CORTEX 197 (2026) 193–210

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Invariant versus context-specific representation of face shape and motion in the face network

S. Sanaz Hosseini [a] and Fabian A. Soto [b,*]

[a] Department of Psychiatry, University of Cambridge, Cambridge, UK
[b] Department of Psychology, Florida International University, Miami, FL, USA

ARTICLE INFO

Article history:
Received 10 July 2025
Revised 6 February 2026
Accepted 6 February 2026
Action editor D. Samuel Schwarzkopf
Published online 13 February 2026

Keywords:
Face perception
fMRI
Decoding
MVPA
Face shape
Face motion
Neural representation

ABSTRACT

A growing consensus suggests that separate brain pathways process facial shape and motion. However, a region may respond preferentially to one type of information (e.g., shape) while still being modulated by another (e.g., motion), similar to modulatory effects observed in non-classical receptive fields of visual neurons. To test this possibility in the face network, we applied a two-test strategy—combining cross-decoding and context-sensitivity analyses—to determine whether face-selective areas encode shape and motion in an invariant or context-specific manner. Twelve participants viewed videos generated from 3D synthetic face models in which facial shape and motion were manipulated independently. We report four key findings. First, shape and motion information could be decoded from all face-selective regions, suggesting overlapping encoding of shape and motion. Some, but not all, findings can be accommodated by models proposing separate pathways. Second, we observed distinct invariance patterns: OFA showed motion-specific shape representations, IFG encoded shape invariant to motion, and FFA encoded both shape and motion invariantly—challenging models that restrict motion processing to dorsal areas. Third, including motion involving different facial poses—in which motion and shape are confounded—inflated invariance estimates across the network, highlighting the importance of controlling for pose-motion confounds. Fourth, individual-level analyses revealed substantial variability. These results call for revised models incorporating representational overlap, context sensitivity, and individual variability.

1. Introduction

Face recognition is undoubtedly fundamental to social interaction. Humans rely heavily on facial cues to guide behavior and support effective communication. Even a brief glimpse of a face can convey a wealth of social information—including a person's identity, emotional state, gaze direction, age, gender, and ethnicity—as well as more abstract social inferences such

* Corresponding author.
E-mail address: fasoto@fiu.edu (F.A. Soto).
https://doi.org/10.1016/j.cortex.2026.02.006

as attractiveness, dominance, and trustworthiness (Little et al., 2011; Mattavelli et al., 2012; Quadflieg et al., 2012).

Neuroimaging studies have revealed multiple clusters of face-selective voxels distributed across the human cortex, including in the fusiform gyrus—known as the fusiform face area (FFA)—the occipital gyrus—home to the occipital face area (OFA)—and the posterior superior temporal sulcus (pSTS), as well as more anterior regions such as the anterior inferior temporal lobe (aIT), anterior superior temporal sulcus (aSTS), and inferior frontal gyrus (IFG) (e.g., Duchaine & Yovel, 2015; Fox et al., 2009). Several models have been proposed to explain the functional organization of this network. The most influential framework, introduced by Haxby et al. (2000), posits a division of labor between two parallel pathways within a core face-processing system: OFA serves as an initial processing stage that feeds into both FFA and pSTS. In this model, FFA primarily encodes invariant facial properties such as identity, while pSTS processes dynamic, changeable aspects such as expression and gaze.

A revised account by O'Toole et al. (2002) emphasized the role of motion-sensitive neurons in the middle temporal area (MT), proposing that dynamic information may reach pSTS via inputs from MT, rather than being routed through OFA. This view has gained empirical support from studies by Sliwinska et al. (2020) and Pitcher and Ungerleider (2021), which demonstrate that pSTS receives direct input from early visual cortex independently of OFA.

Pitcher et al. (2014) found evidence supporting the presence of distinct pathways for processing dynamic and static face stimuli that were not consistent with any of the previously proposed models. To account for this and a variety of other results, Bernstein and Yovel (2015) proposed a model in which the ventral visual pathway—including OFA and FFA—primarily processes shape-related information, while the dorsal pathway—including MT, pSTS, aSTS, and IFG—is specialized for motion processing. This is an important departure from the Haxby et al. (2000) model, as it posits that the ventral pathway processes not only the morphological shape features that determine identity and other invariant aspects of faces (hereafter referred to as *facial shape*), but also the static configurations of facial muscles that convey expressions (hereafter referred to as *facial pose*). Both facial morphology and muscle activity contribute to the overall static shape of a face at any given moment. However, when those muscles move, they generate *facial motion* that is processed in the dorsal pathway, along with other dynamic cues such as head movements and gaze shifts. Crucially, this model suggests that both pathways are engaged by both static and dynamic stimuli, extracting structural (ventral) and dynamic (dorsal) information regardless of whether actual motion is present. According to this view, dynamic information can be conveyed even in static images through implied motion—that is, the perceptual suggestion of movement in the absence of actual motion (Marian & Shimamura, 2013; Senior et al., 2020; Yoshikawa & Sato, 2008). In sum, Bernstein and Yovel (2015) propose that *facial shape* is processed primarily in the ventral pathway, *facial motion* is processed in the dorsal pathway, and *facial pose* engages both pathways. Because most facial motion involves changes in facial pose, testing this hypothesis requires experimental designs that manipulate facial motion while keeping the underlying poses constant.

Liu et al. (2021) conducted a meta-analysis of fMRI activation studies that was generally in line with the Bernstein and Yovel model, revealing that facial pose mainly activates the ventral pathway (OFA, FFA, and IFG) and facial motion activates the dorsal pathway (MT and STS). However, some of their findings were more nuanced than what the model proposes. For example, STS had higher levels of activity for facial motion compared to facial pose involving the same expression. This could be consistent with the model, if one assumes that the motion information implied by a facial pose may not be as effective to drive STS activity as real motion information. They also found that FFA shows a similar level of activation for both facial pose and motion during expressions. Finally, recent work by Wang et al. (2020) expands the model of Bernstein and Yovel, by suggesting the existence of a medial pathway (posterior cingulate cortex, amygdala, and orbitofrontal cortex), mainly processing the social and emotional significance of a face, in addition to the two pathways proposed by Bernstein and Yovel.

The Bernstein and Yovel (2015) model was developed to account for a substantial body of evidence, primarily from fMRI activation and adaptation studies. However, a brain region may respond preferentially—or even exclusively—to one type of facial information (e.g., shape), while still having its activity modulated by another type (e.g., motion), akin to modulatory effects seen in non-classical receptive fields of visual neurons (Spillmann et al., 2015). This raises the broader question of whether representations in the ventral and dorsal visual streams are invariant or context-dependent. Such modulatory interactions, either between pathways or within each, may help explain behavioral findings: that shape perception improves when motion is incorporated (Dobs et al., 2018; Krumhuber et al., 2013; Xiao et al., 2014); that expectations about natural facial motion depend on face shape (Tordjman et al., 2026); and that shape and motion representations are perceptually correlated (Martin et al., 2024).

In neuroimaging research, cross-decoding is a widely used method for assessing representational invariance. An illustration of the procedure is shown in Fig. 1a. In this method, a classifier is trained to distinguish between conditions (e.g., two facial shapes) based on fMRI activity patterns obtained in one context (e.g., a specific facial motion). The trained classifier is then tested on the same conditions presented both in the original context and in a different context (e.g., a different motion).

In Fig. 1a, fMRI activity patterns across two voxels are plotted as coordinates in a two-dimensional space. The two conditions to be discriminated are represented by symbols with different colors and shapes (red "o" and blue "x"), and the dotted line indicates the linear boundary used by the classifier to assign each activity pattern to one condition or the other (i.e., which facial shape produced the pattern). When the classifier is tested on new presentations of the same stimuli in the training context (i.e., the original motion), its performance typically declines somewhat but remains above chance. This above-chance decoding is interpreted as evidence that neural representations within the selected voxels carry information about the experimental conditions (i.e., facial shape) and is shown in the rightmost panel of Fig. 1a as a black bar above chance level.

The cross-decoding test goes one step further by evaluating the same classifier on new presentations of the same

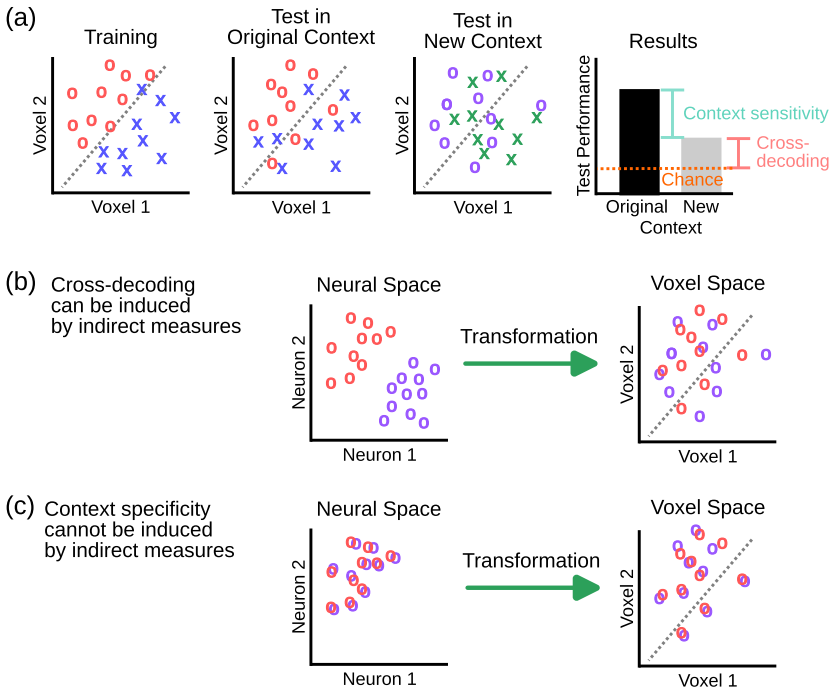

Fig. 1 – Schematic representation of the tests used in this study and their relation to underlying neural representations. (a) Illustration of the cross-decoding and context-sensitivity tests used in this study. Activity patterns across two voxels are represented by symbols with different colors and shapes (red "o" and blue "x"), where color and shape denote the conditions to be discriminated. The dotted line indicates the linear boundary used by the classifier to assign each activity pattern to one condition or the other (i.e., which facial shape produced the pattern). The classifier is trained with data obtained in one context (leftmost panel) and tested with data from both the same context (second panel) and a novel context (third panel). The rightmost panel illustrates classifier performance for new data in the two contexts, where bar heights represent decoding accuracy and the dashed line indicates chance level. The cross-decoding test assesses whether performance is above chance in the new context (pink brackets), whereas the context-sensitivity test assesses whether performance decreases from the training to the new context (cyan brackets). (b) Intuitive explanation of how the cross-decoding test can yield false-positive evidence for invariance when applied in isolation. When indirect measures of neural activity are obtained from fMRI, the transformation from neural space to voxel space (green arrow) can alter the geometry of the underlying representations, creating spurious similarity between conditions. (c) Illustration showing why the context-sensitivity test is less susceptible to such biases. In this case, the transformation from neural space to voxel space (green arrow) cannot make already similar neural representations appear dissimilar at the voxel level.

conditions (i.e., facial shapes) in a different context (i.e., a new facial motion). The corresponding activity patterns are depicted in Fig. 1a with the same symbols but in new colors (purple "o" and green "x"). In this case, the classifier performs slightly worse than before but still above chance, as indicated by the pink brackets in the rightmost panel of Fig. 1a. Such above-chance performance in the new context—cross-decoding—is taken as evidence that the neural representations underlying the decoded variable (e.g., shape) are at least partially shared or invariant across contexts (e.g., motion). Using cross-decoding, it has been found that the STS contains information about face motion (Deen & Saxe, 2019) and valence (Zhang et al., 2016) that is invariant across facial shape, and that FFA and aIT contain information about facial shape that is invariant across expressions (Zhang et al., 2016).

