is consistent with previous reports that classification is just as high for nonpreferred categories of visual stimuli as it is for preferred categories, within the same brain region (Haxby et al. 2001).

Task Instructions as a Means to Modulate Attention

In this experimental design, observers were instructed to attend to particular dimensions of an action vignette without knowing in advance which action was upcoming. One could hypothesize that the failure to modulate classification accuracy in the AON more broadly (outside of the right pSTS) may reflect a failure of task instructions to guide observer behavior and therefore to alter brain state.

To evaluate this, we analyzed behavioral performance in the scanner, in which participants were required to properly identify the label for the action, action goal, or identity of the actor on each trial. Participants were highly accurate in detecting the features of the action vignettes that they were cued to attend (see Fig. 5). A linear mixed-effects logistic regression model predicting the binary outcome of each trial (“correct” vs. “incorrect”) for each task revealed a trend, but no significant effect of task on accuracy ($\chi^2(2)=5.470$, $P=0.065$). An LMM on response latencies, however, yielded a significant main effect of task ($\chi^2(2)=12.068$, $P=0.002$) such that response latencies were longer when participants identified the goal of the action versus the identity of the actor ($b=92.74$, $SE=23.85$, $t(22.12)=3.888$, $P<0.001$), but not when they identified the action category compared with the identity of the actor ($b=34.86$, $SE=21.09$, $t(23.18)=1.653$, $P=0.112$).

In a second analysis, we evaluated whether activation patterns in the AON regions contained information as to the unique attention states of the observer in the three attention conditions, without regard for the action being observed. We trained a single classifier per ROI to perform a three-way classification with classification labels specifying task instruction (attend to action, goal, or identity) rather than action observed (jump or crouch). Figure 6 displays mean classification accuracies for decoding task. All ROIs classified the task demand more accurately than what would be expected by chance (randomized permutation tests; right pSTS, $P=0.01$; left pSTS, $P<0.001$; right LOTC, $P=0.009$; left LOTC, $P=0.007$; right IFC, $P=0.02$; left IFC, $P=0.004$), evidence that participants differentially allocated their attention according to the task instruction, which in turn altered the informational content in each ROI.

Functional Connectivity

Theoretical models propose the inferior frontal cortex (IFC) to function as a biasing agent such that the sensory representations of specific body kinematics are consistent with the observer’s current behavioral goals (Kilner 2011; Koster-Hale and Saxe 2013). We therefore hypothesized that the signature of such feedback may be reflected through increased functional

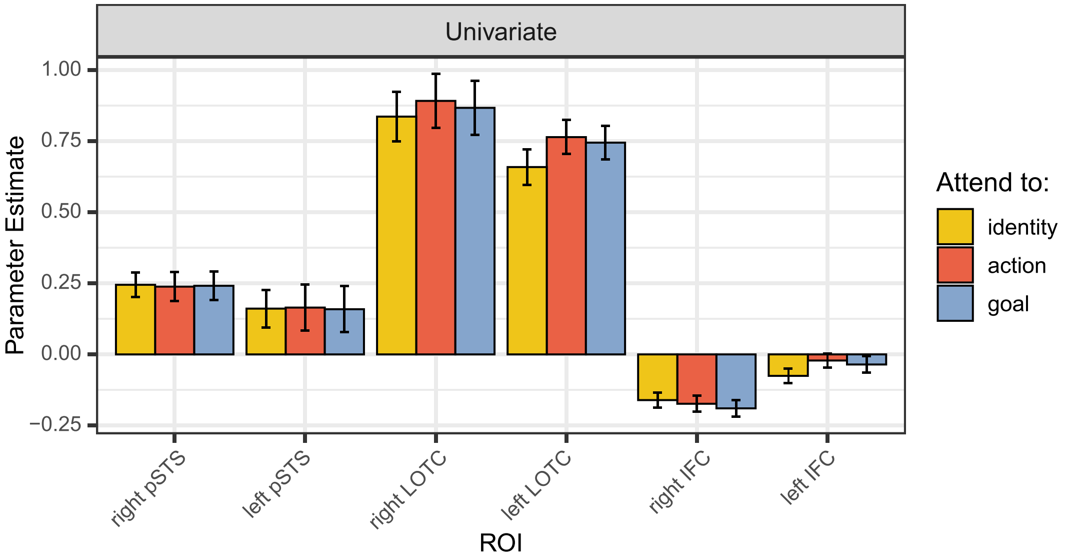

Figure 4. Group univariate responses by task demand modeled during the precue and action observation periods. Univariate activity estimates were produced by averaging the trial-by-trial LSS beta coefficients across trials of each task demand and then averaging the data across voxels within the ROI.

connectivity between the pSTS and IFC during experimental conditions when participants directed their attention to action features.

