Received: 24 October 2018 | Revised: 24 January 2019 | Accepted: 28 January 2019

DOI: 10.1002/hbm.24540

WILEY

RESEARCH ARTICLE

Parts-based representations of perceived face movements in the superior temporal sulcus

Ben Deen | Rebecca Saxe

Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

Correspondence
Ben Deen, Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Room 46-4021, 43 Vassar St., Cambridge, MA 02139.
Email: benjamin.deen@gmail.com

Funding information
National Institutes of Health, Grant/Award Numbers: MH096914-01A1, MH096914-01A; National Science Foundation, Grant/Award Number: CCF-1231216; David and Lucile Packard Foundation

Abstract

Facial motion is a primary source of social information about other humans. Prior fMRI studies have identified regions of the superior temporal sulcus (STS) that respond specifically to perceived face movements (termed fSTS), but little is known about the nature of motion representations in these regions. Here we use fMRI and multivoxel pattern analysis to characterize the representational content of the fSTS. Participants viewed a set of specific eye and mouth movements, as well as combined eye and mouth movements. Our results demonstrate that fSTS response patterns contain information about face movements, including subtle distinctions between types of eye and mouth movements. These representations generalize across the actor performing the movement, and across small differences in visual position. Critically, patterns of response to combined movements could be well predicted by linear combinations of responses to individual eye and mouth movements, pointing to a parts-based representation of complex face movements. These results indicate that the fSTS plays an intermediate role in the process of inferring social content from visually perceived face movements, containing a representation that is sufficiently abstract to generalize across low-level visual details, but still tied to the kinematics of face part movements.

KEYWORDS

face, holistic, motion, MVPA, parts-based, social perception, STS

1 | INTRODUCTION

Facial motion provides a critical source of social information about others, regarding their emotional state, direction of attention, and vocal utterances. Among the set of face-responsive regions in the human brain, it has been argued that regions in the superior temporal sulcus (STS) are specialized for processing face motion and changeable aspects of faces (Allison, Puce, & McCarthy, 2000; Haxby, Hoffman, & Gobbini, 2000). In contrast with ventral temporal regions, face-responsive regions in the STS respond substantially more strongly to moving than to static faces, and prefer naturalistic motion to videos that are temporally scrambled (Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011; Schultz, Brockhaus, Bülthoff, & Pilz, 2013). Studies using static face images have found that these regions adapt to repeated presentations of the same facial expression, even when facial identity is varied, pointing to an identity-invariant representation of expression (Andrews & Ewbank, 2004; Harris, Young, & Andrews, 2012; Winston, Henson, Fine-Goulden, & Dolan, 2004).

Despite compelling evidence for a role of the STS in processing perceived face motion, very little is known about the nature of face movement representations in this region. Multivoxel pattern analysis (MVPA) provides a powerful technique for characterizing neural representations, by asking which stimulus dimensions can be decoded from subtle variations in spatial patterns of response within a region. Said, Moore, Engell, Todorov, and Haxby (2010) found that response patterns to dynamic facial stimuli in anatomically defined anterior and posterior STS regions could be used to classify seven different emotional expressions. Skerry and Saxe (2014) found that responses patterns of a face-responsive STS subregion could classify positively-from negatively-valenced dynamic facial stimuli.

While these studies demonstrate that relevant pattern information can be read out from the STS, many questions remain about the nature of the representations underlying these effects. First, these studies did not attempt to dissociate similarity of facial expression from low-level visual similarity (i.e., pixel-wise similarity of videos, and resulting similarity of early visual representations), insofar as

responses to the same videos were used both for training and testing classifiers. Does the STS contain representations of face movements that are abstracted from low-level visual properties? Second, these studies used full-face emotional expressions that differed in motions of several face parts. Does the STS represent more subtle distinctions in the motion of individual face parts?

Furthermore, how do the representations of complex face movements relate to the movements of individual parts of the face? Facial expressions typically consist of coordinated movements of different face parts, and there is behavioral evidence that expressions are processed holistically: the expression the top or bottom half of a face influences the perceived expression in the other half (Calder, Young, Keane, & Dean, 2000). Does face-responsive STS integrate motion information from multiple face parts to generate a holistic, full-face motion representation? Or are complex movements represented in terms of motion of different parts of the face?

In the current study, we use fMRI and MVPA to address these questions and provide a richer account of the representational content of face-responsive STS. Participants viewed a set of dynamic face movements, including four eye/eyebrow movements, four mouth movements, and combinations of these, performed by one of two actors and presented in one of four spatial positions. A separate behavioral study demonstrated that these stimuli were perceived holistically, based on the presence of a composite effect. To test for parts-based versus holistic neural representations, we asked: can the pattern of response to a combined face movement be predicted from a linear combination of the responses to the eye and mouth component movements? Or is the response to a combined movement distinct from, and not predictable by, responses to component movements? Our results suggest that (a) face-sensitive regions of the STS represent subtle discriminations in type of face movement; (b) these representations generalize across two actors and small differences in visual position; and (c) complex movements are represented as a sum of their parts.

2 | MATERIALS AND METHODS

2.1 | Methods preregistration

In order to reduce the risk of false positive results related to researcher degrees of freedom, and thus bolster the reproducibility of our results, we formally preregistered our experimental methods for both fMRI and behavioral experiments using the Open Science Framework (Deen, 2015, 2016). The stimuli and task, number of participants, acquisition parameters, and most of the analysis pipeline were determined before any data analysis was performed. Analyses that were not part of the preregistration will be explicitly described as such. Stimuli, experimental scripts, and mask files used for analysis can be found as part of the preregistrations.

2.2 | Participants

Twenty-four adults participated in the fMRI study (age 21–36, 10 females), and thirty adults participated in the behavioral study (age 21–37, 15 females), with ten participants shared across both. Participants had no history of neurological or psychiatric impairment, and normal or corrected vision. All participants provided written, informed consent. No further exclusion criteria were used.

2.3 | fMRI paradigm

In the main experiment, participants viewed videos of faces performing a variety of face movements. The videos were generated using Poser 8 character animation software (http://my.smithmicro.com/poser-3d-animation-software.html), allowing tight control over visual properties of the stimuli. The movements included four eye or eyebrow movements (brow raise, eye closing, eye roll, scowl), four mouth movements (smile, frown, mouth opening, snarl), and sixteen combined eye/mouth movements corresponding to all possible combinations of the individual eye and mouth movements (Figure 1). The actions were be performed by two avatars ("actors"), one male and one female––the Poser characters Simon and Sydney. Videos were 2 s long and 30 frames per second, and consisted of a neutral (expression-less) face for 20 frames, a face movement lasting 20 frames, and the face with its final expression for 20 frames (images of the final expression are shown in Figure 1). Throughout the scan, participants fixated centrally, while face videos were presented slightly eccentric, at 0.5° of visual angle in one of four locations: to the upper left, upper right, lower left, and lower right. The faces in these videos subtended 4.2° by 5.9° of visual angle. Across 24 actions, 2 actors, and 4 visual positions, there were 192 distinct stimuli. Throughout a single scan session, four repetitions of each of these stimuli were presented.