However, recent work has shown that cross-decoding can yield *false positives*—apparent invariance that arises not from true invariance in the neural code, but from the way neural activity is transformed into fMRI signals (Sandhaeger & Siegel, 2023; Soto et al., 2018; Soto & Narasiwodeyar, 2023). This issue is illustrated in Fig. 1b, where activity patterns for only one of the two conditions (e.g., one shape) are shown in both the original and new contexts (red and purple "o," respectively). fMRI activity patterns are only indirect measures of the underlying neural representations; that is, the neural representations of interest undergo a transformation before they can be observed in voxel space (represented by the green arrow in Fig. 1b). As depicted in Fig. 1b, even when the true neural representations are *context-dependent*—such that neural activity patterns for the two contexts do not overlap—the transformation from neural space to voxel space can make them appear to overlap. This results in *false-positive invariance*: above-chance cross-decoding performance that reflects properties of the measurement transformation rather than genuine invariance in neural representations. Importantly, this problem is not merely theoretical but is likely to be

widespread in neuroimaging studies (see Soto & Narasiwodeyar, 2023).

To address this limitation, cross-decoding should be complemented with a test of *context sensitivity*. This test uses the same data as the cross-decoding analysis but focuses on whether decoding accuracy significantly decreases when context changes, as illustrated by the cyan brackets in the rightmost panel of Fig. 1a. The key motivation for including this test is to provide a safeguard against the bias toward false-positive invariance inherent in cross-decoding. As illustrated in Fig. 1c, when a representation is truly invariant—so that neural activity patterns in the two contexts completely overlap—the transformation from neural to voxel space cannot artificially separate them and produce apparent context sensitivity. In summary, although significant cross-decoding does not guarantee true invariance in the underlying neural representations, a non-significant context sensitivity test strengthens the case for invariance.

In real data, neural representations are likely to lie somewhere along a continuum between complete invariance and full context sensitivity. Determining exactly where a representation falls on this continuum is unlikely to be feasible with fMRI data alone (see Soto & Narasiwodeyar, 2023). Thus, the most informative approach is to apply both tests jointly and interpret their results together to assess whether the weight of the evidence supports invariance or context sensitivity. Only when one of the two tests yields a significant result—providing evidence exclusively for either invariance or context sensitivity—can we validly conclude that the representations encoded in a given brain region possess that property (Soto & Narasiwodeyar, 2023). Valid interpretations for running these tests jointly are summarized in Table 1.

Here, we applied this two-test strategy to investigate whether neural representations of face shape and motion in face-selective cortical areas (OFA, FFA, STS, aIT, and IFG) are invariant or context-sensitive. To independently manipulate shape and motion while maintaining precise control, we used synthetic face videos in which identical shapes could be animated with different motions, and identical motions could be applied to different shapes (Anzellotti and Caramazza, 2014; Hosseini and Soto, 2024).

Our experimental design, illustrated schematically in Fig. 2, included three distinct facial shapes (rows in the figure), each paired with three distinct motion patterns (columns in the figure). The motion patterns involved transitions from one facial pose to another (linear in the space of pose parameters). Both the facial shapes and poses were selected from previously validated face models (24 shape models representing different identities and six models of basic emotional expressions; see Hays et al., 2020) to ensure that the stimuli were as distinct as possible.[1] For each video included in the study—represented by a film icon in Fig. 2—the figure presents the first and last frames, which together summarize the direction of motion. As discussed above, testing the interaction between facial shape and motion requires experimental designs that manipulate motion while keeping the underlying facial poses constant.

This requirement is met by the conditions outlined by the red boundaries in Fig. 2, where motion is varied while the facial poses remain identical. Specifically, these conditions depict the same two expression endpoints (from slightly anti-surprise to surprise) but differ in the dynamic trajectory connecting them (e.g., reversed sequences). In other words, the conditions within the red boundaries contain the exact same facial poses—the same frames—but presented in different temporal orders. By focusing on these conditions, we can more precisely assess how motion per se is represented in the brain, independent of changes in facial pose.

The additional column on the right introduces a motion sequence that also changes the expression pose (from slightly anti-disgust to disgust). This enables a complementary analysis of how the brain responds when motion cues alter the structural properties of the face, distinguishing effects driven purely by motion from those that are confounded with changes in pose—a type of facial shape variation distinct from the morphological shape differences manipulated across rows in Fig. 2.

By analyzing these conditions using the combined cross-decoding and context-sensitivity approach, we aimed to draw more robust conclusions about the invariance or specificity of neural coding, while minimizing the risk of false-positive findings.

2. Materials and methods

All aspects of this study methods and analyses were preregistered (https://osf.io/snrm4), with the following exceptions: (1) addition of group-level statistical tests to facilitate interpretation of results, and (2) we do not report the results of the decoding separability test (Soto & Narasiwodeyar, 2023) as we have recently discovered that our implementation might be biased toward rejection of the null.

2.1. *Participants*

A total of 12 participants (9 female) with an average age of 26.25 years (range 21-33 years old) were recruited for the study. In designing our experiment, we favored collection of a large amount of data per participant (about 3 h of scanning) rather than a large number of participants, to present results of both individual-level and group-level analyses (see below).

All participants had normal or corrected-to-normal vision and were right-handed. The sample was drawn from the local community at Florida International University and recruited through online advertisements or word-of-mouth. Participants received monetary compensation for their time and effort at a rate of $20/h for fMRI sessions and $10/h for the behavioral session. Each participant completed a total of 3 h of scanning sessions and half an hour of a behavioral session prior to the first scan.

[1] One reason for this choice is that controlled 3D face models tend to appear relatively similar: they share external structural features such as head and ear shape, and they also have identical eye color and skin texture (the latter held constant to avoid confounding face shape with texture differences). In addition, these models lack individualizing features such as hair, which typically aid identity perception. To increase the perceived distinctiveness of the shape models, participants were familiarized with them before testing (see Methods section).

Table 1 – Summary of conclusions made based on conducting cross-decoding and context sensitivity tests jointly. The conclusions refer to how information about a target variable is encoded in a given brain region. The tests assume that there is evidence for the target variable being encoded in the region, which can be shown by significant decoding or cross-decoding of the variable from voxel activity in that region.

		Context Sensitivity	
		Significant	Not Significant
Cross-Decoding	Significant	Inconclusive	Invariant
	Not Significant	Context-Sensitive	Inconclusive

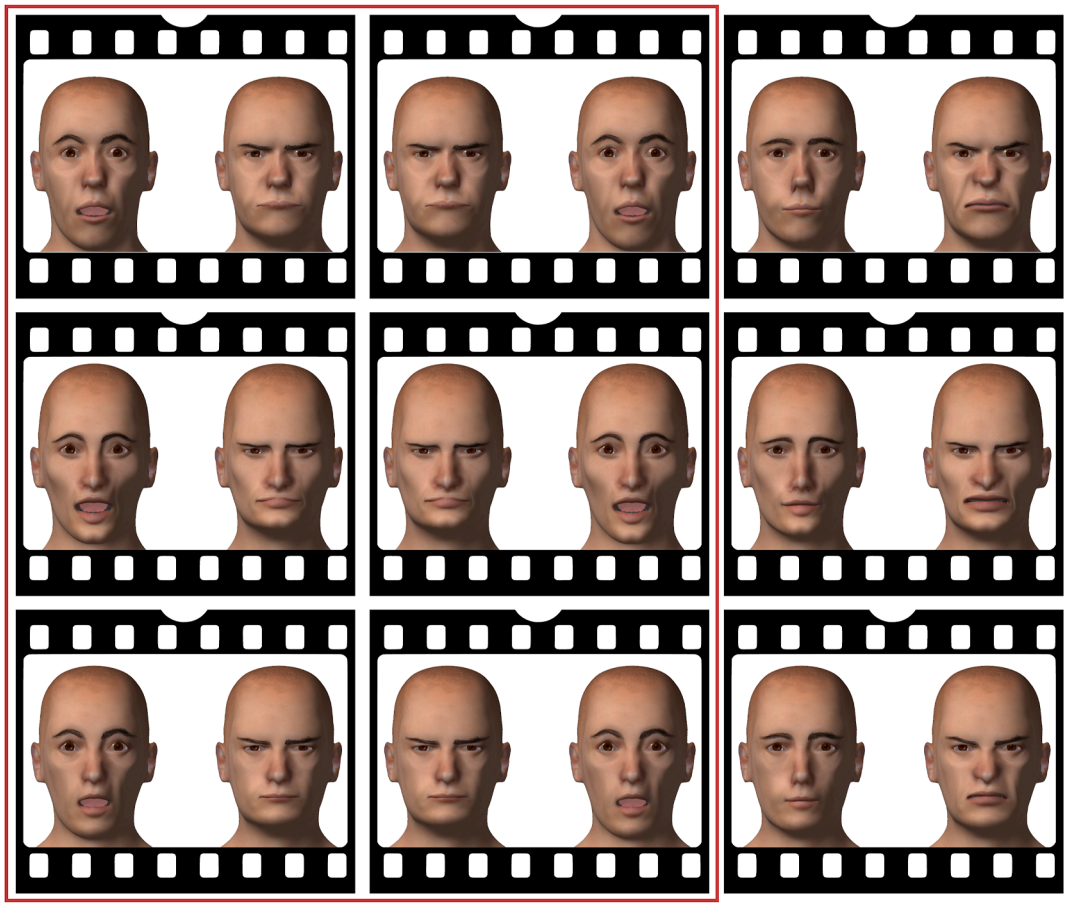

Fig. 2 – Schematic representation of the dynamic stimuli used in this study. Each video is represented by a film icon showing the first and last frame of the dynamic face stimulus (i.e., the starting and ending facial poses). The rows correspond to three different facial shapes, and the columns correspond to three different motion patterns. The conditions outlined by the red rectangle illustrate the key comparison: two motion sequences that differ in their dynamic trajectories but involve the same facial expression poses (from slightly anti-surprise to surprise).

