Face space representations of movement

Nicholas Furl*, Florida Begum, Justin Sulik, Francesca Pizzorni Ferrarese, Sarah Jans, Caroline Woolley

Royal Holloway, University of London, Egham Hill, Egham, Surrey, TW20 0EX, United Kingdom

A B S T R A C T

The challenging computational problem of perceiving dynamic faces “in the wild” goes unresolved because most research focuses on easier questions about static photograph perception. This literature conceptualizes face representation as a dissimilarity-based “face space”, with axes that describe the dimensions of static images. Some versions express positions in face space relative to a central tendency (norm). Are facial movements represented like this? We tested for representations that accord with an a priori hypothesized motion-based face space by experimentally manipulating faces’ motion-based dissimilarity. Because we caricatured movements, we could test for representations of dissimilarity from a motion-based norm. Behaviorally, participants perceived these caricatured expressions as convincing and recognizable. Moreover, as expected, caricature enhanced perceived dissimilarity between facial expressions. Functional magnetic resonance imaging showed that occipitotemporal brain responses, including face-selective and motion-sensitive areas, reflect this face space. This evidence converged across methods including analysis of univariate mean responses (which additionally exhibited norm-based responses), repetition suppression and representational similarity analysis. This accumulated evidence for “representational geometry” shows how perception and visual brain responses to facial dynamics reflect representations of movement-based dissimilarity spaces, including explicit computation of distance from a norm movement.

1. Introduction

Humans somehow recognize faces and other complex stimuli, despite visual input that is moving and changing unpredictably. The face perception literature has primarily sidestepped this problem, using simpler static photographs, which are recognizable without interference from dynamics. However, static images are unlikely to engage all the neural representations needed to achieve reliable performance “in the wild”. Here, we localized neural representations of dynamic faces that are organized as predicted by the face space metaphor (Valentine, 1991). The axes of such spatiotemporal face spaces would reflect component spatiotemporal dimensions (e.g., displacement, speed, timing of facial features). Dynamic faces, then, would give rise to multidimensional patterns over these spatiotemporal dimension values. Distances in the space would encode dissimilarity between faces on these dimensions. Some versions of this theory also specify a dissimilarity metric computed relatively to a “norm” - the expected values of the dimensions and the origin of the face space.

To demonstrate spatiotemporal face space representations, we adapted behavioral and brain imaging methods already established for demonstrating static image-based face spaces. Such studies experimentally manipulated dissimilarity between stimuli and then tested whether perception or brain responses reflect the ensuing dissimilarity structure (Blanz et al., 2000; Calder et al., 2000; Leopold et al., 2001; Skinner and Benton, 2010; Juricevic and Webster, 2012). Caricatures have been an especially popular approach, in which static face shape is morphed to be distant (caricature) or near (anticaricature) a mean over many face shapes (a proxy for the norm). Faces with different caricature levels can also be morphed along the same trajectory but on the opposite side of the mean (i.e., antifaces). Here, we animated a shape-normalized computer head model with movements that were each manipulated along a trajectory at four caricature levels (Fig. 1A): caricatured and anticaricatured versions of a basic expression (e.g., surprise) and caricatured or anti-caricatured versions of its corresponding “antieexpression” (e.g., anti-surprise). This dissimilarity manipulation between caricature levels establishes an a priori dissimilarity structure (Fig. 1B) that can be depicted in a dissimilarity matrix (Fig. 1C). If participants’ representations are organized in this way, such representations will also be detected in participants’ recognition performance (Behavioral Study 1), behavioral dissimilarity ratings (Behavioral Study 2) and brain responses (fMRI responses).

The geometry of face space predicts that larger distances should separate caricatured faces from each other (i.e., within caricature level dissimilarity), relative to distances among anticaricatured faces (Fig. 1D and E). Indeed, for static image caricatures, participants rate pairs of caricatures as more dissimilar than pairs of anticaricatures (Lee et al., 2000; Irons et al., 2014; McKone et al., 2018), participants more accurately recognize caricatured identities (Raufmann and Schweinberger,
represents. The first analysis (“direct effects”) tests whether mean response magnitudes track distances from the average (norm) stimulus. For static images, this method has demonstrated that the averages neural responses reflect distance from a norm in electrophysiology data from monkey inferotemporal cortex (Leopold et al., 2006), and human fMRI data from the face-selective fusiform face area (FFA) (Lofler et al., 2005) and amygdala (Said et al., 2010). These findings are particularly suggestive that visual brain regions explicitly compute distance from a norm during perception. The second analysis uses repetition suppression (Grill-Spector et al., 2006) to localize regions coding the dissimilarity relations among our stimuli. This method has produced convergent evidence for face space coding in FFA, where caricature similarity with preceding faces suppresses fMRI responses (Lofler et al., 2005). The third analysis implements representational similarity analysis (RSA) to detect neural responses where multivariate pattern dissimilarities conform to the “representational geometry” of the caricature set (Kriegeskorte and Kliever, 2013). This method has provided further evidence that FFA and other face-selective areas implement face space representations of static shape.

Facial dynamics place additional computational demands on any observer who needs to ascertain social information from a face. The visual system, therefore, should describe dynamic faces over a wide spectrum of spatiotemporal component dimensions. We therefore expect that our analyses will detect hallmark distances of dissimilarity coding in many occipitotemporal face selective regions of interest (ROIs), including the FFA (Lofler et al., 2005; Carlin and Kriegeskorte, 2017). Beyond the FFA, we expect face space representations to arise in motion-sensitive areas (V5) and the superior temporal sulcus (STS), where responses are sensitive to facial (Furl et al., 2015) or biological motion (Dasgupta et al., 2017).