Analysis of the Pearson's coefficients on the beta time series revealed strong functional connectivity between the right pSTS and the right IFC during action observation under all task instructions (Fig. 7A). When compared across tasks, only the connection between the right pSTS and right IFC varied as a function of task such that it increased significantly when participants attended to action versus when they attended to the actors' identity ($t = 2.21$, $P = 0.018$, uncorrected) (Fig. 7B).

Discussion

The pSTS is increasingly recognized as an integrative hub for decoding social cues that convey essential information for making inferences about actions and intentions (Dasgupta et al. 2016; Sokolov et al. 2018). Contemporary theories propose that the action representations in pSTS are modulated by the attentive state of the observer, and thus identical actions may result in unique representations when viewed under different task goals (Patel et al. 2019). In this study, we test the hypothesis that directed attention to action features sharpens the tuning of neural populations in the pSTS for subsequently viewed actions, reflecting the strengthening of top-down influences acting upon the pSTS.

We found that the attentive state of the observer alters the population code in the right pSTS and the left IFC when viewing action vignettes, as demonstrated by a significant effect of the attention instruction on MVPA accuracy. Specifically, the spatial activation patterns for two distinct actions—jumping and crouching—are more easily differentiated in the right pSTS when observers direct their attention to the kinematic features of the vignette. The spatial activation patterns for the two observed actions are more easily differentiated in the left IFC when observers attended to the implied goals of the actors. The significant improvement in both these ROIs is higher than expected by chance and higher than decodability of the same actions when observers reported the actors' identities. This finding is consistent with models of the AON that propose that pSTS and IFC work together to ascertain implied goals from body kinematics (Hamilton and Grafton, 2008; Kilner 2011), whereas body postures (which specified identity in this experiment) are associated with encoding in the more posterior LOTC and ventral fusiform body area (FBA; Peelen and Downing 2007).

Our findings add to the handful of reports in other sensory domains in which directed feature-based attention refines information in the population response, resulting in more distinct activation patterns that facilitate classification (Kok et al. 2012; Braunlich and Love 2019). A likely mechanism of this attention benefit is the known increased gain in neurons with underlying tuning preferences for the attended features, resulting in an overall sharpening of the population response (for review, see Reynolds and Heeger 2009). In fMRI activation patterns, this has the consequence of warping the representational distinctiveness of the attended items and, when analyzing for information within distributed activation patterns, improving the efficacy of the trained classification algorithm (Çukur et al. 2013; Nastase et al. 2017).

Attention to Action Features Sharpens Neural Tuning in the Right pSTS

Our current findings are consistent with previous fMRI univariate mapping studies that have documented stronger and more widespread activation on the pSTS when attention is directed to social dimensions of an event rather than nonsocial features (Tavares et al. 2008; Safford et al. 2010; Lee et al. 2014). Those studies conclude that directed attention to the social aspects of a scene differentially engages neural populations with tuning to features that promote the interpretation on social events. In our study, we found no change in the univariate response across our three attention tasks, likely because all tasks focused on highly salient social aspects of the stimulus (identity of the actor, actions being conducted, or the goals of the actor). Instead we have documented a shift in the statistical structure of the information within the pSTS multivariate activation patterns,