Within each run, stimuli were presented in a jittered event-related design. Each clip lasted 2 s, with a variable interstimulus interval (ISI) of 2, 4, or 6 s, each occurring with 1/3 probability. To maintain attention, participants performed a one-back task on the action and actor in the video, irrespective of visual position (i.e., pressed a button with their right pointer finger when the actor and action were repeated across two subsequent trials, regardless of whether visual position repeated or not). In a given run, 48 stimuli were presented (such that all 192 are presented across four runs), as well as 6 repeated stimuli, which did not contribute to pattern analyses. Each run contained one example of each action/actor pair, in one of four visual positions. The order of stimuli in each run were randomized across runs and participants. With a fixation block of 10 s at the beginning of the experiment and 8 s at the end of the experiment, each run lasted 5.7 min, and 16 runs were acquired throughout a scan session.

In addition to the main experiment, we ran a face localizer to define face-responsive subregions of the STS. Participants passively viewed videos of dynamic faces and dynamic objects. The dynamic face condition consisted of 60 close-up shots of faces (five females and three males children). These videos showed a range of face movements, including mouth movements (speech, laughter, smiles), eye and head movements, and a variety of full-face expressions (e.g., surprise, excitement). The dynamic object condition consisted of 60 close-up shots of moving objects (15 different children's toys, such as mobiles, windup toys, toy vehicles, and balls rolling down inclines). Each video lasted 3 s and was presented in a block of six videos, lasting 18 s. Blocks were presented in palindromic order, with condition order counterbalanced across runs and participants. There were six blocks each of faces and objects, as well as baseline blocks at the beginning, middle, and end of the experiment, in which six uniform color fields

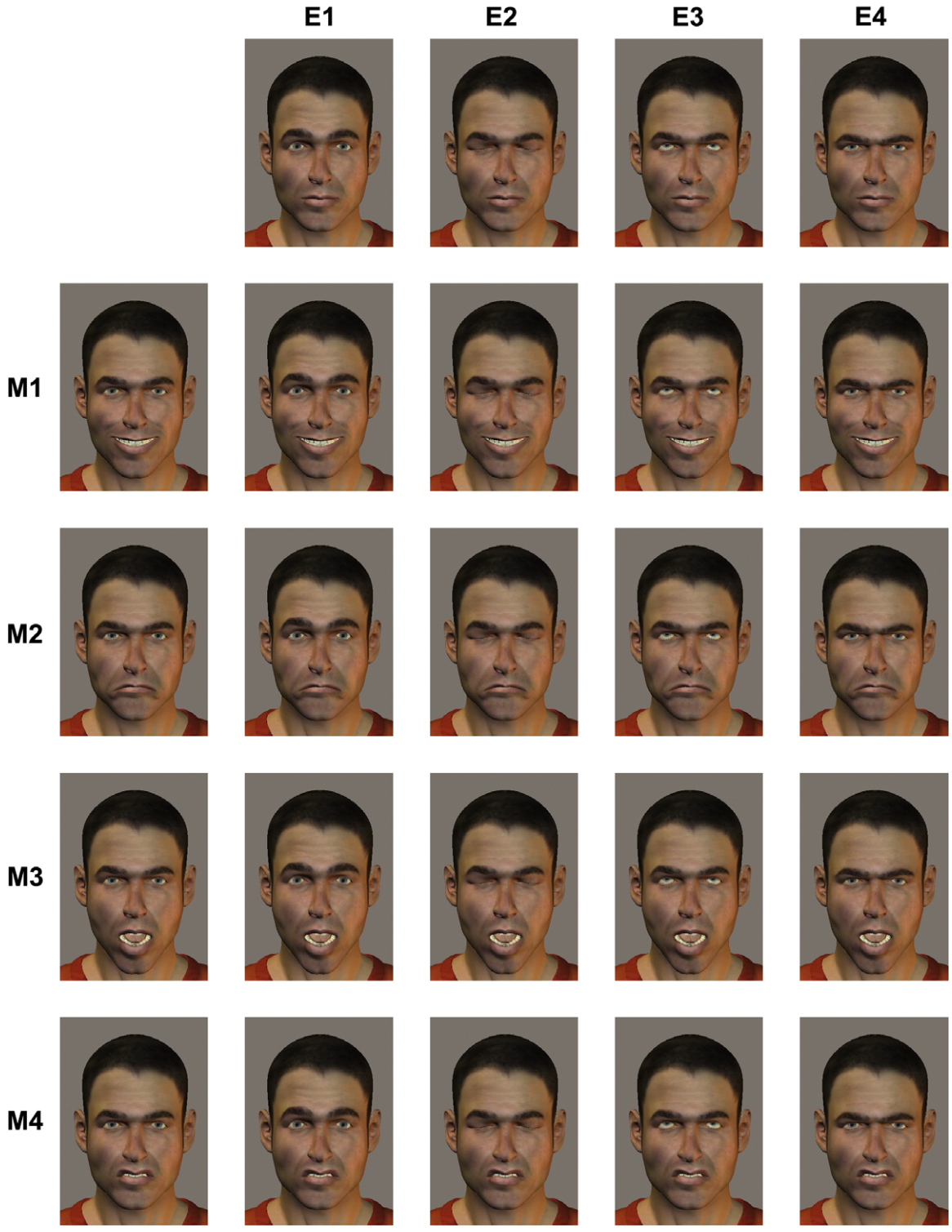

FIGURE 1 Sample frames from video stimuli depicting face movements, from one of two actors. The stimulus set consisted of four eye/eyebrow movements, four mouth movements, and sixteen combined (eye and mouth) movements

were presented for 3 s each. Each run lasted 4.5 min, and 4 runs were acquired throughout a scan session. Further details about the stimuli have been reported previously (Pitcher et al., 2011).

2.4 | Behavioral paradigm

In addition to the fMRI experiment, we ran a behavioral experiment to verify that the specific stimuli used here are perceived holistically, by testing the presence of a composite effect (Young, Hellawell, & Hay, 1987), or an influence of face configuration in the mouth region on perception of the eye region. Participants viewed images corresponding to the final frame of the face movement videos. We used images rather than videos so that stimuli could be presented for a short duration (200 ms rather than 2 s), in order to be consistent with the prior literature on the composite effect, to limit participants' opportunity to saccade, and to increase the difficulty of the behavioral task to avoid

a ceiling effect. Prior research has found comparable composite effects for dynamic stimuli (videos of expression formation) and static images of the expression formed (Tobin, Favelle, & Palermo, 2016). In the "aligned" condition, the top and bottom portions of the face (containing the eye/eyebrow and mouth, respectively) were presented in vertical alignment. In the "misaligned" condition, the bottom half of each face was shifted horizontally to the right, such that the nose is aligned with the right edge of the head. Top and bottom face halves always came from the same identity. Stimuli were presented on a 1440 × 900 resolution 15″ monitor, positioned 80 cm from the participant. Intact faces were 6 × 9 cm (4.3 × 6.4° of visual angle). Face images were positioned such that a black central fixation cross lay roughly between the eyes, and were presented on a gray background.

Participants sat with their head in a chin-rest to stabilize eye position. On each trial, a face appeared for 200 ms, followed by an ISI of 400 ms, and a second face for 200 ms. The movement type in the bottom half of the two faces always differed, while the movement type in the top half could be the same or different, with identity always the same. Participants were asked to fixate centrally throughout the experiment, and to judge whether the appearance of the eyes and eyebrows was identical or different across the two faces, while ignoring the appearance of the mouth, by responding with a key press as quickly and accurately as possible. Responses were recorded within a window of 1 s following stimulus presentation, followed by a random intertrial interval of 1–2 s.