Inclusion criteria included ages between 18 to 35 years old, English-speaker, normal or corrected-to-normal vision, right-handedness, and weighting under 300 pounds. Exclusion criteria included a history of neurological or major medical condition affecting brain function, potential difficulties comprehending study procedures, currently sick or recovering from an illness, not suitable to undergo MRI scanning (due to claustrophobia, implanted devices/metal, pregnancy, color tattoo, color contacts). Participants completed two screening questionnaires in order to assess their health status as well as to ensure they met the inclusion criteria. The study was approved by the institutional review board and all participants provided written informed consent for both behavioral and fMRI sessions.

2.2. Experimental methods

2.2.1. Stimuli

Three-dimensional models of face shape and expression pose were created with MakeHuman (http://www.makehumancommunity.org) and FaReT (Hays et al., 2020), which is a suite of plugins developed to create precisely-controlled face stimuli.

Three face shape models and two face expression pose models (surprise and disgust) were used to create the stimuli. Shape and pose models were selected from those available in FaReT to maximize their distance from one another in the space of model parameters. In addition to using expression pose models, anti-expression pose models were also used. An anti-expression is the direction in FaReT space pointing opposite to its corresponding expression, with reference to neutral (the average of an expression and its corresponding anti-expression pose models will result in neutral). Each dynamic face stimulus was created by concatenating multiple static images into a 1-s video with thirty frames per second. The images were rendered using FaReT's interpolation rendering plugin (Hays et al., 2020), in a sequence going from an expression pose to its corresponding anti-expression pose, which was concatenated and converted into an AVI video.

The final dynamic stimuli were the combination of the three face shapes and three face motion conditions (see Fig. 2). The motion conditions were created by manipulating the pose models in the following pairs: surprise (100%) to anti-surprise (50%), anti-surprise to surprise, and anti-disgust (100%) to disgust (100%). Fifty percent anti-surprise was used because extrapolation beyond this point produced unrealistic renders. This manipulation resulted in a total of 9 stimulus conditions, 3 facial shapes displaying 3 distinct facial motions.

It should be noted that the first two motions ("changing pose from surprise to anti-surprise" and "changing pose from anti-surprise to surprise") include the exact same poses but opposite direction of motion. This provided us with stimuli that changed in motion independent of pose. Face shape models only differed in the shape of inner facial features (i.e., nose, mouth, eyes, forehead, chin, and cheeks) while outer facial features (i.e., head shape, neck, and ears) did not change across conditions. Note that changes in shape features are large enough to be perceived as changes in facial identity.

2.2.2. Procedure

We performed an fMRI experiment with an event-related task design. In designing our experiment, we favored collection of a large amount of data per participant rather than a large number of participants, to present results of both individual-level and group-level analyses, explained below.

Data collection was completed at the Center for Imaging Science (CIS), Florida International University. The experiment consisted of three sessions. During the first session outside the MRI scanner, participants completed a familiarization task. In the second session inside the MRI scanner, participants completed a face localizer and the main task. In the final session, participants completed the main task. These tasks are described below.

Familiarization Task. The first session was conducted outside the scanner and involved a short familiarization task lasting for about half an hour. During familiarization, participants were tasked with learning the face stimuli and names assigned to the different shape models. After receiving instructions on the screen, participants were presented with 30-s animated movies of each shape model changing in viewpoint and expression, with the assigned name on the screen. Participants were asked to memorize each face and its assigned name as they would be tested on recognizing them later. The animated movies were presented 2 times, with the order of presentation of shape models randomized. This was followed by a one-back task that required participants to respond whether the current stimulus (on the screen) was the same or different from the one presented in the previous trial. Participants were presented with 1-s-long animated faces and they were required to press a key whenever they saw a face that was repeated back-to-back. Each of the 9 videos described in the Stimulus section was presented 15 times. Participants were provided with feedback whenever they provided a key response ("CORRECT: That was a repeated face!" or "INCORRECT: That was NOT a repeated face!") and whenever they missed reporting a repeated stimulus ("INCORRECT: You missed a repeated face!"). Feedback was displayed on the screen for 2 sec. Because performance during this familiarization procedure was consistently near ceiling in pilot testing and indicated that the task imposed minimal difficulty, no performance criterion was applied to determine eligibility to proceed to the main study.

Localizer Task. The localizer was performed in order to enable us to functionally localize face-selective regions of interest. Participants were shown a series of videos (as proposed by Fox et al. (2009)) of faces and objects in a block design; each block presenting either faces or objects (Fox et al., 2009). Participants were tasked with performing a one-back repetition detection task, where they were instructed to press a button when they saw two consecutive stimuli (faces or objects) that were identical. Each stimulus (video) was 1 sec long and no feedback was provided. Participants completed 32 blocks, with 16 blocks featuring the face stimuli and 16 blocks featuring object stimuli. A cross-hair (fixation cross) was presented during the inter-block intervals. The experiment lasted approximately 10 min.

Main Task. The main task involved asking participants to complete the same one-back task as in the Familiarization task. Each stimulus was presented 126 times in total, separated into six runs (nine repetitions of each stimulus in each run) during the first MRI session and eight runs during the second MRI session. Each trial started with a .5 sec fixation cross, followed by the presentation of a dynamic stimulus for 1.1 sec, and then a randomly selected inter-stimulus interval (ISI) of 3.4, 5.4, or 7.4 sec. These ISIs were selected to optimize single-trial estimates (Abdulrahman & Henson, 2016; Grootswagers et al., 2017; Nili et al., 2014) for decoding (which require long ISIs) while maximizing the number of trials per session (which requires short ISIs).

2.2.3. MRI

Imaging was performed with a Siemens Magnetom Prisma 3T whole-body MRI system. A volume radio-frequency coil (transmit) and a 32-channel receive array were used to acquire both functional and anatomical images. During the first session, a high-resolution 3D anatomical T1-weighted volume (MPRAGE; TR = 2.4 sec; TI = 1.18 sec; TE = 4.13 msec; flip angle = 8°; voxel size = 111 mm^3; FOV = 256 mm; 176 sagittal slices, encoding direction from anterior to posterior) was obtained for each participant, which served as the reference volume to align functional images. During the main experiment, functional images were collected using a T2*-weighted EPI sequence (TR = 1.5 sec; TE = 30 msec; flip angle = 52°; sensitivity encoding with an acceleration factor of 4). We collected 60 slices with 30° angulation from the transversal plane to the coronal plane (near transversal), voxel size of 2.42.42.4 mm^3, and FOV of 219 mm.

2.3. Analysis

2.3.1. Defining fROIs

The functional localizer data was used to functionally localize the face-selective areas of interest. Preprocessing was performed using FSL (Jenkinson et al., 2012) and included automatic brain extraction (skull stripping) (Smith, 2002), followed by functionally registering runs to the high-resolution anatomical space of the subject using full brain search with the boundary-based registration (BBR) method (Greve & Fischl, 2009), correcting for acquisition timing delays using interleaved slice-timing (Sladky et al., 2011), correcting for motion using MCFLIRT (Jenkinson et al., 2002), spatial smoothing (5 mm) in order to reduce the noise levels without reducing the valid activation in each volume (Smith & Brady, 1997), and filtering using a high-pass temporal filter to remove the low-frequency temporal artifacts including scanner's slow drift, which has been shown to improve model fitting (Friston et al., 2000).

A whole-brain analysis to compare neural activity during face blocks to that during object blocks was carried out using FEAT (fMRI Expert Analysis Tool) Version 6.00, part of FSL. Statistical analysis of activation maps was performed to identify voxels with significantly greater activity ($p < .01$, uncorrected) for face blocks compared to the object blocks. Significant activation was observed in several brain regions, including the bilateral occipital lobes, fusiform gyrus, and temporal gyrus. To identify the fROIs, the Group-Constrained Subject-Specific (GSS) algorithm was used by masking activation maps with group face parcels (Julian et al., 2012); the group masks were obtained from https://web.mit.edu/bcs/nklab/GSS.shtml and transformed from MNI space to each participant's anatomical space using a nonlinear transformation with FNIRT. The regions reported by Julian et al. (2012) included OFA, FFA, and STS. To obtain aIT and IFG masks, we used the Harvard–Oxford Cortical Structural Atlas (Desikan et al., 2006). For aIT, a mask was obtained by including all voxels in the atlas that had a higher probability to include the anterior division of the temporal gyrus than the combined probabilities of the following regions: posterior division of the inferior temporal gyrus, temporooccipital part of the inferior temporal gyrus, anterior division of the temporal fusiform gyrus, posterior division of the temporal fusiform gyrus, anterior division of the middle temporal gyrus, posterior division of the middle temporal gyrus, and the temporal pole. Similarly, for IFG, a mask was obtained by including all voxels in the atlas that had a higher probability to include the pars triangularis or pars opercularis of inferior frontal gyrus than the combined probabilities of the following regions: frontal orbital cortex, frontal pole, frontal operculum cortex, middle frontal gyrus, precentral gyrus, central operculum cortex, and temporal pole. The obtained aIT and IFG masks were non-linearly transformed from MNI space to the individuals' anatomical space and the final fROIs were the overlap between them and the significant activation maps derived from the localizer task.

All fROIs were obtained separately for each hemisphere, and later combined into a single fROI mask.

2.3.2. Preprocessing

For quality control, the time-series for a run with 20% or more outliers were excluded. The frame displacement (FD) measure included in FSL with a threshold of 1 mm was used to detect motion outliers; less than 15% of the total volumes per run for all participants were marked as outliers. Therefore, all runs were included for further analysis.