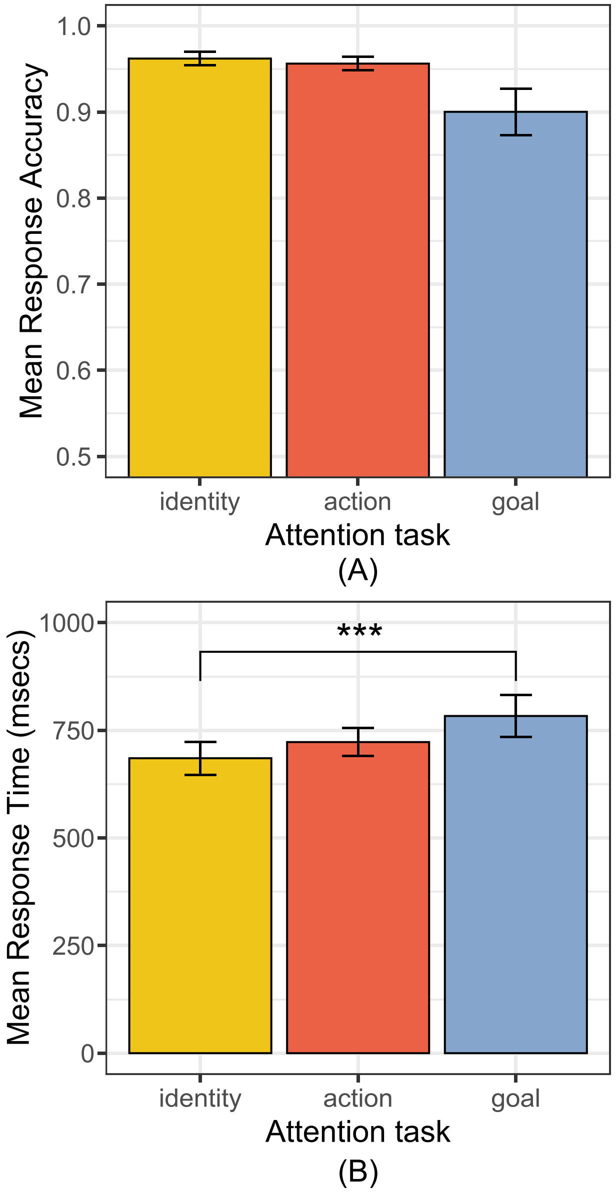

Figure 5. Behavioral results from the performance in the scanner, broken down by attention task. (A) Mean accuracy and (B) response latency (ms) for detecting the feature cued at the beginning of each trial.

without an associated increase (or decrease) in the univariate response. Thus, we conclude that our attention manipulation did not recruit new populations of neurons during action observation but instead altered the information content of the representations constructed during action observation.

In our results, directing attention to the kinematic features of the action vignette improved classification across a wide range of viewing perspectives of the scene, from profile views of the actors with strong lateral movements, to near midline views. Although there is evidence of viewpoint specificity in STS neurons recorded in monkeys (Oram and Perrett 1994), evidence from fMRI strongly favors viewpoint-invariant representations on the human STS (Grossman et al. 2010). In line with this, behavioral research indicates that not all features of an action sequence are equally salient, with key diagnostic features most strongly capturing the attention of the observer (Casile and Giese 2005; Thurman and Grossman 2008). Moreover, with practice, observers can readily identify those salient features and more easily discern action exemplars, with changes in the univariate pSTS response closely following those improvements in training (Grossman et al. 2004; Jastorff et al. 2009). Our current findings are consistent with both of these observations, namely that attention operates on action representations in a manner that is robust to changes in viewpoint and therefore likely reflects the enhanced salience of diagnostic features for action templates, or action categories, rather than specific instances themselves.

Attention to Action Goals Sharpens Neural Tuning in the IFC

Observed directed actions are readily interpreted in the context of the expected outcomes that will be attained and implied intent of the actor (Csibra and Gergely 2007). How this is achieved, however, is a complicated problem as the same action can indicate multiple outcomes and intentions depending on contextual factors, while many actions (variations in body kinematics) have potential to achieve the same ultimate outcome. Evidence indicates that the neural representations in the IFC are closely linked to inferred goals of the actions, even when those goals are specified by body kinematics. For example, the IFC is modulated by the inferred outcomes of observed graping actions rather than the kinematics themselves (Hamilton and Grafton 2008), and neurostimulation of the IFC modulates sensitivity to action outcomes rather than to the body kinematics (Avenanti et al. 2018).

Thus the IFC is important for building mental representations of action goals, whether derived from long-term conceptual representations, the perception of diagnostic action features, or online processing of sensorimotor mappings (Buxbaum and Kalenine 2010). Indeed, there is significant evidence that goal-related neural representations also exist in the inferior parietal lobule (IPL; Grafton and Tipper 2012), likely reflecting the strong connectivity between IPL and IFC (Vry et al. 2015). It is perhaps not surprising then that we observed attending to the goals of the action vignette improved classification accuracy in the IFC.

What is perhaps more surprising is that directed attention to goals did not improve classification accuracy in the pSTS. The reason for this is not entirely clear but may be due to the inherent ambiguity of potential body kinematics to achieve goal outcomes (as noted above) or as a consequence of retaining information as to the vignette goals in a representational state distinct from body kinematics (e.g., as semantic representations).

Attention Does Not Modulate Action Representations in the LOTC

Finally, we did not find evidence for attentional modulation of action representations in other regions of the AON, which is consistent with the proposal that the LOTC is largely sensory driven and tuned to specific body postures and kinematic features (Lingnau and Downing 2015). It is worth noting that each of these regions was able to accurately classify the attentive demands of the three trial instructions, indicating the attentive state of the observer altered how information was processed throughout the AON. Changing task instructions did not, however, render the