In short, we seek to localize dissimilarity-based (i.e., face space) representations in the brain, including spaces that code dissimilarity from a norm. We will comprehensively test this hypothesis by seeking convergent evidence across multiple established behavioral and brain imaging methods. We hypothesize that representations like these will manifest in dissimilarity ratings, in expression ratings and categorizations and in occipitotemporal visual response from face-selective, motion-sensitive and face-motion sensitive areas.

2. Materials and methods

2.1. Participants

In Behavioral Study 1, we enrolled 60 participants. Participants were Royal Holloway University of London students who received course credit from an undergraduate psychology course and participated online. In Behavioral Study 2, we enrolled 592 participants (313 females; three chose to self-describe or to opt out of a question about gender). Participants were recruited from Amazon’s Mechanical Turk (MTurk) via the Turkprime platform (Litman et al., 2017). Participation was limited to those with IP addresses in the USA, with MTurk approval ratings over 95%. In the fMRI study, we enrolled 30 right-handed participants (25 females). One participant had missing data for one of the two localizer runs. For all studies, ethics protocols were approved by the Royal Holloway University of London College Ethics Board and all participants provided informed consent.

2.2. Stimuli

2.2.1. Strategy

We constructed relatively simple animations by manipulating the facial features (Fig. 2A) of a graphical head model (Fig. 2B) based on emotional expressions extracted from facial videos (Fig. 2C and D). Consistent with established terminology (Deffenbacher et al., 1998; Valentine, 1998; Calder et al., 2000; Lee and Perrett, 2000; Lee et al., 2000; Frowd et al., 2007; Nishimura et al., 2010), here (Fig. 1A), we describe as “caricatures” the basic expressions in which the
distinctiveness of their original values is exaggerated. Also consistent with this literature, we describe as “anticareatures” the expressions rendered less distinctive. Previous studies introduced example key feature.

Behavioral Study 1 will con- 2.2.2. Motion tracking

set for which expressions were the most accurately recognized. We previously reported magnetoencephalography (MEG) responses to the movements of six of these identities (Furl et al., 2017). Using established methods implemented in the software Psychomorph (Chen and Tiddeman, 2010), we tracked 141 landmark positions in the facial interior (white and colored points in Fig. 2A) for 50 video frames (i.e., 2 s). Using this full complement of landmarks, we registered the pixels from each identity to a graphical head model, rendering it with the shape-normalized appearance of each identity (Fig. 2B). This registration was performed in Blender by locating the same landmark positions on the surface of the head model and implementing morphing. Although all landmarks were used for this registration process, we animated the head model using a subset of these landmarks, that corresponded to high-level “key facial features”: midline chin/upper lip, right and left: inner middle and outer eyebrow positions, upper and lower eyelids, corners of mouth, and upper lip (colored in Fig. 2A). These key features correspond to (some but not all) of the main points on the face that are directly non-rigidly moveable by the facial musculature (i.e., it excludes facial locations like the nose, forehead, etc. that are less moveable). Movements at some of these locations have also been shown to especially influence facial expression perception (Delis et al., 2016). One male identity anomalously reached apex expression within the first few frames of every video. Although we retained the registered pixel map for this video (i.e., its appearance) and animated the head model using this pixel map, we excluded its landmarks from the analysis of facial motion, described below. For the remaining videos, we removed effects of rigid head movement from the key features, leaving behind the non-rigid motion. We selected three landmarks on the nose, a structure that can only move with whole head, and then corrected key feature position data using an inverse affine transformation.

2.2.3. Component dimensions

We computed Euclidean distances between landmark positions on the successive frames separately for each identity and expression (Fig. 2C). The vertical components of these movements (Fig. 2D) exhibited clear sigmoidal time courses, in which each feature began at a neutral position and, at some time during the video (mid), transitioned at a certain speed (slope) to an asymptotic maximum (max). We summarized these spatiotemporal dimensions using logistic function model fitting with “max” (maximum/asymptote), “slope” (bias) and “mid” (inflation) point as free parameters. We separately fit a logistic function to time-courses for every key feature from every one of the videos we selected from the face set. The three ensuing parameters formed the basis component dimensions from which the face space was constructed. That is, the resultant caricatured videos varied in their overall amount of displacement (max), the speed at which the motion occurred (slope) and the time during the video that the motion occurred (mid).

2.2.4. Caricature

Caricature was applied to each face separately using the average parameters of that face’s expression (across the included identities) plus some Gaussian noise applied to each feature (SD = 20% of mean parameter value). Thus, the starting point of caricature for each individual video was the motion taken from its average expression, with (random) “characteristic” movements added. Subjectively, we observed that animations based these (noisy) expression averages appeared more convincing (as basic expressions) than those based on the original video-specific parameters.