Participants performed six runs of the experiment, each lasting roughly 5 min and containing 96 trials. Runs contained 12 trials for each combination of same/different, aligned/misaligned, and character identity, comprising three trials for each of four top-half movement types (for same trials) or two trials for each of six pairs of movement types (for different trials). Pairs of bottom-half movement types were randomly assigned to corresponding top-half movement types for a given run, but with top/bottom-half combinations matched across same/different and aligned/misaligned trial types, and with each bottom-half pair occurring with equal probability. All trial types were randomly intermixed within each run, with trial order counterbalanced across runs and participants. The second face in each trial were randomly jittered in position within a range of 7 × 7 mm (0.5 × 0.5° visual angle) to prohibit a low-level image matching strategy, with the amount of jitter matched on a trial-by-trial basis across same/different and aligned/misaligned trial types. Prior to the main experiment, participants received a practice phase consisting of 10 sample trials, with feedback provided. For data analysis, all trials with a recorded response were included, and statistics were performed across participants.

2.5 | MRI data acquisition

MRI data were acquired using a Siemens 3T MAGNETOM Tim Trio scanner (Siemens AG, Healthcare, Erlangen, Germany). High-resolution T1-weighted anatomical images were collected using a MPRAGE pulse sequence (repetition time [TR] = 2.53 s; echo time [TE] = 3.48 ms, flip angle $\alpha = 7°$, field of view [FOV] = 256 mm, matrix = 256 × 256, slice thickness = 1 mm, distance factor = .5, 176 near-axial slices, GRAPPA acceleration factor = 2, 24 reference lines). Functional data were collected using a T2*-weighted echo planar imaging pulse sequence sensitive to blood-oxygen-level-dependent (BOLD) contrast (TR = 2 s, TE = 30 ms, $\alpha = 70°$, FOV = 192 mm, matrix = 966 × 96, slice thickness = 2 mm, 42 near-axial slices, multiband acceleration factor = 2, phase partial Fourier = 6/8).

2.6 | Data preprocessing and modeling

fMRI data were processed using the FMRIB Software Library (FSL), version 4.1.8, supplemented by custom MATLAB scripts. Anatomical and functional images were skull-stripped using FSL's brain extraction tool. Functional data were motion corrected using rigid-body transformations to the middle image of each run, and high-pass filtered (Gaussian-weighted least squares fit straight line subtraction, with σ = 50s (Marchini & Ripley, 2000)). Localizer data were also spatially smoothed with a 4 mm-FWHM isotropic Gaussian kernel, while data from the main task were not smoothed. For the purpose of analyzing group-level data in searchlight analyses, functional data were registered to the Montreal Neurological Institute 152 template brain (MNI space) using the following procedure: functional data were registered to anatomical images using a rigid-body transformation determined by Free surfer's bbregister (Greve & Fischl, 2009), and anatomical images were in turn registered to MNI space using a nonlinear transformation determined by FSL's FNIRT.

Whole-brain general linear model-based analyses were performed for each participant, run, and task. Regressors were defined as boxcar functions convolved with a canonical double-gamma hemodynamic response function. All regressors were temporally high-pass filtered in the same way as the data. FSL's FILM was used to correct for residual autocorrelation (Woolrich, Ripley, Brady, & Smith, 2001). Data from each run and task was registered to the middle volume of the first run of the main task using a rigid-body transformation determined by FSL's FLIRT, and further data analysis took place in this space.

For modeling data from the main task, we used the least-squares-single method (Mumford, Turner, Ashby, & Poldrack, 2012). In this approach, a separate model is run for each trial, which consists of one regressor for the trial of interest, and one regressor for all other trials. This provides more accurate and lower variance estimates of response magnitudes for single trials in event-related designs with relatively small ISIs, by reducing collinearity between regressors in each model.

2.7 | Region-of-interest definition

Analysis of the main task data was conducted using independently defined regions-of-interest (ROI). We focused on three functionally defined ROIs: motion-sensitive voxels within the calcarine sulcus (termed early visual cortex, EVC), motion-sensitive lateral occipitotemporal cortex (loosely termed MT+) and face-sensitive right STS (fSTS). The first two ROIs were intended as controls that were not expected to contain action representations. The EVC ROI was defined for each participant by identifying voxels sensitive to visual motion (voxels responding to dynamic faces and objects over a changing color field baseline, $p < 0.001$ voxelwise, in localizer data) within an anatomically defined bilateral calcarine sulcus ROI, from Free surfer's Desikan-Killiany cortical parcellation. The MT+ ROI was defined for each participant by identifying voxels sensitive to visual motion (same criterion

as above), within a bilateral lateral occipitotemporal search space defined from a group-level activation map to visual motion (coherently moving dots vs. static dots) in a separate dataset of 20 participants.[1]

The fSTS ROI was defined functionally by identifying face-sensitive voxels in the right STS. While face responses are most consistently reported in posterior parts of the STS, middle and anterior STS responses have also been reported (Pitcher et al., 2011; Winston et al., 2004). Prior studies have not observed clear functional differentiations between these areas, and thus do not suggest hypotheses as to which contain motion representations. For this reason, we chose to simply consider all face-responsive voxels within the STS. The fSTS in each participant was defined as set of voxels within an anatomical right STS mask that respond significantly ($p < 0.001$ voxelwise) to faces over objects in the localizer task. The anatomical mask was defined by manually drawing STS gray matter on the MNI brain. Any participant who had less than 50 voxels in the resulting ROI was excluded from the fSTS analysis; two participants were excluded based on this criterion.

In addition to predefined ROIs, we performed a hypothesis-neutral search for other brain regions containing action information by using a searchlight analysis across the whole brain. Specifically, we searched for regions whose patterns can discriminate the 24 action conditions, generalizing across position, as described in detail below. We searched across 8 mm-radius spheres centered at each voxel in a gray matter mask, with each sphere intersected with the mask. The mask was defined using the MNI gray matter atlas, thresholded at 0%, and intersected with each individual participant's brain mask. Statistical maps within participants were registered to MNI space to perform inference across participants. Because coverage was only near-whole-brain and differed slightly across participants, we only considered voxels in which every participant had data. The resulting statistical map was thresholded at $p < 0.01$ voxelwise to form contiguous clusters of activation (where two voxels are considered contiguous if they share a vertex). To correct for multiple comparisons across voxels, we used a permutation test with 5,000 iterations to generate a null distribution for cluster sizes, and used this to threshold clusters of activation at $p < 0.05$.

2.8 | Multivoxel pattern analysis

We next used MVPA to determine which features of our face motion stimuli could be discriminated by patterns of response within each ROI. In particular, we used the Haxby correlation method (Haxby et al., 2001). In this approach, the data are first split into two halves, and patterns of response to N distinct conditions are computed in each half. Then, a matrix of Fisher-transformed correlations between patterns from the first half and the second half of the data is computed, and for each participant, a difference score or "discrimination index" is computed: the mean within-condition correlation minus the mean between-condition correlation (i.e., the mean of the diagonal elements of this correlation matrix minus the mean of the off-diagonal elements; depicted in Figure 3a). Lastly, a one-tailed t-test is performed across participants to determine if these difference scores are significantly greater than zero, indicating that patterns in this region discriminate between the conditions tested. We did not correct for multiple comparisons across the three predefined ROIs, insofar as EVC and MT+ were intended as controls, and fSTS was hypothesized to contain action representations.

As a control measure, we first checked for discrimination of visual position, which we expected to find in EVC and MT+, but not fSTS. For this analysis, we split the data in half by trial number (averaging trial repetitions 1 and 3, and 2 and 4), collapsing data over actions and actors. For each region, we constructed a 4×4 split-half correlation matrix, treating each position as a distinct condition, and assessed the difference score for this matrix.