Preprocessing of runs of the main task obtained from different MRI sessions were performed independently. The first six volumes (9 sec) in each run were discarded to allow T1 magnetization to reach steady-state. Moreover, the last 4 volumes (6 sec) in each run were discarded as well. The rest of the preprocessing included automatic brain extraction (skull stripping), registration, and motion correction using MCFLIRT, interleaved slice-timing correction, and high-pass temporal filtering. No spatial smoothing was performed, to preserve fine patterns across voxels. Nipype Python wrappers for FSL were utilized to preprocess and analyze the data (Gorgolewski et al., 2011; Jenkinson et al., 2012). All functional runs of a session for a given subject were then aligned to an averaged volume of the first functional run for the same subject and session. The aligned time-series for each session were concatenated into a single time-series file and masked with the fROI masks for further processing. The obtained time-series were then de-trended using a Savitzky–Golay filter with a polynomial order of 3 and a window length of 81s (Pedregosa et al., 2015).

2.3.3. Deconvolution

Single-trial activity estimates (i.e., betas) were obtained via a data-driven technique in which deconvolved neural activation values and a model of the hemodynamic response function (HRF) are estimated jointly (Pedregosa et al., 2015). Unlike other methods that hold the shape of the HRF constant across voxels, this technique allows the shape of the HRF to be different in each voxel, resulting in more accurate activity estimates. The model is implemented via the *hrf_estimation* Python package (https://pypi.org/project/hrf_estimation/), using a basis function with 3 elements to model the HRF and a Rank-1 General Linear Model with separate designs to perform the deconvolution.

2.3.4. Decoding analyses

To decode stimulus based on voxel activity patterns, we used a Nu-support vector classifier (NuSVC) with a linear kernel, implemented via the Python package scikit-learn (Pedregosa et al., 2011). The deconvolved neural activities obtained from fROI voxels were inputs to the classifier, while trial-specific stimulus values (shape or motion) were provided as labels. To decode a target dimension (e.g., shape), separate classifiers corresponding to each level of the context dimension (e.g., motion) were employed. Each classifier was trained to decode the target dimension (e.g., shape) using only trials in which a specific level of the context dimension (e.g., motion-1) was presented. The trained classifier was then tested with data collected from independent test runs at all different levels of the context dimension (e.g., motion-1, motion-2, and motion-3). This provided us with an accuracy estimate at the training level of the context dimension as well as all other levels.

The data were normalized by voxel using the *RobustScaler* function from *sklearn*, which takes each voxel activation value, subtracts the voxel median and divides by the interquartile range. After normalization, the classifier was trained with a nested cross-validation method where the hyperparameter Nu of the NuSVC model was optimized through leave-one-run-out cross-validation of data from the training runs. The Nu parameter was varied from 0 to 1 in 20 steps, and the Nu resulting in the highest validated accuracy within the training data was used to train a final classifier on all the training data. This classifier was then tested with data from the left-out testing run. This was completed for all runs and the obtained accuracies were averaged.

These procedures were applied to decode shape as the target dimension with motion as a context face dimension, and to decode motion as the target dimension with shape as a context face dimension.

We performed two different decoding analysis. In the main analysis, we included data from only trials in which motion did not produce a change in facial pose (inside the red boundaries in Fig. 2). This allowed us to investigate how motion and shape interact without the confounding influence of facial pose (which, just as morphological shape, momentarily changes the shape of the face surface). In the secondary analysis, we included all data, including trials in which motion did produce a change in facial pose. This allowed us to investigate how the inclusion of this confound, common in the literature, may change our main conclusions.

2.3.5. Decoding tests

Two fMRI decoding tests were used, cross-decoding (also known as cross-classification) and context sensitivity. These tests assess different possibilities for decoding a target dimension in different contexts. Cross-decoding tests against the null of context-specificity (i.e., providing evidence "for" invariance), and context sensitivity tests against the null of context-invariance (i.e., providing evidence "for" context sensitivity). Most previous research used cross-decoding alone to conclude in favor of invariant encoding, but it has been shown that performing both tests leads to more valid conclusions (Soto & Narasiwodeyar, 2023). Valid interpretations for running these tests jointly are summarized in Table 2.

Cross-Decoding Test. For cross-decoding, a decoder trained on a specific level of the context dimension was tested with data from all different levels of the context dimension and the classification accuracy for each level of the context dimension was recorded. Binomial tests were used to assess whether the decoding accuracy for data at any of the levels of the context dimension was above chance, using the Holm-Sidak method to correct for multiple comparisons. Significant accuracy for a level of the context that is not the same as the training level is evidence against the null of context-specificity, suggesting invariant encoding of the target. For example, if the decoder was trained to classify shapes with motion-1, then the decoder would be tested on its accuracy to classify identities in the test data with motion-1, motion-2, and motion-3. If above chance classification for shapes with either motion-2 or motion-3 was achieved, then the test would conclude that representation of shape is invariant from motion (or, more accurately, it is not context-specific). Both significant decoding or cross-decoding are interpreted as evidence that information about the target dimension was encoded in a given fROI.

We also performed a group-level cross-decoding test. A one-sample *t*-test (with Holm-Sidak correction for multiple comparisons) was conducted to determine whether the mean decoding accuracy at each level of the context dimension differed from chance. Above-chance decoding after a change from the training context provides evidence against the null of context-specificity at the group level.

Context Sensitivity Test. The context sensitivity test evaluates whether the probability of correct classification of the target dimension is stable across changes in the levels of the context dimension. A significant violation of context sensitivity provides strong evidence against invariant encoding. An omnibus chi-squared test was used to evaluate whether the decoder's classification accuracy varied across different levels of the context dimension.

We also performed a group-level context sensitivity test. A within-subjects ANOVA was conducted to consider the effect of different levels of the context dimension on the decoding accuracy across participants. A significant effect would provide evidence against the null of context-invariance at the group level.

Group Consistency Test. To summarize the results of individual-level tests, for each fROI we classified participants who provided evidence of accurate decoding (i.e., significant decoding or cross-decoding; total number of participants represented by n) into three groups (Table 2): inconclusive results, invariant decoding, and context-specific decoding. We tested whether the number of participants in the largest group (k) was significantly larger than chance (.33) using a binomial test, to determine whether there was a consistent result across individuals.

3. Results

Given the extensive range of results obtained, here we focus on reporting results from the group tests and the group consistency test (i.e., results of the individual-level analyses that were consistent across participants). Comprehensive details of the individual-level analyses, including statistical tests are provided in the Supplementary Material.

Table 2 – Summary of conclusions made based on conducting cross-decoding and context sensitivity tests jointly in our study. The conclusions refer to how information about a target variable is encoded in a given brain region. The first step is to determine whether there is evidence that the target variable is represented in the region, which can be shown by significant decoding or cross-decoding of the variable from voxel activity in that region. When there is no evidence that the variable can be decoded from a region, the conclusion is that the variable is not represented in the region. We use gray to represent this in our results. When there is evidence that the variable is represented in the region, results from the cross-decoding and context sensitivity tests are used to determine how it is represented.

(a) Interpretation of Shape Decoding Analyses

Was Shape Successfully Decoded?	
No	Shape not Represented
Yes	Shape Represented, See Table Below

		Context Sensitivity	
		Significant	Not Significant
Cross-Decoding	Significant	Not Clear How Shape is Represented	Shape Representation Invariant to Motion
	Not Significant	Shape Representation Sensitive to Motion	Not Clear How Shape is Represented

(b) Interpretation of Motion Decoding Analyses

Was Motion Successfully Decoded?	
No	Motion not Represented
Yes	Motion Represented, See Table Below

		Context Sensitivity	
		Significant	Not Significant
Cross-Decoding	Significant	Not Clear How Motion is Represented	Motion Representation Invariant to Shape
	Not Significant	Motion Representation Sensitive to Shape	Not Clear How Motion is Represented

3.1. *Analysis excluding influence of facial pose*

Here we report the results of our main analysis, including conditions in which motion did not modify facial pose (see stimuli inside the red boundaries of Fig. 2). The group results are summarized in Fig. 3, which shows each of the fROIs in our study colored using the code presented in Table 2 for the conclusion based on the two-test strategy: pink for invariance, cyan for context-sensitivity, yellow for inconclusive results, and gray for no evidence that the variable was represented in the region. As shown in Table 2, inconclusive results regarding how a face variable is represented (highlighted in yellow) occur in two cases: when there is evidence for both invariance and context sensitivity (i.e., both tests are significant), or when there is evidence for neither (i.e., neither test is significant, even though overall decoding is significant). Whenever the individual analyses revealed a consistent conclusion across participants (a prevalent conclusion of invariance, context-specificity, or inconclusive results), we also colored the border of the fROI accordingly.

The main results at the group level are two. First, information about motion and shape can be decoded from most fROIs. The exceptions are aIT, which did not show decodeable information about either motion or shape, and IFG, which did not show decodeable information about motion. Second, the FFA encodes both motion and shape in an invariant way, the IFG encodes shape invariantly, and the OFA encodes shape in a context-specific way. We now present more detailed results for each fROI.

3.1.1. *STS*

Group results showed significant decoding of motion in the context of identities 1 (53.48%, t (11) = 2.49, p = .015) and 3 (51.44%, t (11) = 1.83, p = .048), but there were no significant cross-decoding or context sensitivity tests (all $p > .1$). We found significant results in the individual data of five participants, and a significant prevalence of motion decoding that was specific to shape context ($k = 4$, $n = 5$, $p = .049$).