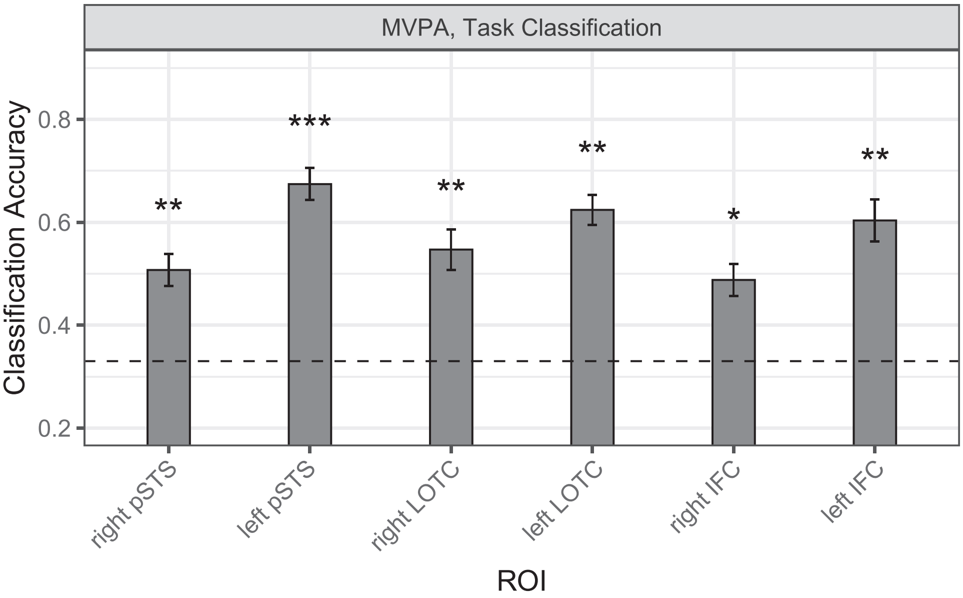

Figure 6. MVPA classification accuracies from decoding task instruction (attend to action, goal, or identity). Error bars indicate SEM. Asterisks indicate statistical significance based on nonparametric permutation tests ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). Dashed line indicates three-way classification accuracy at chance (33%).

two observed actions more or less easily decodable in the most posterior nodes of the AON.

Predictive Coding in Action Observation

Although computational models emphasize bottom-up, feedforward mechanisms of action perception from form and motion features (Hoffman and Flinchbaugh 1982; Mather et al. 1992; Lange and Lappe 2006), biologically inspired models have always noted the top-down influences from prefrontal cortex (Giese and Poggio 2003; Kilner 2011). Prefrontal cortex activation is commonly observed when measuring brain activity during action observation (Saygin et al. 2004; Dasgupta et al. 2016), and interruption to the IFC using noninvasive brain stimulation likewise interferes with action recognition (van Kemenade et al. 2012).

There is mounting evidence that the interpretation of actions, including identifying specific actions and their associated goals, follows a predictive coding framework (Kilner et al. 2007a, 2007b; Urgen and Miller 2015). Empirical studies leveraging the power of dynamic causal modeling (DCM) to infer the direction of causal influence between functionally connected brain regions have revealed both feedforward and feedback connections between IFC and pSTS that are modulated when viewing actions (Gardner et al. 2015; Maffei et al. 2015; Sokolov et al. 2018; Urgen and Saygin 2020).

An important component of predictive coding models is the error signal that is elicited when the observed events mismatch the predicted sensory signals. This error signal has been repeatedly documented in univariate fMRI studies as an increase in the pSTS response when the observed actions violate expectations (Koster-Hale and Saxe 2013; Hillebrandt et al. 2014; Marsh et al. 2014). These include situations in which actors perform irrational reaching and grasping movements (Jastorff et al. 2011), when humans engage unexpectedly in robotlike movements (Saygin et al. 2012; Urgen and Saygin 2020), or when stick figures perform actions at reduced versus normal gravity (Maffei et al. 2015), among other similar violations (Wyk et al. 2009; Gardner et al. 2015; Cardellicchio et al. 2018). Moreover, predictive coding is proposed to operate hierarchically such that cognitively derived internal models can exist in multiple levels of abstraction, from visual kinematic features to the more abstracted action goals (Kilner et al. 2007b; Bach and Schenke 2017).

The pathways by which error signals propagate through the AON are an active area of investigation. Dual pathway models propose distinct structural and functional pathways for action understanding, with a ventrodorsal pathway for action identification further split into a caudal route that codes diagnostic action features and a ventral route that processes action goals (Buxbaum and Kalenine 2010; Binkofski and Buxbaum 2013). Tracing studies in monkeys support the notion of dual pathways, with a dorsal route connecting the upper bank of the STS to premotor cortex via parietal connections, and a second ventral route direct between the lower bank of the pSTS and premotor cortex (Nelissen et al. 2011). In humans, undirected functional connectivity analyses reveal strong connectivity between the IFC and pSTS that carries information unique from that in other segments of the AON network (Dasgupta et al. 2016), and dynamic causal modeling shows that this top-down pathway is strongly modulated by viewing biological motion (Sokolov et al. 2018). As further indirect evidence in support of these top-down models, in this study, we observe attention to mediate this pathway such that directed attention to actions increases functional connectivity between the right IFC and pSTS. Although functional connectivity does not imply direct structural connectivity, it is nonetheless consistent with a model in which neural information is biased along processing pathways contingent on the attentive state of the observer.