To test for the presence of action representations, we performed a hierarchy of analyses, in which we first tested whether a region's patterns could discriminate among the 24 action conditions, and if this was the case, tested several more specific discriminations to detail the nature of action representations. Each of these tests were run in two ways, requiring generalization across either position (left vs. right) or actor, by splitting the data across this dimension to compute the split-half correlation matrix. Generalization across position was considered a prerequisite for an abstract action representation, and therefore we only tested further hypotheses if a region's patterns contained action information that generalized across position. For the initial test for action information, we constructed a 24×24 split-half correlation matrix, treating each action as a distinct condition, and tested the difference score for this matrix.

This analysis revealed that fSTS, but not MT+ or EVC, contained patterns that discriminated actions across position. Thus for this region, we next performed further specific tests. Three of these assessed the nature of representations of isolated eye and/or mouth movements, termed single movements (as opposed to combined eye and mouth movements). First, we tested for discrimination of eye versus mouth movements, by considering the 8×8 submatrix of correlations between isolated movements, and treating eye to eye and mouth to mouth correlations as within condition, but eye to mouth and mouth to eye correlations as between condition. We also tested for discrimination of specific eye movements, by computing a difference score from the 4×4 submatrix of eye movements, and did the same for mouth movements (termed eye type and mouth type).

These tests for eye and mouth type information were relatively underpowered, using only 4×4 submatrices of a 24×24 correlation matrix. We thus ran two additional unplanned analyses to test for discrimination of eye and mouth type, taking advantage of the larger amount of data provided by responses to combined movements. First, we tested for discrimination of eye and mouth type within combined movements. Second, we tested for discrimination of eye and mouth type across single and combined movements—that is, by assessing correlations between patterns of response to single and combined movements.

We next ran two analyses to probe the nature of representations of combined eye/mouth movements. One possibility is that these

[1]Our planned analysis strategy was to use the anatomical regions themselves (calcarine sulcus, lateral occipitotemporal search space) as control ROIs, without an additional functional criterion. We added the functional criterion to make control ROIs more comparable to the fSTS ROI, based on reviewer feedback. Results with the originally planned ROIs were qualitatively identical (Supporting Information Figure S1).

movements are encoded in a parts-based manner, such that the neural response to a combined movement is roughly the sum of neural responses to eye and mouth movements; this might be expected of a region that encodes the kinematics of face movements. Another possibility is that these representations are holistic, in that the neural response to combined movements cannot be decomposed into responses to individual components; this would be expected, for instance, of a region that encodes the emotion expressed by a face movement. These alternatives are not mutually exclusive: a region could contain neural subpopulations with both types of code.

We tested for the presence of parts-based representations by asking whether patterns of response to combined eye/mouth movements could be discriminated by linear combinations of patterns of response to the isolated movements. Within the first half of the dataset, we used linear regression to find the linear combination of eye and mouth patterns that best predicted the combined pattern (depicted in Figure 4a). We then computed a 16 × 16 split-half correlation matrix between these "simulated" combined patterns from the first half, and empirical combined patterns from the second half. To maximize power, we computed two such matrices, where the simulated patterns were computed from either the first or second half of the dataset, and averaged these together. Finding a significant difference score from this matrix would indicate that combined patterns could be discriminated by linear combinations of eye and mouth patterns.

To test for the presence of holistic representations, we asked whether combined patterns themselves do a better job of discriminating responses to combined movements in left out data than the simulated patterns do. In particular, we computed a difference score for split-half correlations between responses to the 16 combined movements, and asked whether this was significantly greater than the difference score for simulated-to-combined correlations, described above.[2] Finding a significant difference score in this matrix would indicate the presence of discriminative pattern information in responses to combined movements that is not captured by the simulated patterns, pointing to a holistic representation.

2.9 | Univariate analysis

To address whether differences in fSTS patterns across conditions were accompanied by differences in mean response magnitude of the region, we added an unplanned control analysis. We analyzed the mean response of the fSTS to each of the 24 action conditions, by averaging beta values across voxels in the region, as well as trials, actors, and positions. Post hoc repeated measures ANOVAs were used to assess modulation of mean fSTS responses by action type.

[2]This approach differed slightly from our planned analysis, which compared within-condition correlations, rather than within/between difference scores. Upon analyzing the data, it became clear that simulated patterns had lower variance than responses to combined movements, which biases toward increased split-half correlations for simulated patterns. Because both within- and between-condition correlations are similarly influenced by differences in variance between simulated and combined patterns, the approach reported here is less influenced by this bias. This difference in analysis did not influence our conclusion regarding holistic processing.

3 | RESULTS

3.1 | Behavioral results: Composite effect

The present study used dynamic face movement stimuli, constructed as combinations of isolated eye/eyebrow and mouth movements. To verify that the face movements depicted in our stimuli were perceived holistically, we first ran a behavioral experiment. Specifically, we tested whether the movement type in the mouth region influenced the perceived movement type in the eye region, only when top- and bottom-halves are vertically aligned. (While we use the term "movement type" for consistency, the behavioral experiment used still images of formed expressions rather than full movements.) Following prior studies of the composite effect (Goffaux & Rossion, 2006; Mondloch & Maurer, 2008; Robbins & McKone, 2007), we compared performance between same-aligned and same-misaligned trials as an index of holistic processing. Because we expected that performance on the task would be near ceiling, our pre-planned analysis strategy focused on reaction time (RT; Deen, 2016). Indeed, we found that RTs were significantly longer for same-aligned trials than for same-misaligned trials (Figure 2; $t[29] = 7.45$, $p < 10^{-7}$, RT difference 29.5 ms), indicating that the differing mouth-movement-types disrupted the perception of eye-movement-types as the same, specifically when top and bottom face halves were vertically aligned. While accuracy was near ceiling as expected, an unplanned analysis also revealed a similar effect on accuracy, which was lower for same-aligned trials than same-misaligned trials ($t[29] = 2.48$, $p < 0.01$, accuracy difference 2%). These results indicate that the face movements depicted in our stimuli were indeed perceived holistically, despite being created by combining distinct animated eye/eyebrow and mouth movements.

3.2 | Action representations in fSTS

These stimuli were next used in an fMRI experiment, to assess the nature of cortical representations of perceived face movements using MVPA. We first asked whether patterns of response in face-sensitive regions of the superior temporal sulcus (fSTS) contained information about face movement (action) type (Figure 3b). Action information was observed, both when requiring generalization across visual position ($t[21] = 1.90$, $p < 0.05$), and across actor ($t[21] = 4.37$, $p < 10^{-3}$). This indicates that the fSTS contains a position-tolerant representation of perceived face movements, more tied to the movements themselves than to actor-movement pairs.