Regarding shape, we found significant decoding of shape in the context of motion 1 (35.88%, t (11) = 3.35, p = .003), but

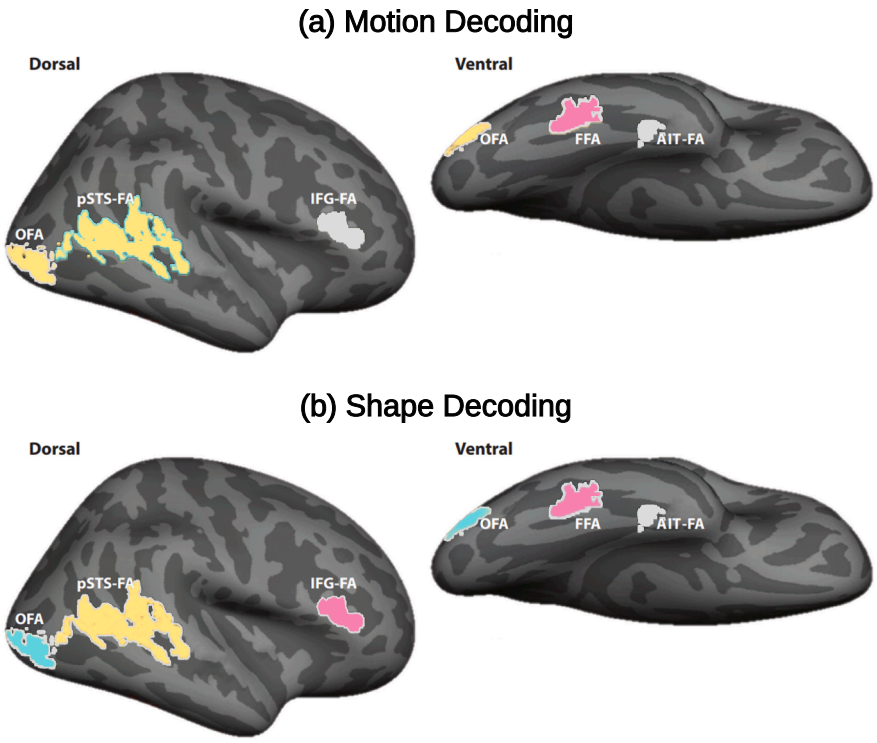

Fig. 3 – Results of the main analysis involving motion that did not modify expression shape. The area of each fROI is colored based on results for group analyses and peripheries of ROIs are colored based on results for individual analyses. Pink represents evidence for invariant encoding, cyan represents evidence for context-sensitive encoding, yellow represents inconclusive results, and gray indicates that there is no evidence that the variable was represented in the region.

again there were no significant cross-decoding or context sensitivity tests (all $p > .1$). We found significant results in the individual data of four participants, no conclusion regarding context invariance or specificity was prevalent ($k = 2$, $n = 4$, $p > .1$).

In summary, we found evidence that information about both motion and shape could be decoded from STS, but no evidence at the group level of either invariance or specificity. At the individual level, we found some evidence of context-specific encoding of motion in STS (i.e., encoding of motion information that is specific to shape).

3.1.2. FFA

Group results showed significant cross-decoding of motion (generalization from shape 1 to 3: 52.47%, t (11) = 3.63, $p = .002$), but all other tests were not significant (all $p > .1$). We found significant decoding in the individual data of only one participant, with inconclusive results regarding context invariance or specificity.

We also found significant cross-decoding of shape (generalization from motion 1 to 2: 34.60%, t (11) = 2.15, $p = .027$), but all other tests were not significant (all $p > .1$). We found significant results in the individual data of four participants, but no conclusion regarding context invariance or specificity was prevalent ($k = 3$, $n = 4$, $p > .1$).

In summary, as cross-decoding implies significant decoding that is invariant across contexts, we found some evidence that information about both motion and shape could be decoded from FFA in an invariant way. At the individual level, results were inconclusive, probably due to low power as significant decoding was found only for a few participants in each analysis.

3.1.3. OFA

We found significant cross-decoding of motion (generalization from shape 1 to 3: 52.44%, t (11) = 2.59, $p = .013$; generalization from shape 3 to 1: 52.02%, t (11) = 3.36, $p = .003$), but also significant context sensitivity when training was in the context of shape 3 (F (2, 22) = 4.53, $p = .019$), leading to inconclusive results. We found significant results in the individual data of four participants, without a prevalent conclusion about context invariance or specificity ($k = 3$, $n = 4$, $p > .1$).

We did not find significant decoding or cross-decoding of shape information (all $p > .05$), but we did find context sensitivity when training was in the context of motion 2 (F (1, 11) = 5.469, $p = .029$). We found significant results in the individual data of four participants, with no prevalent conclusion about context invariance or sensitivity ($k = 2$, $n = 4$, $p > .1$).

In summary, we found evidence that information about motion can be decoded from the OFA and, although group results were not significant, individual results also suggested that information about shape can be decoded from this area. Regarding context invariance and specificity, the results were in general mixed, with the only consistent result coming from the group analysis and showing that decoding of shape is sensitive to motion.

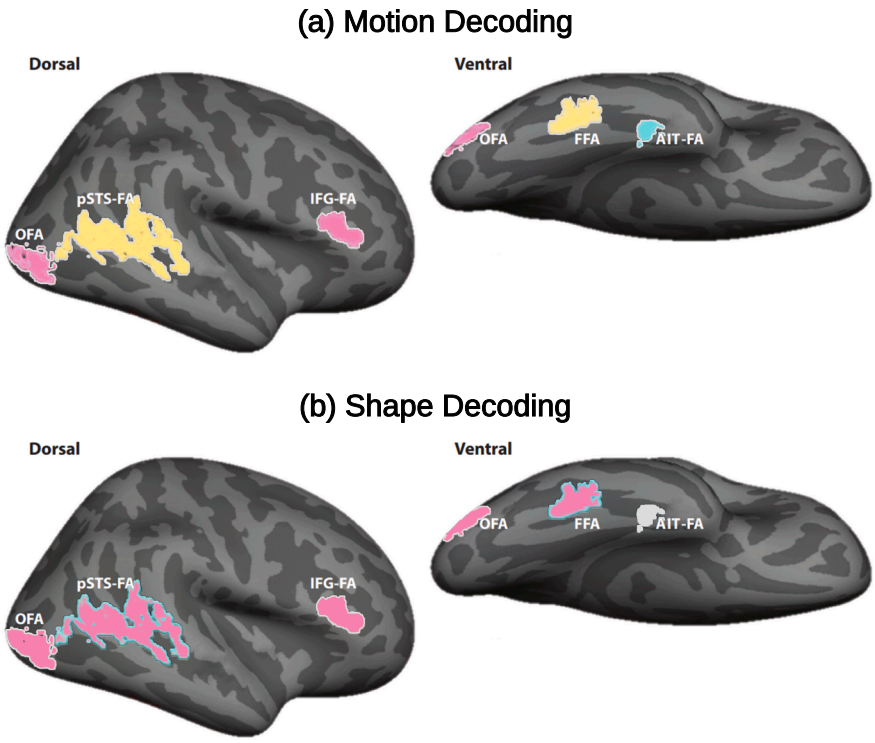

Fig. 4 – Results of the secondary analysis involving motion that changed expression shape. The area of each fROI is colored based on results for group analyses and peripheries of ROIs are colored based on results for individual analyses. Pink represents evidence for invariant encoding, cyan represents evidence for context-sensitive encoding, yellow represents inconclusive results, and gray indicates that there is no evidence that the variable was represented in the region.

3.1.4. *IFG*

We found no evidence of decoding, cross-decoding, or context sensitivity at the group level (all $p > .05$). Five participants showed significant results in their individual data analyses, without a prevalent conclusion about context invariance or sensitivity ($k = 3$, $n = 5$, $p > .1$).

Regarding shape, we found significant evidence of decoding (in the context of motion 1: 35.15%%, t (11) = 1.91, $p = .041$) and cross-decoding (generalization from motion 2 to 1: 34.89%, t (11) = 2.33, $p = .019$) from IFG. Five participants showed significant results in their individual data analyses, without a prevalent conclusion about context invariance or sensitivity ($k = 3$, $n = 5$, $p > .1$).

In summary, the strongest evidence from IFG comes from group results showing that information about shape can be decoded invariantly. While group results were not significant, we also found some evidence that information about motion can be decoded from IFG, but whether the code is invariant or context-specific is inconclusive due to mixed results across participants.

3.1.5. *aIT*

We found no evidence of decoding, cross-decoding, or context sensitivity at the group level, for either the motion or shape analyses (all $p > .1$). Results from individual analyses were also limited, with only one participant in each analysis showing any evidence of decoding or cross-decoding. In both cases, no conclusions about context-sensitivity or invariance could be made.

3.2. *Analysis including influence of facial pose*

In this section we describe results when all three levels of shape and all three levels of motion were used to train and test the decoders (decoders 1 and 2).

The group results are summarized in Fig. 4, which shows each of the fROIs in our study colored using the code presented in Table 2 for the conclusion based on the two-test strategy: red for invariance, blue for context-specificity, yellow for inconclusive results, and gray for no significant results. We considered inconclusive results both the case in which decoding was significant at the training context, but neither the cross-decoding or context sensitivity tests were significant, or when both of these tests were significant. Whenever the individual analyses revealed a consistent conclusion across participants (a prevalent conclusion of invariance, context-specificity, or inconclusive results), we also colored the border of the fROI accordingly.

The main result of including shape-changing motion in the analysis was an increase in conclusions of invariance throughout the network, not only for motion but also for shape. The only exception was aIT, where we found that including shape-changing motion led to decoding of motion that was shape-specific. We now present more detailed results for each fROI.

3.2.1. *STS*

Group results showed significant decoding of motion in the context of shape 1 (36.43%, t (11) = 3.11, $p = .005$), but there

were no significant cross-decoding or context sensitivity tests (all $p > .1$). We found significant results in the individual data of six participants, but no conclusion about invariance or context-sensitivity was prevalent among them ($k = 2$, $n = 6$, $p > .1$).

Regarding shape, we found significant decoding (in the context of motion 1: 35.88%, t (11) = 3.35, $p = .003$) and cross-decoding (generalization from motion 3 to 1: 34.58%, t (11) = 2.07, $p = .032$), suggesting encoding of shape in the STS that is invariant across changes in motion. Surprisingly, different results were found in the analysis of individual data. We found significant test results in the data of five participants, and a significant prevalence of context-specific decoding of shape ($k = 4$, $n = 5$, $p = .041$).

In summary, the inclusion of different facial poses did not produce important modifications in the conclusions regarding decoding of motion in STS, but it changed importantly conclusions regarding the decoding of shape, leading to evidence of invariant decoding at the group level and context-specific decoding at the individual level.

3.2.2. FFA

Group results showed significant cross-decoding of motion (generalization from shape 1 to 3: 52.47%, t (11) = 3.63, $p = .002$), but all other tests were not significant (all $p > .1$). We found significant decoding in the individual data of only one participant, with inconclusive results regarding context invariance or specificity. We found significant results in the individual data of only two participants, without a significant prevalence ($k = 1$, $n = 2$, $p > .1$).