By taking these parameter values and finding their differences with the averaged parameters over all videos (the norm), we could compute new movement parameters that were 170% (caricature) or 85% (anti caricature) caricature levels (Fig. 2E). Using the negative differences from the norm parameters, we also computed parameters that were the same 85% (antimovement anticaricature) and 170% (antimovement caricature) distances from the average, but on the other side of the

Fig. 2. Animations. A, We quantified motion from facial videos for 15 key features (colored). B, We registered pixel data to a graphical head model with moveable facial features (yellow boxes). C, Quiver plots show landmark veloci- ties for an example movement. D, Timecourses of the vertical components of the key features for the movement in B. E, Caricatured timecourses for an example key feature.
average set of parameters. We selected these caricature values because they produced equally-spaced (85%) increments around the norm and because the caricature levels of the animations subjectively appeared visually distinguishable.

Caricatures and antimovement caricatures exaggerate differences in logistic parameters (max, slope, mid) from average parameters, but caricatures do not necessarily “move more” than anticaricatures. Fig. 2E shows logistic functions reconstructed for one example feature. The veridical timecourse (black solid line) has a smaller displacement (max) and is slower (slope) and moves later (mid) than the norm (black dashed line). Consequently, this feature’s caricature (red line) moves an even shorter distance, more slowly, and later than the veridical or anti-caricature (orange line) timecourses. The antimony movements (cyan and blue line) reverse this pattern.

2.2.5. Animation

We implemented our graphical head model in the Blender software. The head model was equipped with 15 “drivers” (yellow boxes in Fig. 2B), each effectively a high-level feature of the face, that shifted the locations of surrounding model vertices. Each driver directly corresponded to a key feature (Fig. 2A). Logistic functions reconstructed from each key feature’s caricatured parameter values served as the movement trajectory of the corresponding Blender driver. To map movements in landmark (pixel) space to Blender driver space, we linearly-normalized the logistic-shaped movements to the range of motion for each Blender driver, such that the maximum and minimum of each landmark range (pixels) mapped to 75% of the maximum and minimum of the Blender range. Then, we programmed Blender drivers to follow linear interpolations of these normalized logistic functions.

2.2.6. Stimulus numbers

In Behavioral Study 2, participants rated dissimilarities between pairs of 180 of our resultant videos: nine identities, five expression trajectories (anger, disgust, happy, fear and surprise) and the four caricature levels. With this amount, each participant on-line could comfortably rate subsets of stimulus pairs. We used six identities (three female) and the same expression trajectories and caricature levels (120 videos) in the fMRI experiment and Behavioral Study 1. This number neared the limit of a comfortable scanning duration that could be achieved using a Type 1 Index 1 counterbalanced design (Aguirre, 2007).

The norm, which derived from the landmark data of 66 videos spanning diverse facial categories (10 identities, two sexes, six expressions), compares reasonably to previous brain imaging studies using facial caricatures. For example, Loftier et al. (2005) used norms with 40 same-sex, all-neutral, computer-generated artificial faces. Wutke and Schweinberger (2019) used a norm with 20 all-neutral identities and Zheng et al. (2012) used a norm with 32 all-female, all-neutral identities. Lee et al. (2000) used a norm computed more in range with our numbers and Schweinberger (2019), was high (mean r = 0.56). The behavioral study was programmed in jsPsych (De Leeuw, 2015) and delivered on Amazon’s Mechanical Turk to participants in the United States.

There are more pairs of faces than feasible for participants to rate in an online setting. One way we managed the number of pairs was dividing the rating task into two between-participants conditions: in one of these conditions (pairs differ in identities), each participant (N = 309) viewed faces with only one of the expressions or anticaricatures, while the two identities in each pair differed from each other on every trial. In the second condition (pairs differ in expressions), each participant (N = 283) viewed videos, all with the same one of nine identities (including the six from the fMRI study), while the two identities in each pair differed from

2.3. Behavioral study 1 procedures

We tested how convincing the animations appeared as basic expressions, compared to the original videos. We also tested our hypothesis that participants perceive caricatured videos as more convincing and recognize them more accurately than anticaricatures. These latter hypotheses derive from the geometry of face space (Fig. 1D), in which caricatured expressions should be spaced more widely than anticaricatured expressions (and are therefore less confusable), and from previous empirical studies that show similar effects using caricature of static face shape along expression trajectories in face space (Calder et al., 2000; Lane et al., 2019).

Participants viewed three versions of the same identities and expressions used in the fMRI experiment: original videos, anticaricatured animations and caricatured animations. Each stimulus was presented three times, sequentially and in random order. Next to each face (Fig. 3A) a drop box was presented, which allowed participants to choose one of the five expression categories, and a slider, which allowed them to rate the face as “convincing” on a 0 to 100 scale. Participants made slider choices using a mouse click (with the chosen numerical rating displayed above) with no initial slider position visible. Participants could advance via button press to the next face after completing categorization and rating.

2.4. Behavioral study 2 procedures

We used motion-based caricature to create a face space with defined predictions for human dissimilarity perception. Behavioral Study 2 tested whether participants perceive the expected dissimilarity that we manipulated between different levels of caricatures, such as those shown by the dendrogram and dissimilarity matrix in Fig. 1B and C (For example, participants should perceive the most dissimilarity between caricatured expression and antiexpression pairs). It also tested predictions, based on the geometry of face space (Fig. 1D and E), about how dissimilar participants perceive pairs of expressions with the same caricature level: Caricatured expressions should be more dissimilar from each other than anticaricatures are from each other. This analysis was based on that of the finding, already-replicated for static shape caricatures, that dissimilarity ratings increase with caricature level (Lee et al., 2000; Irons et al., 2014; McKone et al., 2018). Last, Behavioral Study 2 provided a measure of face space representation that could be detected in fMRI responses using fMRI.