Subsequent analyses tested alternative a priori hypotheses about the face movement representations in the fSTS. First, we found that patterns of fSTS response could discriminate eye from mouth movements, generalizing across both position ($t[21] = 2.90$, $p < 0.01$) and actor ($t[21] = 4.12$, $p < 10^{-3}$). Can fSTS patterns make the more fine-grained discrimination between different specific eye movements, and specific mouth movements? Within single (eye- or mouth-only) movements, we found no evidence for discrimination of specific movements, either when generalizing across position or actor (P's > 0.05). However, this negative result could result from a lack of power in these analyses, which focused on 4 × 4 submatrices of a 24 × 24

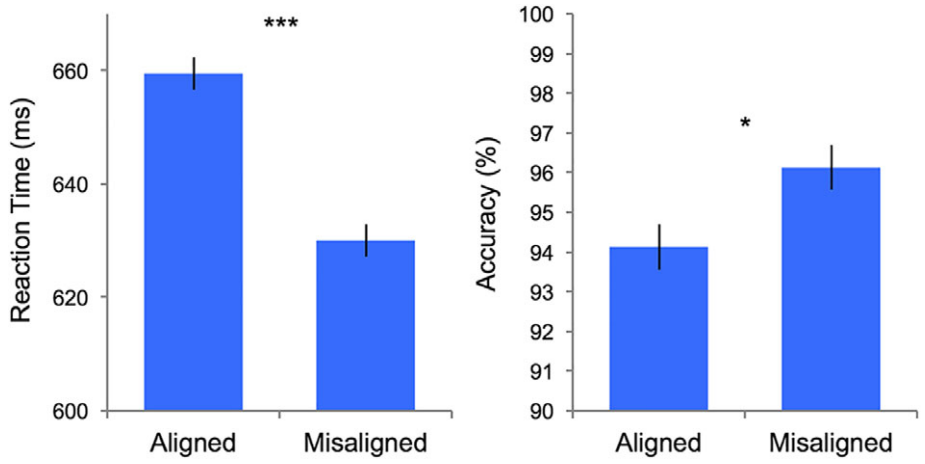

FIGURE 2 Behavioral results from a composite effect paradigm. When top and bottom face halves are vertically aligned (such that differing mouth movement types interferes with the perception of eye movement types as identical), RT is longer, and accuracy is lower. Error bars show within-subject standard error, computed following the strategy of Morey (2008). *Denotes $p < 0.05$ and *** denotes $p < 10^{-3}$

correlation matrix. To address this possibility, we performed an unplanned analysis to ask whether fSTS patterns discriminated type of eye or mouth movement within the combined (eye and mouth) movements, of which there were 16 rather than 4. We found significant discrimination of eye motion type, generalizing across both position ($t[21] = 2.32$, $p < 0.05$) and actor ($t[21] = 2.48$, $p < 0.05$), as well

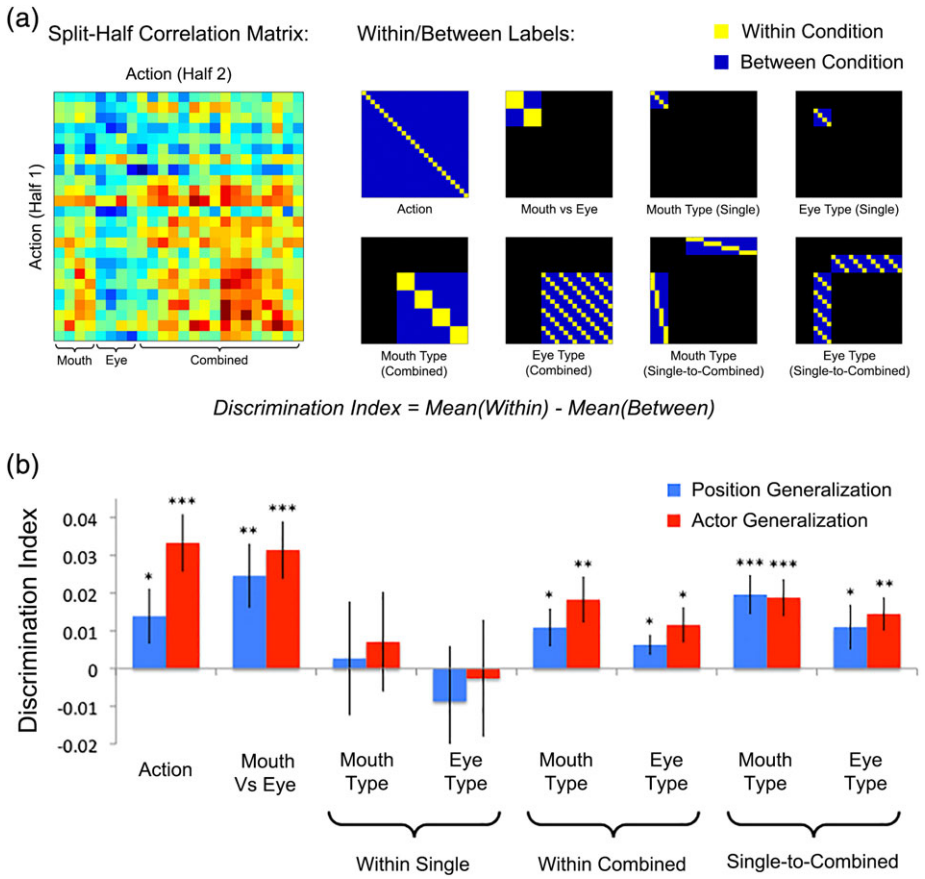

FIGURE 3 (a) Depiction of correlation difference method used for MVPA. On the left is a matrix of split-half correlations of patterns of response to each action (where this split is either across visual position or actor). On the right are a set of matrices indicating which cells are within-condition correlations, and which are between-condition correlations, for a number of tests. Discrimination indices are computed as the difference between within-condition and between-condition correlations (Fisher-transformed). (b) Discrimination indices for various analyses of fSTS patterns. "Mouth type" and "eye type" refer to discrimination of one of four specific mouth (or eye) movements. "Single" refers to individual eye and mouth movements, while "combined" refers to stimuli with both eye and mouth motion. Single-to-combined analyses assessed correlations between patterns of response to single and combined stimuli. *Denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 10^{-3}$

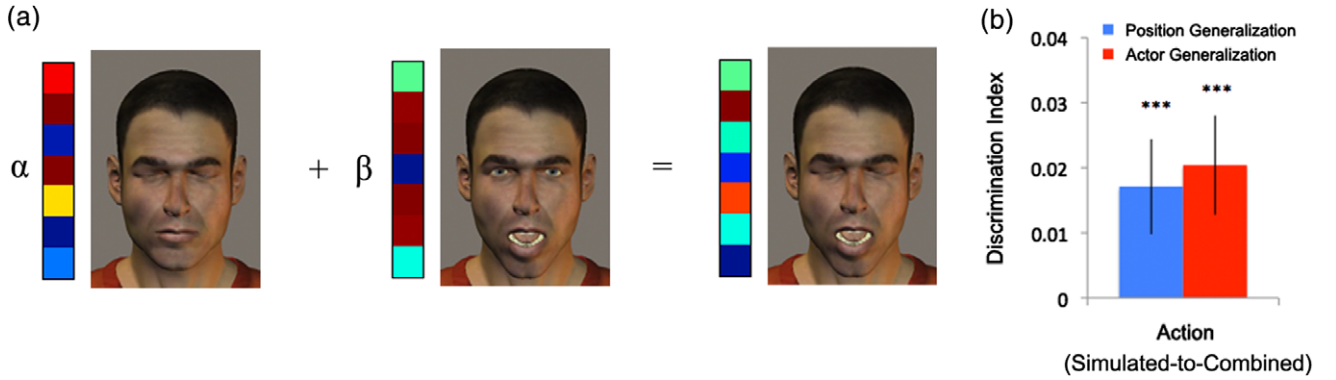

FIGURE 4 Evidence for a parts-based representation of combined face movements in the fSTS. (a) Method: In one half of the dataset, "simulated patterns" were constructed for each combined movement, as a linear combination of responses to the corresponding individual eye and mouth movements. These simulated patterns were then used to discriminate patterns of response to combined movements in the second half of the dataset. (b) Results from the simulation analysis. ***Denotes $p < 10^{-3}$

as discrimination of mouth motion type, generalizing across both position ($t[21] = 2.19$, $p < 0.05$) and actor ($t[21] = 3.03$, $p < 0.01$). This results indicates that fSTS represents subtle distinctions between types of perceived eye and types of mouth movement.