Additional Considerations

An important consideration is the possibility that the attentive state guides not just which features to which an observer

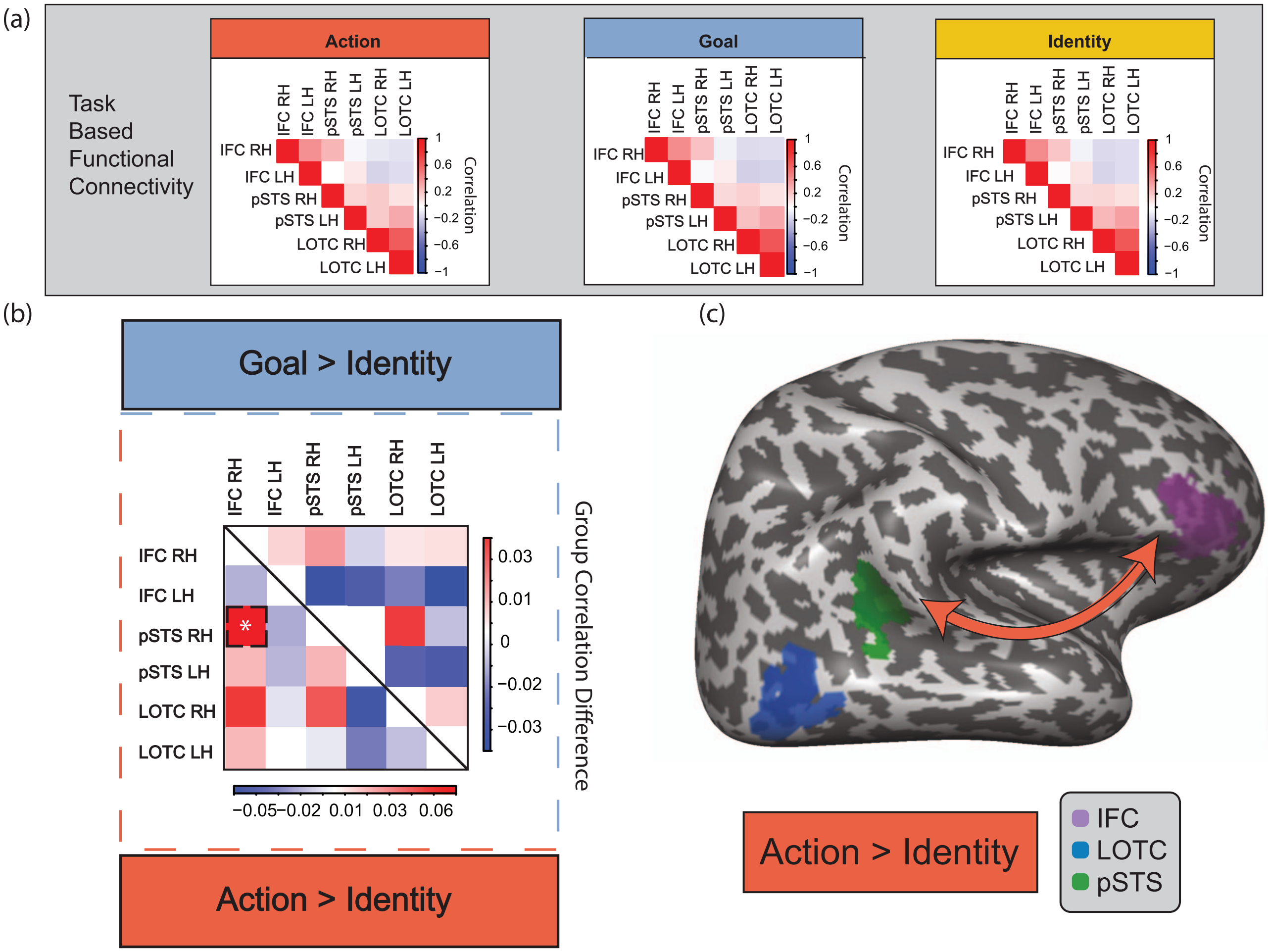

Figure 7. Functional connectivity results. (A) Task-based functional connectivity correlation matrices for each attention task. (B) Right upper matrix is the difference in functional connectivity between attending to goal than attending to identity. Lower left matrix is the difference in correlation between all ROIs when subjects attending to action than attending to identity. (C) ROIs in the current functional connectivity analysis: IFC, LOTC, pSTS. IFC connectivity to pSTS, indicating in red arrow, is significantly more strongly connected during attending to action than identity.