Participants viewed pairs of videos, presented side by side, in a pseudorandom sequence, and then made self-paced ratings of the similarity of each pair by adjusting a slider on a line representing a 100-point scale (similarity ratings were converted to dissimilarity for analysis). The initial slider position was at 50. Attention-check trials randomly appeared four times during each participants’ session. In each attention-check trial, one of the videos was replaced by a heavily-pixelated version of a new video participants did not rate. Participants were instructed to rate all attention-check pairs as “very similar” (i.e., a rating of 100). Participants were attentive (mean accuracy 82%), and the person-total correlation (recommended for checking attentiveness in subjective ratings, especially in data collected online (Curran, 2016; Dupuis et al., 2019), was high (mean r = 0.56). The behavioral study was programmed in jsPsych (De Leeuw, 2015) and delivered on Amazon’s Mechanical Turk to participants in the United States.
each other on every trial. Note that participants never rated pairs where both identities and expressions differ (i.e., there are empty cells in the behavioral dissimilarity matrix in Fig. 4D).

Participants in the “pairs differ in identities” condition saw pairs that were both caricatures, were both anticaricatures, or where one was a caricature and one was an anticaricature (because trials were constrained to only one expression/antiexpression). Participants in “pairs differ in expressions” saw all possible combinations of the four caricature levels for one identity.

2.5. fMRI study procedures

We follow the counterbalancing procedure proposed by Aguirre (2007) for implementing multiple measures of dissimilarity-based brain representations. This procedure, when applied to our paradigm, facilitates concurrent analyses of “direct effects” (univariate responses that track caricature/distinctiveness), repetition suppression related to caricature dissimilarity and RSA. This procedure prevents direct effects from arising due only to repetition suppression of common, typical faces (Aguirre, 2007; Said et al., 2010). Counterbalancing can avoid this problem, if each stimulus condition (e.g., location in the hypothetical dissimilarity space) is equally likely to precede and follow the others. Aguirre recommends an algorithm for ordering counterbalanced sequences (i.e., Type 1 index 1 sequences). For each identity × expression combination, we derived a new Type 1 index 1 sequence (Aguirre, 2007) to counterbalance the four caricature levels, null events (blank screen) and target events (small white plus sign). In response to target events, participants pressed a response pad key with their right index fingers. All events appeared against a black screen for 2000 ms followed by a blank screen for 500 msec. These sequences were concatenated in a pseudorandom order and the total sequence was divided into five scanning sessions of 216 trials each. A structural scan was taken after the third session and localizer scans were run after the second and fourth sessions.

We designed localizer sessions according to previously published methods (Furl et al., 2013) to localize face-selective, motion-sensitive and face-motion-sensitive functional ROIs in individual participants. There were two localizer sessions. In each session, participants viewed four types of block, each containing grayscale presentations of a stimulus category: dynamic faces, objects or static versions of the same faces or objects (taken from the last frame of each video). There were four blocks of each block type per session and block order was pseudo-random. Each block was 11 s. Each stimulus was presented for 1375 ms with a 150 ms inter-block interval containing a fixation cross. Face videos (Van der Schalk et al., 2011) included eight identities (half female), presented in grayscale and transitioning from a neutral expression to either dynamic disgust, fearful, happy or sad. Each block presented all eight identities, with each expression presented twice. Dynamic objects (Fox et al., 2009a) included spinning globes, ceiling fans, machinery, a candle and plants moving in wind. Participants identified via button press when a fixation dot located in the center of each movie turned from white to red on one-third of trials.

2.6. fMRI data acquisition, preprocessing and general analysis

Data were collected on the 3T Allegra scanner (Siemens, Munich, Germany) at Royal Holloway, University of London. Volumetric data included a 1 mm^3 spatial resolution MPRAGE (flip angle 11°; TE 3.03 ms; TR 1900 ms; image matrix 256 × 256 × 176) for each participant. Echo-planar volumes were collected using a high-resolution multi-band sequence (56 slices, 2 mm^3 voxels, TR = 1.2 s, TE = 36.8 ms, FA = 30°). Scans were slice-time corrected, realigned, coregistered to anatomic scans, spatially normalized to the Montreal Neurological Institute (MNI) standardized space and smoothed to 5 mm^3 fullwidth half maximum using conventional procedures in SPM12 (Wellcome Trust Centre for Neuroimaging, UCL, UK). Each first-level, individual-participant, general linear model (GLM) described below used AR(1) correction and a 128 ms high-pass filter. We computed contrasts in first-level GLMs then tested...
Between caricature level average dissimilarity

Within caricature level average dissimilarity

All pairs of videos

Fig. 4. Behavioral Study 2. A, Dendrogram of participants’ between caricature level dissimilarity structure. Compare to Fig. 1B. B, Dissimilarity matrix of participants’ between caricature level dissimilarity ratings. Compare to Fig. 1C. C, Kernel density plots of within caricature level dissimilarity ratings. White circles are mean values. Compare to Fig. 1E. Color scheme in legend is used throughout Figure. D, Matrix of dissimilarity ratings for all pairs of videos used in fMRI study. This matrix is used for RSA of the fMRI data. Pairs in white cells were not rated. E, Multidimensional scaling solution. (A = anger, D = disgust, F = fear, H = happy, S = surprise). Expressions in larger bold font are caricatured and expressions in smaller font are anticaricatured.

them for significance at the group level using mass univariate second-level one-sample t-tests. We implemented multiple comparison correction using best practice for mass univariate procedures and Gaussian random field theory (Brett et al., 2003) (Woo et al., 2014; Eklund et al., 2016). We first identified clusters at a cluster defining threshold of $P < 0.001$ uncorrected and then report and interpret these clusters only when they are also significant at a $P < 0.05$ cluster-level family-wise error rate.