To bolster this result, we performed a further unplanned analysis, which attempted discriminate specific eye and mouth movements by assessing correlations between patterns of response to single and combined movements. From a machine learning perspective, this corresponds to training a movement type classifier on single movements, and testing on combined movements (and vice versa). In this analysis, both eye and mouth movements could be discriminated, both generalizing across position (eye: $t[21] = 1.88$, $p < 0.05$; mouth: $t[21] = 3.78$, $p < 10^{-3}$) and actor (eye: $t[21] = 3.29$, $p < 0.01$; mouth: $t[21] = 3.81$, $p < 10^{-3}$). This result demonstrates the presence of information about specific movement type even in patterns of response to single movements. Furthermore, this demonstrates that pattern information about eye movement type and mouth movement type generalize from responses to individual eye and mouth movements to combined movements.

3.3 | Parts-based versus holistic representations

How do patterns of fSTS response to combined movements relate to patterns of response to single movements? If the fSTS represents face movements in a parts-based fashion, responses to combined movements should reflect a combination of the responses separately evoked by the eye and mouth movements. In contrast, a holistic representation would predict that responses to combined movements cannot simply be decomposed into responses to parts. In order to assess the presence of parts-based and holistic representations in the fSTS, we generated "simulated" patterns of responses to combined movements, by finding an optimal linear combination of evoked responses to the corresponding eye and mouth movements, and asked to what extent these simulations predicted patterns of response to combined movements.

We found that patterns of response to combined movements could be discriminated by linear combinations of responses to individual eye and mouth movements (Figure 4), both when requiring generalization across position ($t[21] = 3.63$, $p < 10^{-3}$) and actor ($t(21) = 4.68$, $p < 10^{-4}$). This provides strong evidence for a parts-based representation of face movements in the fSTS.

Is there action information in fSTS responses to combined movements that cannot be captured by combinations of responses to single movements, pointing to holistic representations? To address this question, we asked whether measured patterns of response to combined movements do a better job of discriminating between the same patterns in left-out data than simulated patterns do. We found no difference between discrimination ability based on simulated or measured patterns, generalizing across position ($t(21) = -0.99$, $p = 0.83$) or actor ($t(21) = 1.45$, $p = 0.08$). Thus, our data do not provide evidence for holistic representations of face movements in the fSTS.

3.4 | Univariate analysis

The above results demonstrate that distinct face movements evoke different patterns of response in the fSTS. Do they also evoke different mean responses, or are these effects only measurable in the spatial patterns of the STS response? As an unplanned control analysis, we compared mean responses magnitudes to different actions. A one-way, two-level repeated measures ANOVA comparing responses to single and combined face movements revealed significantly stronger responses to combined movements (15% stronger responses to combined; $F[1,505] = 17.11$, $p < 10^{-4}$). Based on this difference, we subsequently looked for effects of action within single and combined movements. One-way repeated measures ANOVAs showed no effect of action condition on response magnitude, for either single movements ($F[7,147] = 0.95$, $p = 0.47$) or combined movements ($F[15,315] = 1.52$, $p = 0.10$). Thus, in contrast to pattern information, mean responses did not differentiate movement types, apart from the distinction between single and combined movements.

3.5 | Control ROI analyses

To what extent are the face movement representations reported above unique to the fSTS? We asked this question in two ways: by analyzing two early visual control ROIs, and by performing a whole-brain searchlight analysis. Unlike the fSTS, position-tolerant

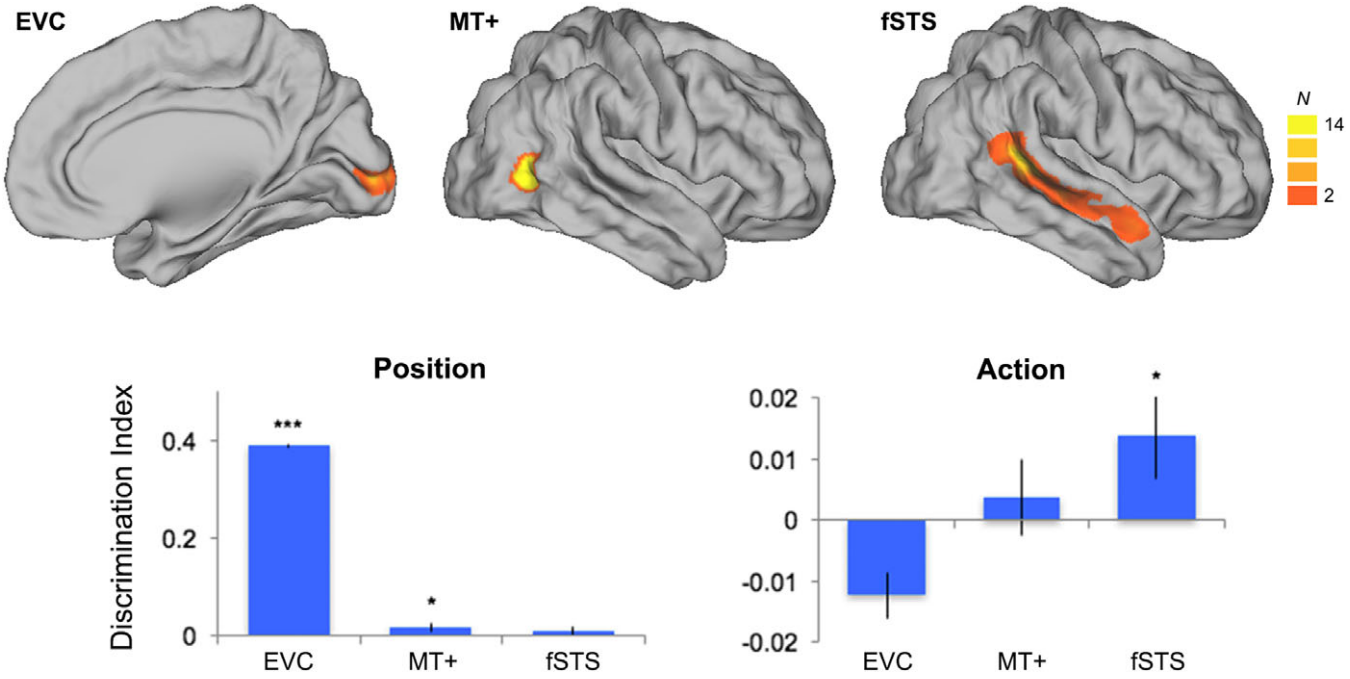

FIGURE 5 Upper: region of interest (ROI) locations, depicted as maps of the number of participants whose ROI included a given location. Lower: discrimination indices (correlation difference scores) for information about visual position (left), and action, generalizing across position (right). *Denotes $p < 0.05$ and *** denotes $p < 10^{-3}$

action information was not observed in EVC ($t[23] = -3.33$, $P \approx 1$) or MT+ ($t[23] = 0.57$, $p = 0.28$). We were able, however, to decode visual position from patterns of response in both EVC ($t[23] = 6.78$, $p < 10^{-6}$) and MT+ ($t[23] = 2.18$, $p < 0.05$), but not fSTS ($t(21) = 0.94$, $p = 0.18$), demonstrating that our approach was sufficiently sensitive to recover a well-established functional property of early visual regions. Split-half correlations between patterns of activity were generally high in all regions ($r = 0.88$ for fSTS; $r = 0.91$ for EVC; $r = 0.93$ for MT+; computed by averaging elements in the 4×4 matrix of split-half correlations across different visual positions).