attends, but also the temporal intervals during which attention is directed to the action vignette. Indeed, not all intervals of actions are equally salient (Lu and Liu 2006; Thurman and Grossman 2008), and thus, an efficient observer would deploy attention to the vignette only during the critical moments that are diagnostic to the directed attention task. Ideally, one would be able to identify the diagnostic intervals participants used for the judgments on each trial and model those as independent events. Unfortunately, our experimental design precluded separating different intervals of the action vignette, all of which occurred within 1 sec, without jittered timing, of adjacent events. Moreover, observers withheld all motor responses until after the response cue appears (to prevent motor planning during the vignette), precluding any estimates as to when observers made perceptual judgments based on RTs. To evaluate the importance of critical time windows during action recognition would also benefit considerations of ROI-specific hemodynamic responses, which have known variations in temporal integration windows that will require individualized modeling (Hasson et al. 2008). With those caveats nonwithstanding, the field would benefit from future research identifying and tracing the neural substrates of dynamic attention allocation during action observation.

Conclusions

The pSTS supports the initial perceptual encoding of dynamic body states that underlie particular goals (e.g., hand movements during reaching actions, decoding dynamic facial expressions, and the encoding of limb kinematics during whole-body movements). These perceptual representations are subsequently interpreted by higher-level cognitive systems to support action understanding and intentional states for social interactions. Our findings indicate that the converse is also true: Cognitive systems shape the coding of action representations in the pSTS when observers attend to action features. We propose that the putative functional heterogeneity of pSTS may be accounted for, in part, by top-down influences reflecting the observer's goals when engaged in action observation.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Funding

This work was supported by the National Science Foundation grant BCS-1658560 to EG and BCS-1658078 to JP.

Notes

Conflict of Interest:None declared.

References

Avenanti A, Paracampo R, Annella L, Tidoni E, Algioti, SM. 2018. Boosting and decreasing action prediction abilities through excitatory and Inhibitory tDCS of inferior frontal cortex. *Cerebral Cortex*.. 28:1282–1296.

Bach P, Schenke KC. 2017. Predictive social perception: towards a unifying framework from action observation to person knowledge. *Soc Personal Psychol Compass*. 11:1–17.

Bahnemann M, Dziobek I, Prehn K, Wolf I, Heekeren HR. 2009. Sociotopy in the temporoparietal cortex: common versus distinct processes. *Soc Cogn Affect Neurosci*. 5:48–58.

Bates D, Maechler M, Bolker B, Walker S. 2019. lme4: linear mixed-effects models using 'Eigen' and S4. https://cran.r-project.org/web/packages/lme4/index.html.

Binkofski F, Buxbaum LJ. 2013. Two action systems in the human brain. *Brain Lang*. 127:222–229.

Brainard D. 1997. The psychophysics toolbox. *Spat Vis*. 10: 433–436.

Braunlich K, Love BC. 2019. Occipitotemporal representations reflect individual differences in conceptual knowledge. *J Exp Psychol Gen*. 148:1192–1203.

Buxbaum LJ, Kalenine S. 2010. Action knowledge, visuomotor activation, and embodiment in the two action systems. *Ann N Y Acad Sci*. 1191:201–218.

Cardellicchio P, Hilt PM, Olivier E, Fadiga L, Ausilio AD. 2018. Early modulation of intra-cortical inhibition during the observation of action mistakes. *Sci Rep*. 8:1–9.

Carter RMK, Huettel SA. 2013. A nexus model of the temporal-parietal junction. *Trends Cogn Sci*. 17:328–336.

Casile A, Giese MA. 2005. Critical features for the recognition of biological motion. *J Vis*. 5:348–360.

Csibra G, Gergely, G. 2007. 'Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*. 124:60–78.

Çukur T, Nishimoto S, Huth AG, Gallant JL. 2013. Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci*. 16:763–770.

Dasgupta S, Tyler S, Wicks J, Srinivasan R, Grossman E. 2016. Network connectivity of the right STS in three social perception localizers. *J Cogn Neurosci*. 29:221–234.

Deen B, Koldewyn K, Kanwisher N, Saxe R. 2015. Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb Cortex*. 25:4596–4609.

Downing PE, Wiggett AJ, Peelen MV. 2007. Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *J Neurosci*. 27:226–233.

Fristen KJ, Frith CD, Fletcher P, Liddle PF, Frackowiak RS. 1996. Functional topography: multidimensional scaling and functional connectivity in the brain. *Cereb Cortex*. 6:156–164.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. 1998. Event-related fMRI: characterizing differential responses. *NeuroImage*. 7:30–40.

Frost MA, Goebel R. 2012. Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*. 59: 1369–1381.

Gardner T, Goulden N, Cross XES. 2015. Dynamic modulation of the action observation network by movement familiarity. *J Neurosci*. 35:1561–1572.

Geng JJ, Vossel S. 2013. Re-evaluating the role of TPJ in attentional control: contextual updating? *Neurosci Biobehav Rev*. 37:2608–2620.