For the localizer runs, the first-level GLMs included four regressors: dynamic and static faces and objects. We localized bilateral face-selective areas Occipital Face Area (OFA) and Fusiform Face Area using a contrast of face-selectivity [dynamic faces, static faces] > [dynamic objects, static objects]. We localized the bilateral motion-sensitive area V5 using the contrast [dynamic faces, dynamic objects] > [static faces, static objects]. We localized the face-motion sensitive area in the right STS using the contrast [dynamic faces > dynamic objects] > [static faces > static objects]. We localized ROIs by searching for the voxel in each individual participant with peak contrast inside a 6 mm radius sphere around the peak identified in the group level contrast. Then we extracted data from 5 mm radius spheres around these new, individual-participant peaks.

2.7. Direct effects

Following Aguirre (2007), we defined direct effects as those where the fMRI univariate response is modulated by the distinctiveness/caricature of stimuli. We entered the four caricature levels as four regressors in a first-level GLM and tested for direct effects with the contrast [caricature, antimovement caricature] > [anti-caricature, antimovement ant caricature]. To control for variations in absolute motion, we used the differences between each parameter (max, slope, mid) and its average to explain absolute parameter magnitudes in three linear regressions. Their residuals (i.e., the variability of absolute magnitude that does not also covary with norm-based differences) were then entered as parametric modulators into the first-level GLM as three nuisance variables. We also entered the caricature dissimilarity with each preceding face (and zero if no preceding face) as a nuisance variable, as prescribed for carryover designs (Aguirre, 2007).

In addition to direct effects of caricature level, we also quantitatively assessed direct effects along the separate spatiotemporal dimensions that we caricatured. The purpose of this analysis was to learn which brain regions exhibit response magnitudes that signal the value of the dimensions of our face space. Such brain regions may contain the information needed to identify a face’s position in a movement face space (i.e., the dimensional values). Related to this purpose, we also tested whether any such signals relate to distance from the norm (as predicted by norm-based face space models), along each dimension or whether they signal the absolute values of each dimension. Movements in the original videos exhibited a sigmoidal pattern (Fig. 2D) which we quantified using logistic model fits as (1) the asymptote or total distance moved (max), (2) the speed or “slope” and (3) the time at which the movement occurred (the “mid” or inflection point). We localized fMRI responses that covaried separately with the degree of caricature (absolute difference from the average) along each of the three dimensions (slope norm, mid norm and max norm) and the absolute magnitude of these dimensions (slope mag, mid mag and max mag). That is, we used one first level GLM with six regressors: each of these six quantities, with the remaining five already partialed out.

2.8. Repetition suppression

In theory (Grill-Spector et al., 2006), neurons selectively tuned to a stimulus dimension produce suppressed fMRI signals when repeated stimuli are similar along the tuned dimension. This allows neurons that code for these dimensions to be localized using fMRI by repeating faces that vary along this dimension. Following Aguirre (2007), we used repetition suppression to localize responses that conformed to the spatiotemporal dissimilarity structure built into our stimulus set. That is, when a face space is defined according to spatiotemporal dimensions, will pairs of faces suppress fMRI responses more when the faces are nearby in face space and less so when the faces are farther apart? This technique has also been previously used with static shape caricatures of artificial faces (Loeffler et al., 2005). For this purpose, we implemented a GLM where all face repetitions in the study (i.e., any event where a face was preceded by another face within its Type 1 Index 1 sequence) were treated as events. These events were parametrically modulated by the dissimilarity with the preceding face (faces could differ by 0, 1, 2 or 3 caricature levels). The results of this “all faces” analysis measure neural responses that code the position along the caricatured trajectory.

To ensure that any findings from the analysis using all face events in fact resulted from dissimilarity between caricature levels and not confounding factors, we performed control analyses. In the “all faces” analysis, caricature level dissimilarity between successive pairs of faces is partially confounded with changes in expression between those faces. This is because the largest caricature dissimilarities always involve a change between a caricatured expression and its caricatured anti-expression counterpart, which appears as a change in expression from one face to the next. However, smaller caricature dissimilarity changes (e.g., from caricature to ant caricature) may or may not also involve an expression change. Thus, neurons with responses tuned to discrete expression categories might give rise to apparent effects of movement-based caricature, even in brain regions that do not represent positions
in movement-based face space. We excluded this possibility by testing whether caricature similarity persisted in suppressing responses whether or not there was an expression change between successive faces. We divided all face events into “between expression” events (the face’s predecessor had a different expression, for example, a transition from caricatured surprise to caricatured antisurprise) and “within expression” events (the face’s predecessor had the same expression, for example, a transition from caricatured antisurprise to anticaricatured antisurprise). In all three analyses, facial identities were always repeated across successive faces. We then re-ran our analysis of caricature dissimilarity separately for between-expression and within-expression face events. If neural responses encode a movement-based face space, then they will be parametrically modulated by caricature dissimilarity in all faces, between- and within-expression analyses. Note that our interest here was in localizing repetition suppression effects of caricature level dissimilarity, rather than repetition suppression effects of expression or identity category, as numerous studies have already done (Winston et al., 2004; Fox et al., 2009b; Cohen Kadosh et al., 2010; Xu and Biederman, 2010; Harris et al., 2014).