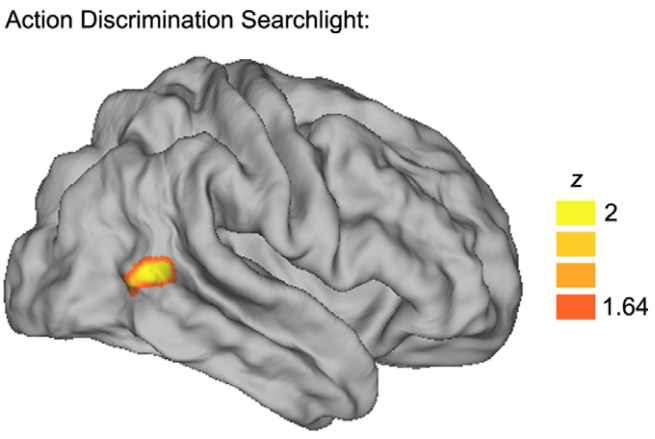

FIGURE 6 Searchlight analysis for position-tolerant action information. A significant effect indicates that patterns in an 8 mm-radius sphere around a given location contain information that discriminates perceived action, in a manner that generalizes across visual position. Thresholded at $p < 0.05$ voxel-wise, with an additional permutation-based cluster-wise threshold of $p < 0.05$ to correct for multiple comparisons across voxels

Would EVC or MT+ contain action information if generalization across visual position were not required? To test this, we ran an exploratory, unplanned analysis, testing for action information while generalizing across actor but not position. In this case, action information was observed in MT+ ($t[23] = 5.83$, $p < 10^{-5}$) but not EVC ($t[23] = 1.33$, $p = 0.20$). In other words, patterns of response in MT+ can discriminate face movements, but this discrimination is abolished when generalizing across small differences in visual position, pointing to a movement representation tied to retinotopic position, rather than an abstract representation of movement type. This result highlights the importance of requiring generalization across distinct stimulus conditions to infer higher level representations from decoding.

Do any brain regions outside of our planned ROIs contain pattern information that discriminates perceived face movements? To address this question, we performed a whole-brain searchlight analysis for position-tolerant action information. At our planned threshold ($p < 0.01$ voxel-wise, $p < 0.05$ cluster-wise), we did not observe any regions with significant decoding. To check whether any marginal effects could be observed, we additionally applied a threshold of $p < 0.05$ voxel-wise, $p < 0.05$ cluster-wise. In this analysis, we observed a single region in the right posterior STS and middle temporal gyrus (Figure 6). This region was nearby and overlapping with the location of the posterior STS face response, but was centered slightly posterior and inferior to the face response. Furthermore, a supplementary analysis analyzing face-responsive voxels in the posterior, middle, and anterior right STS found face movement information in the posterior, but neither middle not anterior regions (Supporting Information Figure S2). Thus, position-tolerant representations of perceived face movements appear to be particularly pronounced in the posterior STS and adjacent cortex.

4 | DISCUSSION

Our results demonstrate that the face-sensitive cortex in the STS (fSTS) represents face movements, in a manner that is robust to changes in actor and small changes in visual position. Such representations were not observed in earlier visual regions in the calcarine sulcus and lateral occipitotemporal cortex, where responses are not expected to be position-tolerant. Indeed, a search across the whole brain for position-tolerant action information revealed just a single region of posterior STS and middle temporal gyrus, roughly consistent with the location of fSTS. Action representations in fSTS were sufficiently fine-grained to discriminate subtle differences between specific eye motions (e.g., closing or rolling eyes) and specific mouth motions (e.g., a smile or frown). Finally, responses to combined eye and mouth movements could be well predicted by responses to the isolated eye and mouth movements, pointing to a parts-based representation of face movements. Taken together, these results indicate that fSTS contains a representation of the kinematics of face movements, which is sufficiently abstract to generalize across actor and across small variations in visual position, but which is nevertheless decomposable into the movements of separate face parts.

These results are consistent with prior findings of STS responses to perceived eye and mouth movements (Pelphrey, Morris, Michelich, Allison, & McCarthy, 2005; Puce, Allison, Bentin, Gore, & McCarthy, 1998), and extend these findings by identifying differences in response patterns to distinct types of motion. Strikingly, we find that linear combinations of fSTS responses to individual eye and mouth movements can be used to discriminate responses to combined movements, and can do so as well as responses to combined movements themselves in independent data. This is consistent with an underlying neural code in which responses to specific eye and mouth movements sum linearly, as has been argued for the coding of an object's shape and its color or material in macaque inferotemporal cortex (Köteles, De Maziere, Van Hulle, Orban, & Vogels, 2008; McMahon & Olson, 2009). This approach for assessing parts-based versus holistic processing could be equally well applied in other domains of cognitive neuroscience, for characterizing representations in perceptual as well as high-level cognitive domains (see also Baldassano, Beck, & Fei-Fei, 2016).

Evidence for parts-based coding was observed in spite of the fact that by design, many of the combined stimuli differed in terms of semantic interpretation from the individual movements they were composed of. For instance: a smile accompanied by an eyebrow raise appears happy or eager; accompanied by eyes closing, appears relaxed; and accompanied by a scowl, appears as an evil grin. Our behavioral study provided evidence that our stimuli were indeed perceived holistically, consistent with prior studies on holistic processing of facial expression (Calder et al., 2000; Flack et al., 2015). Thus, finding a parts-based representation of combined motions in fSTS is more consistent with a kinematic representation than a categorical representation of the interpretation of the movement.

Our results are consistent with several prior studies on fSTS responses to static images, that have been suggestive of a parts-based or kinematic representation. Liu, Harris, and Kanwisher (2010) found that the fSTS responded identically to intact face images as to images with spatially scrambled face parts. Using a composite effect paradigm, Flack et al. (2015) found that vertical alignment of top and bottom face halves did not influence the magnitude of release from adaptation of the fSTS response upon changing the expression of one face half. Harris et al. (2012) measured fSTS responses to face images defined along a morph continuum between two categorically perceived emotional expressions, and found that the fSTS response released from adaptation whenever there was a change in physical properties of the expression, regardless of whether perceived emotion differed. Similarly to our results, this indicates that the fSTS does not contain a categorical representation of perceived emotion, but a continuous representation of facial expression. Lastly, Srinivasan, Golomb, and Martinez (2016) found that spatial patterns of activity in an anatomically defined pSTS region could discriminate facial expressions containing different action units (facial muscle actions) more effectively than expressions from different emotion categories.

In interpreting others' face movements, we begin with a two-dimensional input on the retina, and are ultimately able to infer abstract social properties, such as others' mental or bodily states, from this visual input. The representation characterized in the present study appears to correspond to an intermediate stage in this inferential process: it is sufficiently abstract to generalize over low-level visual details, but still relates more to the properties of face motion itself than to their social interpretation. On this interpretation, where is social information from face movements extracted and represented? Prior evidence indicates that these processes involve both the STS and downstream regions. Watson et al. (2014) observed cross-modal adaptation in the right pSTS to the emotional content of faces and voices, pointing to a representation that generalizes beyond kinematic properties. Similarly, Peelen, Atkinson, and Vuilleumier (2010) found emotion information in a region of left pSTS/STG that generalized across dynamic face, body, and vocal inputs. The STS may thus contain both neural populations that represent face movements in a kinematic format, as well neural populations that encode a more general representation, pooled across multiple input modalities.