We found significant cross-decoding of shape (generalization from motion 1 to 2: 52.47%, t (11) = 3.63, $p = .002$); generalization from motion 3 to 2: 52.47%, t (11) = 3.63, $p = .002$)) at the group level. Once again, different results were found in the analysis of individual data. We found significant test results in the data of three participants, and a significant prevalence of context-sensitive decoding of shape (3 = 4, $n = 3$, $p = .033$).

In summary, the inclusion of different facial poses altered the conclusions regarding invariant decoding of motion from the FFA, and at the individual level it produced evidence of context-specific decoding of shape.

3.2.3. OFA

We found significant cross-decoding of motion (generalization from shape 1 to 2: 34.86%, t (11) = 3.63, $p = .002$; generalization from shape 1 to 3: 35.02%, t (11) = 3.63, $p = .002$; generalization from shape 3 to 1: 34.84%, t (11) = 3.63, $p = .002$) but no other tests were significant (all $p > .1$). We found significant results in the individual data of six participants, but no significant prevalence of conclusions regarding invariance or context-specificity ($k = 3$, $n = 6$, $p > .1$).

Similarly, we found significant evidence for cross-decoding of shape (generalization from motion 1 to 3: 35.57%, t (11) = 2.73, $p = .009$; generalization from motion 2 to 3: 34.84%, t (11) = 1.84, $p = .046$; generalization from motion 3 to 2: 35.19%, t (11) = 2.86, $p = .008$) but no other tests were significant (all $p > .1$). We found significant results in the individual data of eight participants, but no significant prevalence of conclusions regarding invariance or context-specificity ($k = 3$, $n = 8$, $p > .1$).

In summary, significant decoding and cross-decoding of motion was obtained, indicating the encoding of facial motion information in OFA. The evidence regarding the independence of encoding from shape information was inconclusive. Also, significant decoding and cross-decoding of shape was obtained, indicating the encoding of facial shape information in OFA. The evidence regarding the independence of encoding from motion mostly pointed to sensitive encoding. Although results at the individual level had some variations, results at the group level pointed to shape information encoding being tolerant to motion information.

In summary, the inclusion of different facial poses altered the conclusions from our analyses, leading to conclusions of invariant decoding of both motion and shape from OFA.

3.2.4. IFG

We found evidence of motion decoding (in the context of shape 2: 34.84%, t (11) = 1.93, $p = .039$) and cross-decoding (generalization from shape 2 to 3: 35.21%, t (11) = 2.28, $p = .022$; generalization from shape 3 to 2: 34.89%, t (11) = 2.01, $p = .035$), but no evidence of context sensitivity (all $p > .1$). We found significant results in the individual results of four participants, but no significant prevalence of conclusions regarding invariance or context-specificity ($k = 3$, $n = 4$, $p > .1$).

Similarly, we found evidence of shape decoding (in the context of motion 1: 35.15%, t (11) = 1.91, $p = .041$) and cross-decoding (generalization from motion 2 to 1: 34.89%, t (11) = 2.33, $p = .020$; generalization from motion 2 to 3: 34.86%, t (11) = 3.17, $p = .004$; generalization from motion 3 to 2: 35.30%, t (11) = 3.44, $p = .003$), but no evidence of context sensitivity (all $p > .1$). We found significant results in the individual results of seven participants, but no significant prevalence of conclusions regarding invariance or context-specificity ($k = 3$, $n = 7$, $p > .1$).

In summary, the inclusion of different facial poses altered the conclusions from our analyses, leading to conclusions of invariant decoding of motion from OFA.

3.2.5. aIT

We found evidence of context sensitivity of motion when training was in the context of shape 1 (F (2, 22) = 4.78, $p = .016$), with all other tests being nonsignificant (all $p > .1$). At the individual level, we found significant test results for six participants, but no significantly prevalent conclusion ($k = 4$, $n = 6$, $p = .097$).

Regarding motion, none of the group tests were significant (all $p > .1$). We found significant test results for four participants, with no significantly prevalent conclusion ($k = 3$, $n = 4$, $p > .1$).

In summary, the inclusion of different facial poses in this analysis altered the conclusions from our main analysis, leading to the conclusion of context-specific decoding of motion from aIT.

4. Discussion

In this study, we used a two-test strategy (Soto & Narasiwodeyar, 2023) to examine whether neural representations of face shape and motion in face-selective cortical

regions (OFA, FFA, STS, aIT, and IFG) are invariant or context-specific. This approach mitigates the risk of false-positive findings of invariance, a known concern in neuroimaging analyses (Sandhaeger & Siegel, 2023; Soto et al., 2018; Soto & Narasiwodeyar, 2023). To independently manipulate shape and motion, we generated videos from 3D synthetic face models (Anzellotti & Caramazza, 2014; Hosseini & Soto, 2024). Our design included conditions in which motion varied without changing the involved facial poses, as well as a condition in which both motion and pose were unique. The main findings can be summarized in four key points, which we outline and discuss below.

4.1. *Ubiquitous representation of shape and motion in the face network*

First, we found that both facial motion and shape information could be decoded—either at the group or individual level—-from all regions within the face-selective network. Based on the model proposed by Bernstein and Yovel (2015), we pre-registered the hypothesis that facial shape would be decodable from ventral pathway areas (OFA, FFA, and aIT), while facial motion would be decodable from dorsal pathway areas (STS and IFG). These hypotheses were supported by our results. Additionally, the finding that motion information could also be decoded from OFA is not surprising, given its proposed role as an early-stage hub that sends output to both the ventral and dorsal pathways (e.g., Nguyen et al., 2014). However, the broader observation that all regions in the network encoded both shape and motion was unexpected and not fully anticipated by existing models of face processing.

Decoding of shape information from dorsal stream areas can be reconciled with the Bernstein and Yovel (2015) model, which posits that the STS represents implied motion derived from static facial images. Recent findings suggest that observers form expectations about how motion should unfold in differently shaped faces (Tordjman et al., 2026). If such motion expectations are encoded in the dorsal stream in a manner similar to implied motion, then these representations may implicitly carry information about the underlying facial structure that generates them.

Decoding of motion information from the ventral stream is consistent with prior findings showing that the FFA responds more strongly to dynamic than static facial stimuli. Bernstein and Yovel (2015) reviewed this literature and argued that such effects likely reflect sensitivity to pose variations over time (which temporarily modifies the superficial shape of the face), rather than direct encoding of motion per se. In our study, however, we generated two motion conditions using identical pose sequences presented in reverse order, thereby holding pose variation constant while reversing motion direction. Despite this control, we still observed reliable decoding of motion from FFA activity, suggesting that this region may encode aspects of motion beyond changes in pose.

The model proposed by O'Toole et al. (2002) may account for our findings with the addition of some reasonable assumptions. To explain how face identification can be enhanced by motion, or even achieved using motion information alone, these authors suggested that the dorsal pathway provides motion input to the ventral stream, enabling structure-from-motion representations in the FFA. In our study, participants were familiarized with faces through dynamic videos, giving them the opportunity to learn how motion altered facial shape. These dynamic signatures could, in principle, be encoded in both motion directions (e.g., from neutral to surprise and back), but may emphasize different structural aspects of the face depending on motion direction. This could result in distinguishable FFA activity patterns across motion conditions. Under this interpretation, one can maintain the view that the FFA primarily encodes facial shape, while also receiving input from motion-sensitive regions that inform or refine those shape representations through dynamic cues.

4.2. *Invariance and context-sensitivity of shape and motion representations in the face network*

This brings us to our second set of key findings, concerning the modulation of face representations by context. We found that the OFA encodes shape representations that are specific to motion context, the FFA encodes invariant representations of both shape and motion, and the IFG encodes shape representations that are also invariant across motion contexts.

With regard to the FFA, we had originally hypothesized—based on the Bernstein and Yovel (2015) model and prior work (Rostalski et al., 2022)—that this region would exhibit invariant shape representations. This prediction was confirmed. However, we did not anticipate finding invariant motion representations in FFA, a result that challenges influential models of face processing (e.g., Duchaine & Yovel, 2015; O'Toole et al., 2002), which assign motion processing primarily to dorsal-stream regions. One possible explanation lies in our modified structure-from-motion account: if different motion trajectories emphasize different structural features of a face, then using identical motion trajectories across identities may have facilitated generalization. In this case, the FFA may encode structural information revealed through motion, rather than motion per se, leading to apparent motion-invariant representations.

The IFG results are more challenging for models that characterize it as an extension of the dorsal stream receiving input from STS (Bernstein & Yovel, 2015; Kegel et al., 2021; Uono et al., 2017). We found invariant decoding of facial shape—but not motion—in IFG, and evidence for motion decoding only emerged when the stimuli involved motion that altered facial pose. This pattern suggests that IFG forms shape representations that are invariant to motion, but does not support invariant representations of motion itself. This interpretation aligns with prior findings showing that IFG activity does not correlate with STS (Foley et al., 2012), but does correlate with FFA (Liu et al., 2021); that IFG responds robustly to facial expressions/poses (e.g., Engell & Haxby, 2007; Muukkonen & Salmela, 2022; Uono et al., 2017) without a clear preference for dynamic over static stimuli (Liu et al., 2021); and that IFG encodes face shape in a manner invariant to attributes that do not change facial structure (Visconti di Oleggio Castello et al., 2017). Additionally, IFG is implicated in higher-level social cognitive processes (Diveica

et al., 2021; Perry et al., 2017), which may further support its role in abstract, shape-based representations of faces. Taken together, these findings suggest that IFG may underlie the motion-invariant shape perception reported in prior psychophysical studies (Martin et al., 2024). However, our results do not clearly indicate where motion representations that are invariant to facial shape—also observed in psychophysical work—might reside in the brain.

In our results, we did not find evidence that the STS encodes motion information exclusively, nor that its motion representations are invariant to changes in facial shape. Prior research has implicated the STS in a range of functions related to face processing, including identity (shape) recognition (O'Toole et al., 2002), emotional expression (Bernstein et al., 2018; Muukkonen & Salmela, 2022; O'Toole et al., 2002; Sellal, 2022), and facial motion (Bernstein & Yovel, 2015; Hoffman & Haxby, 2000; O'Toole et al., 2002). These findings support the view of STS as a flexible and multifunctional node within the face-processing network (Thome et al., 2022).