2.9. Representational similarity analysis

Aguirre (2007) recommended RSA as a third method that could be used to study dissimilarity spaces using Type 1 Index 1 sequences. We conducted a first-level GLM using a separate regressor for each face presentation on unsmoothed data and implemented RSA on the beta values of these trial-specific regressors using CoSMoMVPA (Oosterhof et al., 2016). RSA proceeded by correlating the matrix of dissimilarity ratings between all the video pairs that we measured in our behavioral study (Fig. 3D) with dissimilarities (one minus Pearson’s r of the trial-specific beta values) among local multivariate response patterns within 6 mm radius searchlights across the whole brain (Kriegeskorte et al., 2006).

As in the repetition suppression analysis, a confound might arise because the largest dissimilarity distances between caricature levels always involves a change between an expression and an antipexpression, while this is not necessarily true for some smaller distances. This makes it possible for brain response patterns that code for differences between expressions and antipexpressions (but do not necessarily instantiate a movement-based face space) to appear significant in this analysis. To address this, we constructed a control dissimilarity matrix that encoded dissimilarity between expression and antipexpression pairs as 1 and zero dissimilarity otherwise. We then used a partial correlation measure in the searchlight analysis to control for nuisance variance with this control matrix when we correlated our behavioral dissimilarity matrix with that of brain responses.

We smoothed these statistical parametric maps over searchlight locations to 4 mm isotropic, full width half maximum and entered them into second level mass univariate t-tests. Searchlight results are reported thresholded at P < 0.05, family-wise error corrected (Brett et al., 2003). We also performed RSA on the response patterns within each ROI separately. These are evaluated using two-tailed t-tests at P < 0.05 Bonferroni corrected.

3. Results

3.1. Behavioral study 1

Participants viewed three video types: the original videos of basic expressions and animated head models where movements could be caricatured or anticaicatured. These animations were also used in the fMRI study and Behavioral Study 2. For each video, participants rated whether the expression was “convincing” and categorized the expression as anger, disgust, fear, happy or surprise. We did not study animations of antimovements, as they do not have either corresponding original videos or “correct” expression labels, which could be used as a basis for comparison.

For all expressions except happy (Fig. 3B), participants did not rate caricatured animations as any less convincing than original videos. Only for happy expressions did participants perceive original videos as significantly more convincing. Also, as hypothesized, for all five expressions, participants rated anticaricatures as less convincing than both the original videos and the caricatured animation. This pattern was supported statistically by a significant interaction between expression and caricature R(8,472) = 43.046, P < 0.001 and main effects of expression F(4,236) = 50.68, P < 0.001 and video type F(2,118) = 62.37, P < 0.001 and the pattern of post hoc two-tailed t-tests (black lines in Fig. 3B) at P < 0.05 Bonferroni-corrected for 5 expressions × 5 pairs is 15 tests. Note that the post hoc null effects between original videos and caricatured animations were not due to the conservative Bonferroni correction. They remained nonsignificant even without correction (P > 0.13), except for surprise, where caricatured animations were significantly more convincing at an uncorrected level (P < 0.001).

Participants also categorized each of the videos into one of the five expression categories (chance = 1/5 = 0.2). Participants’ hit rates were always above chance, even for the lowest-performing participants in the most difficult, anticaicatured animation condition (Fig. 3C). As hypothesized, participants more accurately recognized original videos and caricatured animations than anticaricatured animations, although hit rates on caricatured animations did not reach the performance level of original videos. This pattern resulted in a main effect of video type F(2,118) = 378.69, P < 0.001 and the pattern of post hoc two-tailed t-tested (Bonferroni corrected for three tests) shown by the black lines in Fig. 3C. Inspection of the behavioral categorization confusion matrices (Fig. 3D) confirms that anticaricatured animations appear more confusable than caricatured animations and original videos (i.e., proportion responses are more distributed throughout cells).

Together, these data suggest that the expression categories in the animated stimulus set are perceived and recognized much like the original videos. More importantly for our hypotheses about the structure of face space (Fig. 1D), and consistent with previous findings using static image shape caricatures of expressions (Calder et al., 2000; Lane et al., 2019), caricatured animations were more convincing and better recognized (i.e., less confusable) than anticaicatured animations.

3.2. Behavioral study 2

We assessed whether participants’ dissimilarity ratings were as predicted by the dissimilarity structure built into our caricatured face space. Our predictions concerned two main comparisons. First, we hypothesized that participants’ perceived dissimilarity between caricature levels, to some degree, will be structured in a way resembling the dendrogram in Fig. 1B and the off-diagonal cells of the dissimilarity matrix shown in Fig. 1C. Our second comparison concerns dissimilarity within each caricature level (i.e., the diagonal of the dissimilarity matrix in Fig. 1E). Here, we predicted that caricatured pairs should have longer distances between themselves than anticaricatured pairs, as predicted by the geometry of caricatured facial expression trajectories (Fig. 1D) and by replicated findings using static shape caricatures (Lee et al., 2000; Irons et al., 2014; McKone et al., 2018).