Genovese CR, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate 1. *NeuroImage*. 878:870–878.

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

Glover GH. 1999. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*. 9:416–429.

Goebel R, Esposito F, Formisano E. 2006. Analysis of FIAC data with BrainVoyager QX: from single-subject to cortically aligned group GLM analysis and self-organizing group ICA. *Hum Brain Mapp*. 27:392–401.

Grafton S, Tipper C. 2012. Decoding intention: a neuroergonomic perspective — Elsevier enhanced reader.

Grossman ED, Blake R, Kim CY. 2004. Learning to see biological motion: brain activity parallels behavior. *J Cogn Neurosci*. 16:1669–1679.

Grossman ED, Jardine NL, Pyles JA. 2010. fMR-adaptation reveals invariant coding of biological motion on the human STS. *Front Hum Neurosci*. 4:1–18.

Hamilton AF, Grafton ST. 2008. Action outcomes are represented in human inferior frontoparietal cortex. *Cereb Cortex*. 18:1160–1168.

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. 2008. A hierarchy of temporal receptive windows in human cortex. *J Neurosci*. 28:2539–2550.

Haxby JV, Gobbini IM, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 293: 2425–2430.

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

Hillebrandt H, Friston KJ, Blakemore SJ. 2014. Effective connectivity during animacy perception - dynamic causal modelling of human connectome project data. *Sci Rep*. 4:1–9.

Hoffman DD, Flinchbaugh BE. 1982. The interpretation of biological motion. *Biol Cybern*. 204:195–204.

Huk A, Dougherty F, Heeger D. 2002. Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci*. 22:7195–7205.

Jastorff J, Clavagnier S, Gergely G, Orban GA. 2011. Neural mechanisms of understanding rational actions: middle temporal gyrus activation by contextual violation. *Cereb Cortex*. 21:318–329.

Jastorff J, Kourtzi Z, Giese MA. 2009. Visual learning shapes the processing of complex movement stimuli in the human brain. *J Neurosci*. 29:14026–14038.

Jezzard P, Balaban RS. 1995. Correction for geometric distortion in echo planar images from B0 field variations. *Magn Reson Med*. 34:65–73.

Kilner JM. 2011. More than one pathway to action understanding. *Trends Cogn Sci*. 15:352–357.

Kilner JM, Friston KJ, Frith CD. 2007a. Predictive coding: an account of the mirror neuron system. *Cogn Process*. 8:159–166.

Kilner JM, Friston KJ, Frith CD. 2007b. The mirror-neuron system: a Bayesian perspective. *NeuroReport*. 18:619–623.

Kok P, Rahnev D, Jehee JF, Lau HC, De Lange FP. 2012. Attention reverses the effect of prediction in silencing sensory signals. *Cereb Cortex*. 22:2197–2206.

Koster-Hale J, Saxe R. 2013. Theory of mind: a neural prediction problem. *Neuron*. 79:836–848.

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

Lee S, Kable JW. 2018. Simple but robust improvement in multivoxel pattern classification. *PLoS One*. 13:1–15.

Lee SM, Gao T, McCarthy G. 2014. Attributing intentions to random motion engages the posterior superior temporal sulcus. *Soc Cogn Affect Neurosci*. 9:81–87.

Lingnau A, Downing PE. 2015. The lateral occipitotemporal cortex in action. *Trends Cogn Sci*. 19:268–277.

Lu H, Liu Z. 2006. Computing dynamic classification images from correlation maps. *J Vis*. 6:475–483.

Maffei V, Indovina I, Macaluso E, Ivanenko YP, Orban GA, Lacquaniti F. 2015. Visual gravity cues in the interpretation of biological movements : neural correlates in humans. *NeuroImage*. 104:221–230.

Marsh LE, Mullett TL, Ropar D, Hamilton AFC. 2014. Responses to irrational actions in action observation and mentalising networks of the human brain. *NeuroImage*. 103:81–90.

Mather G, Radford K, West S. 1992. Low-level visual processing of biological motion. *Proc R Soc B Biol Sci*. 249:149–155.

Maunsell JH, Treue S. 2006. Feature-based attention in visual cortex. *Trends Neurosci*. 29:317–322.

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. 2018. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*. TU Wien. https://cran.r-project.org/web/packages/e1071/index.html.

Morris JP, Pelphrey KA, McCarthy G. 2008. Perceived causality influences brain activity evoked by biological motion. *Soc Neurosci*. 3:16–25.

Mumford JA, Turner BO, Ashby FG, Poldrack RA. 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*. 59: 2636–2643.

Nastase SA, Connolly AC, Oosterhof NN, Halchenko YO, Guntupalli JS, Di Oleggio Castello MV, Gors J, Gobbini MI, Haxby JV. 2017. Attention selectively reshapes the geometry of distributed semantic representation. *Cereb Cortex*. 27:4277–4291.