As aIT is located anterior to the FFA within the ventral stream and is generally thought to receive input from the FFA (Pyles et al., 2013), some researchers have proposed that it serves as the endpoint of the ventral stream where invariant identity representations are formed (e.g., Anzellotti et al., 2014; Kriegeskorte et al., 2007; Nasr & Tootell, 2012; Nestor et al., 2011). From this perspective, the FFA may primarily support face detection rather than identification (Iidaka, 2014; Kriegeskorte et al., 2007). However, our findings do not support this view. We found no evidence of decoding or cross-decoding in aIT in any of the group-level analyses, and only limited evidence of decoding at the individual level. If anything, the individual data showed a tendency toward context-specific identity (shape) decoding in the secondary analysis.

In contrast, we observed robust cross-decoding of shape in the FFA across both analyses, suggesting that information necessary for identifying individual faces—not merely detecting their presence—is represented in this region. These results challenge the notion that invariant identity representations are formed exclusively in aIT and instead highlight a prominent role for FFA in supporting face identification.

4.3. *Confounds between motion and pose increase conclusions of invariance*

Our third key finding was that including pose-changing motion in the analysis substantially altered the conclusions of our decoding results. Specifically, this led to an overall increase in observed invariance across the face-processing network—not only for motion, but also for shape. While this outcome is not entirely surprising, it highlights the critical importance of controlling for pose-change confounds when studying facial motion. In the context of shape decoding, increased invariance may arise from similarities in the initial or final expression poses across videos. For example, as illustrated in Fig. 2, the added dynamic condition (right column; slightly anti-disgust to disgust) ends with a shape that closely resembles the final frame of the first video (surprise to anti-surprise), potentially facilitating generalization of shape decoding across contexts. For motion decoding, the result can be explained by the confound between motion and pose in the stimuli. Using the same example, the final expressions in the first and third column have a similar and distinct pose, which can produce better decoding and more generalization across identities.

To be sure, pose-changing motion reflects the dynamics most commonly encountered in natural face perception. In this sense, continuing to study it—as previous research has done—is both appropriate and necessary. However, by holding pose constant, as we did in our main analysis, it becomes possible to isolate and identify what is genuinely being encoded in each face-selective area. Thus, while the results from our fixed-pose analysis provide greater interpretive precision, those from the changing-pose condition may better reflect how the visual system operates in real-world settings.

4.4. *High individual variability in the representation of motion and shape*

Our fourth and final key finding was the high degree of heterogeneity observed in individual-level analyses. Most individuals showed a mixture of context-sensitive and invariant representations, with only context-sensitivity emerging as a consistent result across participants. Notably, this individual-level prevalence of context-sensitivity sometimes contradicted the conclusions drawn from group-level analyses. This highlights the potential for divergent interpretations depending on the level of analysis and underscores the need for more multi-level approaches in face perception research. Prior studies have demonstrated substantial individual variation in face processing ability, which is not only heritable (Wilmer et al., 2010; Zhu et al., 2010) but also genetically distinct from other perceptual processes (Shakeshaft & Plomin, 2015). These findings suggest that individual differences should play a central role in the development of future models of the functional architecture of face perception. Moreover, future studies should report and interpret individual-level variability alongside group-level trends, rather than treating group averages as universally representative.

4.5. *Limitations*

Here, we focused on testing the interaction between representations of facial shape and motion within the face-processing network. To examine motion per se—that is, motion independent of changes in facial pose—we used two conditions in which dynamic changes between the same facial poses were presented in reverse order (anti-surprise surprise *vs.* surprise anti-surprise). A limitation of this manipulation is that the perception of emotion from facial motion depends not only on static poses but also on the dynamics of the movement (e.g., Jack et al., 2014). Consequently, one sequence of expression poses may be interpreted as conveying a more coherent or prototypical emotional expression than the other. Consistent with this, internal emotional states inferred from dynamic expressions have been shown to follow characteristic temporal transitions (Jeganathan et al., 2022).

In our stimuli, this implies that one motion sequence could be perceived as a transition from an emotion with negative valence (see anti-surprise in Fig. 2) to one without such valence (i.e., surprise), whereas the reverse sequence might be perceived as a transition in the opposite affective direction.

Thus, the difference between our two main motion conditions could, in principle, be represented in some brain regions not as a difference in motion itself, but rather as a difference in the emotion conveyed or in the emotional transition implied by the sequence.

Our primary goal, however, was to test theories of visual face perception rather than affective processing, as framed in the Introduction. Nevertheless, it remains possible that some areas in the face network encode not only facial shape and motion but also contribute to the inference of affective states from dynamic facial expressions. Because both familiarity and affective transitions are difficult to control using naturalistic emotional expressions, a promising direction for future work would be to employ unfamiliar expression poses—such as those that can be generated using 3D face modeling software like the one used in our study—to better isolate perceptual from affective influences on neural representations of facial motion.

An important caveat to our conclusions regarding invariance and context-specificity concerns the limited temporal resolution of fMRI. Initial feedforward processing of facial information occurs rapidly—within a few hundred milliseconds of face onset (e.g., Jacques et al., 2007; Rossion & Jacques, 2008)—and is followed by recurrent interactions among face-selective areas and top-down modulation. These fast dynamics are not captured by the BOLD signal, which reflects neural activity on the scale of seconds. As a result, the context sensitivity or invariance observed within a given region may, in part, reflect interregional communication or top-down influences that fMRI cannot temporally resolve. Applying our two-test strategy to data with high temporal resolution, such as M/EEG, could clarify the timing of these neural representations as they unfold during face perception (Miki et al., 2022). For studies seeking both high temporal and spatial resolution, approaches such as simultaneous EEG-fMRI (Dellert et al., 2021) or source-localized MEG or high-density EEG offer promising avenues for future research.

Multiple studies have subdivided the STS into posterior (pSTS), middle (mSTS), and anterior (aSTS) subregions, reporting functional differences among them during face perception (Pitcher et al., 2011; Winston et al., 2004; Zhang et al., 2020). For instance, Zhang et al. (2020) found that aSTS is significantly more sensitive to non-rigid facial motion than either mSTS or pSTS. This distinction presents a limitation in our study, as we defined the STS using a pSTS parcel (Julian et al., 2012). It remains an open question whether our findings would generalize to more fine-grained analyses that differentiate STS subregions. Future work may benefit from adopting region-specific parcellation strategies to clarify the distinct functional roles of these subdivisions in face processing.

Finally, our study included a relatively small sample of participants, each tested extensively over long sessions (approximately 3 h, with 126 presentations of each stimulus). This design was deliberately chosen to enable direct comparison between group-level and individual-level analyses, providing a more detailed view of how face representations may vary across participants. While this approach offers rich within-subject data, it necessarily limits the generalizability of our findings.

Future progress in understanding of the face-processing network will likely require large, systematically designed datasets that combine extensive sampling—both across participants and longitudinally—with parameterized facial stimuli that vary along theoretically relevant dimensions. Such datasets should include trials that isolate specific components of facial information (e.g., shape, motion, expression pose, gaze, emotion, valence) to allow more precise tests of competing theoretical models. Importantly, large shared datasets of this kind would enable multiple research groups to address complementary questions using common data resources, fostering cumulative progress.

Until such large-scale efforts become feasible, studies like ours will continue to face practical trade-offs imposed by limited budgets and participant availability. These constraints necessarily limit the generality of our conclusions, yet they also provide focused insights that can inform and motivate the design of future collaborative, large-scale investigations into the organization of the face-processing network.

5. Conclusion

In summary, our study provides several key insights into how shape and motion information are represented across the face-selective cortical network. First, we found that all core and extended face areas encode both motion and shape information, with evidence of overlapping functionality across the traditionally defined ventral and dorsal pathways. Second, our analysis revealed differences in invariance and context sensitivity across regions: OFA exhibited motion-specific shape representations, IFG showed motion-invariant representations of shape, and FFA displayed invariant representations of both shape and motion—challenging prevailing models that limit motion processing to the dorsal stream. Third, we showed that including shape-changing motion in the stimuli significantly increases observed invariance, underscoring the importance of controlling for motion-shape confounds in face perception research. Fourth, individual-level analyses revealed substantial heterogeneity, emphasizing the need for multi-level modeling and greater attention to individual differences. Finally, our results question common assumptions about the functional hierarchy of the ventral stream, with little evidence of invariant shape representation in aIT and strong evidence of such representations in FFA. Together, these findings call for a more nuanced understanding of how face shape and motion are encoded in the brain, and they highlight the value of an approach to study brain representations that overcomes the dangers of false positive invariance inherent in the application of cross-decoding alone.

CRediT authorship contribution statement

S. Sanaz Hosseini: Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Fabian A. Soto:** Writing – review & editing, Writing – original draft, Supervision, Project administration,

Methodology, Funding acquisition, Formal analysis, Conceptualization.

Funding

This work was supported by the National Science Foundation under grant No. 2020982 to Fabian A. Soto. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. All research at the Department of Psychiatry in the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312) and the NIHR Applied Research Collaboration East of England. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Scientific transparency statement

DATA: All raw and processed data supporting this research are publicly available: https://openneuro.org/datasets/ds007369, https://osf.io/q2v68.

CODE: All analysis code supporting this research is publicly available: https://osf.io/q2v68.

MATERIALS: Materials openly available include the PsychoPy tasks and stimuli used in the familiarization and main fMRI experiments described in the manuscript. These materials can be accessed via OSF: https://osf.io/q2v68.

The Dynamic Localizer task and associated stimuli were provided to the authors by the Human Vision and Eye Movement Laboratory at the University of British Columbia (Director: Dr. Jason J. S. Barton) for research use only, under conditions that prohibit redistribution. As a result, these materials cannot be shared by the authors. Researchers wishing to obtain access should contact the original providers directly.

DESIGN: This article reports, for all studies, how the author(s) determined all sample sizes, all data exclusions, all data inclusion and exclusion criteria, and whether inclusion and exclusion criteria were established prior to data analysis.

PRE-REGISTRATION: At least part of the study procedures was pre-registered in a time-stamped, institutional registry prior to the research being conducted: https://osf.io/snrm4 At least part of the analysis plans was pre-registered in a time-stamped, institutional registry prior to the research being conducted: https://osf.io/snrm4 The analyses that were undertaken deviated from the preregistered analysis plans. All such deviations are fully disclosed in the manuscript.

For full details, see the *Scientific Transparency Report* in the supplementary data to the online version of this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2026.02.006.

REFERENCES

Abdulrahman, H., & Henson, R. N. (2016). Effect of trial-to-trial variability on optimal event-related fmri design: Implications for beta-series correlation and multi-voxel pattern analysis. *Neuroimage, 125*, 756–766.

Anzellotti, S., & Caramazza, A. (2014). The neural mechanisms for the recognition of face identity in humans. *Frontiers in psychology, 5*, 672.