We averaged all dissimilarity ratings for pairs with different caricature levels. Visual inspection of the dendrogram of these data (Fig. 4A) and the dissimilarity matrix (Fig. 4B) suggests that, qualitatively, the caricature levels are ordered broadly as expected (Compare to Fig. 1B and C). Most importantly, caricatures and anticaicatured caricatures pairs, as expected, are the most distant caricature levels (the dark red cell in the bottom left-hand corner), while smaller distances connect adjacent caricature levels (e.g., orange cells caricatures and anticaicatures). The perceptual behavioral matrix somewhat differs quantitatively from predictions, however, because the distance between anticaricatures and antimaicatures caricatures (i.e., between faces closest to the norm) is smaller than expected. Thus, the main features of our predictions with
respect to between caricature distances/dissimilarity (Fig. 1B and C) are broadly confirmed.

Fig. 4C plots the values on the diagonal of the matrix shown in Fig. 4B (compare to Fig. 1E), in which each value contains the average dissimilarity rating of faces of a given caricature level with other faces at the same caricature level. However, we have separately averaged these on-diagonal values for pairs where expressions differed or pairs where identities differed. For pairs of faces where expressions differed, participants rated caricatured expressions as more dissimilar than ant caricatured expressions, although this effect was not present for pairs that differ in identity. This pattern was demonstrated by a 4 (caricature condition, within-participant) × 2 (pair type, between-participant; either identities or expressions differ across pairs) × 2 (identity or expression category, nested in pair type) ANOVA, with a significant caricature condition × pair type interaction F(3,1694) = 88.34, P < 0.001. Two-tailed post-hoc pairwise t-tests, Bonferroni-corrected for twelve tests at P < 0.05 contrasted the caricature levels separately for each of the two pair type conditions. For pairs that differed in identity, none of the caricature levels showed significant mean differences (Ps > 0.08). However, we found a different pattern for pairs that differed in expression (Fig. 4A). Here, both caricatures and antimovement caricatures showed greater dissimilarity than either ant caricatures or ant movement ant caricatures (all P < 0.001). In contrast, caricatures and antimovement caricatures did not significantly differ in dissimilarity from each other (P = 0.15), nor did the two ant caricature conditions (P = 0.4). This latter finding with respect to pairs that differ in expressions replicates findings reported in Irons et al. (2014) and replicated in McKone et al. (2018) for static shape caricatures and is consistent with expectations about the geometry of face space (Fig. 1D).

Other work (e.g., Lee et al., 2000) has examined both within and between caricature level dissimilarity by visualizing dissimilarity matrices (Fig. 4D) using multidimensional scaling (MDS). Indeed, Lee et al. (2000) used this technique to show that static image shape caricatures of facial identities have longer distances in face space than ant caricatures. We repeated this technique here, after averaging dissimilarities for each expression, using a two-dimensional nonclassical multidimensional scaling with Kruskal’s normalized stress 1 criterion. Visual inspection (Fig. 4E) shows, with respect to between caricature level distances, that caricatured expressions and their ant expression counterparts are the most separated, with each expression occupying the opposite end of the plot as its ant expression. With respect to within expression level distances, faces are more spread out (have larger distances between them) when caricatured than when ant caricatured. MDS, therefore, can be used to visually illustrate that our artificially-manipulated face space broadly influences perception as expected theoretically (Fig. 1) and from previous research with static caricatures (Lee et al., 2000).

### 3.3. fMRI study

A mass univariate analysis (Fig. 5A) revealed direct effects of caricature distance from the norm [ caricatures + antimovements > ant caricatures + antimovement ant caricatures] in two large clusters, one in each hemisphere, spreading throughout lateral and ventral occipitotemporal cortex. The larger cluster (4569 voxels when thresholded at P < 0.001 uncorrected) in the right hemisphere peaked at 50–64 0 MNI in Brodmann area 37. The left hemisphere cluster (3878 voxels) peaked at nearly the same location -58 -60 4. Every functional ROI reached significance using two-tailed t-tests at P < 0.05, even when applying conservative Bonferroni correction for seven ROIs (Fig. 5B). Note that this analysis uses nuisance variables to control for both variation in the absolute amount of motion and repetition suppression effects, as recommended in Aguirre (2007).

In a related analysis, instead of localizing direct effects of deviation from the norm of the overall caricature level, we localized deviations from the norm separately for each motion component dimension (max norm, slope norm and mid norm). We also tested for absolute magnitude (rather than norm-based) coding along these same dimensions (max mag, slope mag and mid mag). This analysis provided further evidence that lateral occipitotemporal regions – especially in a dorsal temporal region in Brodmann area 37 near V5 – contributes to a variety of both norm-based and absolute motion coding. Max norm (Fig. 6A) gave rise to a large right hemisphere dorsolateral occipitotemporal cluster (1162 voxels) peaking at 46–62 2 MNI in Brodmann area 37 and a smaller cluster (640 voxels) peaking in about the same location -42 -76 2 MNI in the left hemisphere. Max norm also produced positive effects in bilateral OFA and V5 (Fig. 6F). Slope norm did not produce any effects in the mass univariate analysis or corrected effects in any ROIs (Fig. 6F). Mid norm (Fig. 6B) gave rise to one right hemisphere cluster, peaking at a similar location as max norm (Fig. 6A) 48–62 4 MNI, although none of the ROIs survived correction (Fig. 6F). Max mag exhibited this similar right hemisphere cluster (Fig. 6C), with a similar peak at 46–68 4 MNI. At a corrected level, right OFA and bilateral V5 exhibited max mag effects (Fig. 6F). Like max norm (Fig. 6A), mid norm (Fig. 6B) and max mag (Fig. 6C), slope mag (Fig. 6D) gave rise to a right hemisphere cluster peaking in a similar region: 44–60 0 MNI, with responses in bilateral V5 reaching significance with correction (Fig. 6F). Mid mag (Fig. 6E) produced the most widespread results throughout bilateral dorsal and ventral lateral occipitotemporal cortex, peaking in a similar place as the other motion dimensions (right hemisphere: 48–66 0 MNI, left hemisphere: -50 -68 6 MNI), and with responses in all ROIs reaching corrected significance levels (Fig. 6F). In summary occipitotemporal cortex, centered on a right lateralized dorsolateral temporal area, perhaps overlapping V5, seems involved in both norm-based and absolute coding for most motion dimensions.