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.

Ogawa K, Inui T. 2011. Neural representation of observed actions in the parietal and premotor cortex. *NeuroImage*. 56:728–735.

Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE. 2010. Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *J Neurophysiol*. 104:1077–1089.

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

Patel GH, Sestieri C, Corbetta M. 2019. The evolution of the temporoparietal junction and posterior superior temporal sulcus. *Cortex*. 118:38–50.

Peelen MV, Downing PE. 2007. The neural basis of visual body perception. *Nat Rev Neurosci*. 8:636–648.

Pyles JA, Grossman ED. 2013. Neural mechanisms for biological motion and animacy. In: Johnson KL, Shiffrar M, editors. *People watching: social, perceptual, and neurophysiological studies of body perception. Chapter 17*. New York: Oxford University Press, pp. 304–317.

Reynolds JH, Heeger DJ. 2009. The normalization model of attention. *Neuron*. 61:168–185.

Rissman J, Gazzaley A, D'Esposito M. 2004. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*. 23:752–763.

Saenz M, Buracas GT, Boynton GM. 2002. Global effects of feature-based attention in human visual cortex. *Nat Neurosci*. 5:631–632.

Safford AS, Hussey EA, Parasuraman R, Thompson JC. 2010. Object-based attentional modulation of biological motion processing: spatiotemporal dynamics using functional magnetic resonance imaging and electroencephalography. *J Neurosci*. 30:9064–9073.

Saygin AP, Chaminade T, Ishiguro H, Driver J, Frith C. 2012. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc Cogn Affect Neurosci*. 7:413–422.

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

Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo XN, Holmes AJ, Eickhoff SB, Yeo BTT. 2018. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb Cortex*. 28:3095–3114.

Sokolov AA, Zeidman P, Erb M, Ryvlin P, Friston KJ, Pavlova MA. 2018. Structural and effective brain connectivity underlying biological motion detection. *Proc Natl Acad Sci U S A*. 115:E12034–E12042.

Steffener J, Tabert M, Reuben A, Stern Y. 2010. Investigating hemodynamic response variability at the group level using basis functions. *NeuroImage*. 49:2113–2122.

Stelzer J, Chen Y, Turner R. 2013. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage*. 65:69–82.

Stigliani XA, Weiner XKS, Grill-spector XK. 2015. Temporal processing capacity in high-level visual cortex is domain specific. *J Neurosci*. 35:12412–12424.

Summerfield C, Egner T. 2009. Expectation (and attention) in visual cognition. *Trends Cogn Sci*. 13:403–409.

Tavares P, Lawrence AD, Barnard PJ. 2008. Paying attention to social meaning: an fMRI study. *Cereb Cortex*. 18:1876–1885.

Thompson EL, Bird G, Catmur C. 2019. Conceptualizing and testing action understanding. *Neurosci Biobehav Rev*. 105: 106–114.

Thompson J, Parasuraman R. 2012. Attention, biological motion, and action recognition. *NeuroImage*. 59:4–13.

Thurman S, Grossman ED. 2008. Temporal "bubbles" reveal key features for point-light biological motion perception. *J Vis*. 8:1–11.

Treue S, Martínez Trujillo JC. 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*. 399:575–579.

Turner BO, Mumford JA, Poldrack RA, Ashby FG. 2012. Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage*. 62: 1429–1438.

Urgen BA, Saygin AP. 2020. Predictive processing account of action perception: evidence from effective connectivity in the action observation network. *Cortex*. 128:132–142.

Urgen BA, Miller LE. 2015. Towards an empirically grounded predictive coding account of action understanding. *J Neurosci*. 35:4789–4791.

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.

Vangeneugden J, De Maziere PA, Van Hulle MM, Jaeggli T, Van Gool L, Vogels R. 2011. Distinct mechanisms for coding of‘ visual actions in macaque temporal cortex. *J Neurosci*. 31:385–401.

Vry MS, Tritschler LC, Hamzei F, Rijntjes M, Kaller CP, Hoeren M, Umarova R, Glauche V, Hermsdoerfer J, Goldenberg G *et al*. 2015. The ventral fiber pathway for pantomime of object use. *NeuroImage*. 106:252–263.

Weiner KS, Grill-Spector K. 2011. Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *NeuroImage*. 56: 2183–2199.

Weiner KS, Grill-Spector K. 2013. Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychol Res*. 77: 74–97.

Wurm MF, Lingnau A. 2015. Simultaneously learning at different levels of abstraction. *J Neurosci*. 35:7727–7735.

Wyk BCV, Hudac CM, Carter EJ, Sobel DM, Pelphrey KA. 2009. Action understanding in the superior temporal sulcus region. *Psychol Sci*. 20:771–777.