Prior evidence also suggests that other regions encode inferred social information with a higher degree of abstraction. The theory of mind network, a set of regions thought to be involved in the representation of mental states of others, provides a plausible candidate for the substrate of such a downstream representation (Fletcher et al., 1995; Saxe & Kanwisher, 2003). For instance, Skerry and Saxe (2014) found that the medial prefrontal cortex (mPFC), part of the theory of mind network, contained abstract emotion representations (of positive vs. negative valence), which generalized across emotions depicted from dynamic facial expressions to emotions inferred from animations of geometric shapes mimicking social interactions. In contrast, fSTS contained emotion representations within each domain that did not generalize across domains.

Our study has several limitations, which should be noted. First, although MVPA provides a powerful method for assessing the representational content of human brain regions, the method is intrinsically limited by the spatial resolution of fMRI. MVPA can only detect neural representations that are spatially organized at a scale that can be detected with the 2–3 mm resolution of fMRI. There are known

representations that lack such a spatial organization, such as representations of place in the hippocampus or representations of face identity in macaque face patches, which would not be detectable with MVPA (Dombeck, Harvey, Tian, Looger, & Tank, 2010; Dubois, de Berker, & Tsao, 2015). Thus, it is not valid to make strong negative claims from MVPA data. In particular, the lack of evidence for a holistic representation in our study does not imply that no such representation exists. Nevertheless, our data do provide positive evidence for the presence of a parts-based representation in fSTS.

Another potential limitation of the current study was the use of animated stimuli. We chose to use animated stimuli to ensure tight visual control over the stimuli, and so that combined movements would be exact combinations of individual eye and mouth motions, which was critical for the logic our analyses. However, the animated stimuli are somewhat nonnaturalistic, and might be less likely to evoke meaningful emotion attributions than real actors would be. We cannot rule out the possibility that holistic representations would be observed in response to naturalistic face movement stimuli. Thus, studies using video-recorded stimuli might be better suited for studying emotion representations in fSTS.

Lastly, while we tested whether face movement information in fSTS generalizes across two actors (one male and one female) and two visual positions (slightly to the left and to the right of fixation), we cannot say whether these representations would generalize over a wider range of actors and visual positions––for example, they may well not generalize to visual positions farther from the center of fixation. Thus, the term "generalization" as used in this report should be taken only to refer to the range of conditions used in this experiment; subsequent work will be needed to determine the full scope of generalization.

Our results point to a number of interesting directions for future research. If the fSTS primarily contains an intermediate representation of face movements, how does this region interact with other areas, such as the amygdala or mPFC, to support social inferences? Research on effective connectivity between these regions, or using combined TMS and fMRI to provide a causal manipulation, may be able to address this question. Beyond the dimensions considered in the present study, is the fSTS representation tolerant to other relevant dimensions, such as size, viewpoint, or larger changes in position? And lastly, if the fSTS representation is largely actor-invariant, corresponding to action type rather than an action-actor pairing, where does action information become associated with actor to form a representation of a specific agent's motion or implied internal state?

To conclude, the present research provides evidence that the fSTS represents the face movements of others, in a manner that is abstracted from low-level visual details, but tied to the kinematics of face part movements. Future research should further detail the nature of motion representations in the fSTS, and clarify the role of this region in the inferential process that takes us from raw visual input to socially meaningful inferences about other humans.

ACKNOWLEDGMENTS

This research was funded by grants from the David and Lucile Packard Foundation, National Institutes of Health (MH096914-01A1), and National Science Foundation (Center for Brains, Minds, and Machines, CCF-1231216) to R.S. B.D. was supported by a National Science Foundation graduate research fellowship. The authors declare no competing financial interests.

ORCID

Ben Deen https://orcid.org/0000-0001-6361-6329

REFERENCES

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*, 267–278.

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, *23*, 905–913.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2016). Human–object interactions are more than the sum of their parts. *Cerebral Cortex*, *27*, 2276–2288.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology. Human Perception and Performance*, *26*, 527–551.

Deen, B. (2015) FMVPA. Retrieved from osf.io/gqhk9.

Deen, B. (2016) FMVPA Behavioral. Retrieved from osf.io/mc7pd/.

Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L., & Tank, D. W. (2010). Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neuroscience*, *13*, 1433–1440.

Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *The Journal of Neuroscience*, *35*, 2791–2802.

Flack, T. R., Andrews, T. J., Hymers, M., Al-Mosaiwi, M., Marsden, S. P., Strachan, J. W., … Young, A. W. (2015). Responses in the right posterior superior temporal sulcus show a feature-based response to facial expression. *Cortex*, *69*, 14–23.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*, 109–128.

Goffaux, V., & Rossion, B. (2006). Faces are" spatial"—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology. Human Perception and Performance*, *32*, 1023–1039.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*, 63–72.

Harris, R. J., Young, A. W., & Andrews, T. J. (2012). Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. *Proceedings of the National Academy of Sciences*, *109*, 21164–21169.

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Shouten, J., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233.

Köteles, K., De Maziere, P. A., Van Hulle, M., Orban, G. A., & Vogels, R. (2008). Coding of images of materials by macaque inferior temporal cortical neurons. *The European Journal of Neuroscience*, *27*, 466–482.

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience*, *22*, 203–211.

Marchini, J. L., & Ripley, B. D. (2000). A new statistical approach to detecting significant activation in functional MRI. *NeuroImage*, *12*, 366–380.

McMahon, D. B., & Olson, C. R. (2009). Linearly additive shape and color signals in monkey inferotemporal cortex. *Journal of Neurophysiology*, *101*, 1867–1875.

Mondloch, C. J., & Maurer, D. (2008). The effect of face orientation on holistic processing. *Perception*, *37*, 1175–1186.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*, 2636–2643.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *The Journal of Neuroscience*, *30*, 10127–10134.

Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: An fMRI study of eye, mouth and hand movements. *Cerebral Cortex*, *15*, 1866–1876.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*, 2356–2363.

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*, 2188–2199.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*, 34–79.

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*, 11.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*, 1835–1842.

Schultz, J., Brockhaus, M., Bülthoff, H. H., & Pilz, K. S. (2013). What the human brain likes about facial motion. *Cerebral Cortex*, *23*, 1167–1178.

Skerry, A. E., & Saxe, R. (2014). A common neural code for perceived and inferred emotion. *The Journal of Neuroscience*, *34*, 15997–16008.

Srinivasan, R., Golomb, J. D., & Martinez, A. M. (2016). A neural basis of facial action recognition in humans. *The Journal of Neuroscience*, *36*, 4434–4442.

Tobin, A., Favelle, S., & Palermo, R. (2016). Dynamic facial expressions are processed holistically, but not more holistically than static facial expressions. *Cognition and Emotion*, *30*, 1208–1221.

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. *The Journal of Neuroscience*, *34*, 6813–6821.

Winston, J. S., Henson, R., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, *92*, 1830–1839.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, *14*, 1370–1386.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747–759.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Deen B, Saxe R. Parts-based representations of perceived face movements in the superior temporal sulcus. *Hum Brain Mapp*. 2019;1–12. https://doi.org/10.1002/hbm.24540