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2014). Decoding representations of face identity that are tolerant to rotation. *Cerebral cortex, 24*, 1988–1995.

Bernstein, M., Erez, Y., Blank, I., & Yovel, G. (2018). An integrated neural framework for dynamic and static face processing. *Scientific reports, 8*, 7036.

Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience and Biobehavioral Reviews, 55*, 536–546.

Deen, B., & Saxe, R. (2019). Parts-based representations of perceived face movements in the superior temporal sulcus. *Human Brain Mapping, 40*, 2499–2510. https://doi.org/10.1002/hbm.24540

Dellert, T., Müller-Bardorff, M., Schlossmacher, I., Pitts, M., Hofmann, D., Bruchmann, M., & Straube, T. (2021). Dissociating the neural correlates of consciousness and task relevance in face perception using simultaneous eeg-fmri. *Journal of Neuroscience, 41*, 7864–7875.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Albert, MS (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage, 31*, 968–980.

Diveica, V., Koldewyn, K., & Binney, R. J. (2021). Establishing a role of the semantic control network in social cognitive processing: A meta-analysis of functional neuroimaging studies. *Neuroimage, 245*, Article 118702.

Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies—A review of behavioral findings and methodology. *Frontiers in Psychology, 9*, 1355.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual review of vision science, 1*, 393–416.

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia, 45*, 3234–3241.

Foley, E., Rippon, G., Thai, N. J., Longe, O., & Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: A connectivity analysis study. *Journal of cognitive neuroscience, 24*, 507–520.

Fox, C. J., Iaria, G., & Barton, J. J. (2009). Defining the face processing network: Optimization of the functional localizer in fmri. *Human brain mapping, 30*, 1637–1651.

Friston, K., Josephs, O., Zarahn, E., Holmes, A., Rouquette, S., & Poline, J. B. (2000). To smooth or not to smooth?: Bias and efficiency in fmri time-series analysis. *Neuroimage, 12*, 196–208.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics, 13*.

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

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on

multivariate pattern analysis applied to time series neuroimaging data. *Journal of cognitive neuroscience, 29*, 677–697.
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.
Hays, J., Wong, C., & Soto, F. A. (2020). Faret: A free and open-source toolkit of three-dimensional models and software to study face perception. *Behavior Research Methods, 52*, 2604–2622.
Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience, 3*, 80–84.
Hosseini, S. S., & Soto, F. A. (2024). Multidimensional signal detection modeling reveals Gestalt-like perceptual integration of face emotion and identity. *Emotion, 24*, 1494–1502.
Iidaka, T. (2014). Role of the fusiform gyrus and superior temporal sulcus in face perception and recognition: An empirical review. *Japanese Psychological Research, 56*, 33–45.
Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology, 24*, 187–192. https://doi.org/10.1016/j.cub.2013.11.064. URL: https://www.cell.com/current-biology/abstract/S0960-9822(13)01519-4.
Jacques, C., d'Arripe, O., & Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. *Journal of vision, 7*, 3–3.
Jeganathan, J., Campbell, M., Hyett, M., Parker, G., & Breakspear, M. (2022). Quantifying dynamic facial expressions under naturalistic conditions. *eLife, 11*, Article e79581. https://doi.org/10.7554/eLife.79581 (publisher: eLife Sciences Publications, Ltd).
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*, 825–841.
Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage, 62*, 782–790.
Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage, 60*, 2357–2364.
Kegel, L. C., Frühholz, S., Grunwald, T., Mersch, D., Rey, A., & Jokeit, H. (2021). Temporal lobe epilepsy alters neural responses to human and avatar facial expressions in the face perception network. *Brain and Behavior, 11*, Article e02140.
Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences, 104*, 20600–20605.
Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review, 5*, 41–46.
Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). The many faces of research on face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*, 1634–1637.
Liu, M., Liu, C. H., Zheng, S., Zhao, K., & Fu, X. (2021). Reexamining the neural network involved in perception of facial expression: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 131*, 179–191.
Marian, D. E., & Shimamura, A. P. (2013). Contextual influences on dynamic facial expressions. *The American Journal of Psychology, 126*, 53–66.
Martin, E. R., Hays, J. S., & Soto, F. A. (2024). Face shape and motion are perceptually separable: Support for a revised model of face processing. *Psychonomic Bulletin & Review*, 1–10.
Mattavelli, G., Andrews, T. J., Asghar, A. U., Towler, J. R., & Young, A. W. (2012). Response of face-selective brain regions to trustworthiness and gender of faces. *Neuropsychologia, 50*, 2205–2211.
Miki, K., Takeshima, Y., Watanabe, S., & Kakigi, R. (2022). Human face perception using electroencephalography and magnetoencephalography. *Frontiers in Physiology, 13*, 416.
Muukkonen, I., & Salmela, V. (2022). Representational structure of fmri/eeg responses to dynamic facial expressions. *Neuroimage, 263*, Article 119631.
Nasr, S., & Tootell, R. B. (2012). Role of fusiform and anterior temporal cortical areas in facial recognition. *Neuroimage, 63*, 1743–1753.
Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences, 108*, 9998–10003.
Nguyen, V. T., Breakspear, M., & Cunnington, R. (2014). Fusing concurrent eeg–fmri with dynamic causal modeling: Application to effective connectivity during face perception. *Neuroimage, 102*, 60–70.
Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *Plos Computational Biology, 10*, Article e1003553.
O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in cognitive sciences, 6*, 261–266.
Pedregosa, F., Eickenberg, M., Ciuciu, P., Thirion, B., & Gramfort, A. (2015). Data-driven hrf estimation for encoding and decoding models. *Neuroimage, 104*, 209–220.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
Perry, A., Saunders, S. N., Stiso, J., Dewar, C., Lubell, J., Meling, T. R., Solbakk, A. K., Endestad, T., & Knight, R. T. (2017). Effects of prefrontal cortex damage on emotion understanding: Eeg and behavioural evidence. *Brain: a Journal of Neurology, 140*, 1086–1099.
Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic *vs.* static information in face-selective cortical regions. *Neuroimage, 56*, 2356–2363.
Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined tms and fmri reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology, 24*, 2066–2070.
Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences, 25*, 100–110.
Pyles, J. A., Verstynen, T. D., Schneider, W., & Tarr, M. J. (2013). Explicating the face perception network with white matter connectivity. *Plos One, 8*, Article e61611.
Quadflieg, S., Todorov, A., Laguesse, R., & Rossion, B. (2012). Normal face-based judgements of social characteristics despite severely impaired holistic face processing. *Visual cognition, 20*, 865–882.
Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the n170. *Neuroimage, 39*, 1959–1979.
Rostalski, S. M., Robinson, J. E., Ambrus, G. G., Johnston, P., & Kovács, G. (2022). Person identity-specific adaptation effects in the ventral occipito-temporal cortex. *European Journal of Neuroscience, 55*, 1232–1243.
Sandhaeger, F., & Siegel, M. (2023). Testing the generalization of neural representations. *Neuroimage, 278*, Article 120258.

Sellal, F. (2022). Anatomical and neurophysiological basis of face recognition. *Revue Neurologique, 178*, 649–653.
Senior, C., Hassel, S., Waheed, A., & Ridout, N. (2020). Naming emotions in motion: Alexithymic traits impact the perception of implied motion in facial displays of affect. *Emotion, 20*, 311.
Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences, 112*, 12887–12892.
Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional mri. *Neuroimage, 58*, 588–594.
Sliwinska, M. W., Bearpark, C., Corkhill, J., McPhillips, A., & Pitcher, D. (2020). Dissociable pathways for moving and static face perception begin in early visual cortex: Evidence from an acquired prosopagnosic. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 130*, 327–339.
Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping, 17*, 143–155.
Smith, S. M., & Brady, J. M. (1997). Susan—A new approach to low level image processing. *International journal of computer vision, 23*, 45–78.
Soto, F. A., & Narasiwodeyar, S. (2023). Improving the validity of neuroimaging decoding tests of invariant and configural neural representation. *PLOS Computational Biology, 19*, Article e1010819.
Soto, F. A., Vucovich, L. E., & Ashby, F. G. (2018). Linking signal detection theory and encoding models to reveal independent neural representations from neuroimaging data. *Plos Computational Biology, 14*, Article e1006470.
Spillmann, L., Dresp-Langley, B., & Tseng, C. H. (2015). Beyond the classical receptive field: The effect of contextual stimuli. *Journal of Vision,* 15. https://doi.org/10.1167/15.9.7. 7–7.
Thome, I., Alanis, J. C. G., Volk, J., Vogelbacher, C., Steinsträter, O., & Jansen, A. (2022). Let's face it: The lateralization of the face perception network as measured with fmri is not clearly right dominant. *Neuroimage, 263*, Article 119587.
Tordjman, R., Martin, E. R., & Soto, F. A. (2026). Motion-from-Structure in Face Perception: Expectations of Natural Face Motion Depend on Face Shape. Retrieved from osf.io/preprints/psyarxiv/jt3h7_v1.
Uono, S., Sato, W., Kochiyama, T., Sawada, R., Kubota, Y., Yoshimura, S., & Toichi, M. (2017). Neural substrates of the ability to recognize facial expressions: A voxel-based morphometry study. *Social Cognitive and Affective Neuroscience, 12*, 487–495.
Visconti di Oleggio Castello, M., Halchenko, Y. O., Guntupalli, J. S., Gors, J. D., & Gobbini, M. I. (2017). The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Scientific reports, 7*, Article 12237.
Wang, Y., Metoki, A., Smith, D. V., Medaglia, J. D., Zang, Y., Benear, S., Popal, H., Lin, Y., & Olson, I. R. (2020). Multimodal mapping of the face connectome. *Nature human behaviour, 4*, 397–411.
Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of sciences, 107*, 5238–5241.
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.
Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology, 5*, 633.
Yoshikawa, S., & Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cognitive, Affective, & Behavioral Neuroscience, 8*, 25–31.
Zhang, H., Japee, S., Nolan, R., Chu, C., Liu, N., & Ungerleider, L. G. (2016). Face-selective regions differ in their ability to classify facial expressions. *Neuroimage, 130*, 77–90.
Zhang, H., Japee, S., Stacy, A., Flessert, M., & Ungerleider, L. G. (2020). Anterior superior temporal sulcus is specialized for non-rigid facial motion in both monkeys and humans. *Neuroimage, 218*, Article 116878.
Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology, 20*, 137–142.