As hypothesized, brain responses exhibited a repetition suppression pattern (Fig. 7) that approximated the dissimilarity structure we built into our caricatured face space. Fig. 7A shows the “all faces” mass-univariate analysis of caricature repetition suppression, which includes all repeated face events, excluding face events following target or null events. Affected regions are similar to those observed for direct effects (Figs. 5 and 6) and include bilateral lateral and ventral occipitotemporal areas with peaks at 54–60 -4 and 30 -62 -14 MNI, right inferior frontal gyrus (Brodmann area 9) with peak 46 6 28 MNI and all functional ROIs except
right STS (Fig. 7D and E).

In addition to demonstrating this primary finding, we also performed control analyses. Results from an analysis using all faces as events (Fig. 7A) may confound the widest caricature level differences with changes between basic expressions versus antiexpressions. So we repeated our analysis of caricature dissimilarity, but separately using only face events involving a change between expressions and Between expressions analyses. Right OFA and FFA responses were suppressed. This suppression cannot be accounted for by an expression change confound. A, Mass univariate analysis that tested for parametric modulation of caricature level dissimilarity with preceding face. This “All faces” analysis included all repeated face events. For display, mass univariate analyses are thresholded at $P < 0.001$ uncorrected but reported clusters were significant at $P < 0.05$ cluster-level family-wise error rate. LH = Left hemisphere, RH = Right hemisphere. B, “Within expression” analysis, which used the subset of face events that were preceded by the same expression. C, “Between expression” analysis, which used the subset of face events that were preceded by different expressions. D, Response magnitudes of different dissimilarity values, where increasing dissimilarity is plotted from left to right for each ROI. Zero represents repetitions of the same caricature level (the most suppressed), 1 represents repetitions of adjacent caricature levels, 2 and 3 represent the number of caricature levels separating faces from their predecessors. Shaded areas represent 95% confidence intervals over participants. Points derive from “all faces” analysis. E, Mean parametric modulation of dissimilarity is plotted for All faces, Within expressions and Between expressions analyses. Right OFA and FFA responses were modulated by caricature repetition, regardless of confounding expression changes. Asterisks represent one-tailed $t$-tests where parametric modulation is greater than zero, Bonferroni-corrected for seven ROIs. Error bars represent 95% confidence intervals.

Small correlations were expected based on previous reports including (for example) split-half Spearman correlations of object recognition similarity fMRI data (Walther et al., 2016). The lower and upper bounds of the estimated noise ceilings for these ROIs, the highest achievable correlation, given noise in the data (Nili et al., 2014), were indeed in this low range: left OFA [0.12 0.24], right OFA [0.13 0.26], left FFA [0.12 0.25], right FFA [0.12 0.25], left V5 [0.11 0.25], right V5 [0.13 0.26], right STS [0.13 0.26].

4. Discussion

How people perceive facial movements like facial expressions is a computationally difficult visual problem. Yet computational frameworks for vision have hardly scratched the surface. Although the face space framework is a popular computational metaphor for human face perception, the possibility of face space representations of facial movements in the human has not yet been tested. Here, we demonstrate converging lines of behavioral and brain imaging evidence for face space...
representations of movements and computation of differences in this space from a central tendency or norm movement. Like previous studies of face space representations of static 2D images, we implemented caricaturing to derive specific predictions about participants’ “representational geometry”, based on a pre-defined hypothetical face space. Our caricatured basic expression animations are reasonable stimuli for this purpose, as participants rated them no less convincing than the original videos the animations derive from and, moreover, the participants categorized even the antiexpressions relatively accurately. These spatiotemporal caricatures, we show, yielded similar spatiotemporal information from natural movements in video and dynamic face caricature (Panis et al., 2011) and the cortical regions we observed across analyses could overlap with non-face related findings like these. Indeed, the direct effects and RSA findings involving V5 might to some degree reflect relatively domain-general motion representations. This possibility is especially prominent for the direct effects of the various motion dimensions, which showed peaks near V5 and/or clusters restricted to its vicinity (Fig. 6A–E). Likewise, the STS manifested in our direct effects analyses and RSA. Some regions in the STS are responsive to biological motion more generally as well as for facial motion (Grosbras et al., 2012; Dasgupta et al., 2017). However, note that our ROI localizer functionally defined V5 and STS only using faces, without contribution of other types of motion. Moreover, face-selective ROIs also played some role in every analysis and were the only ROIs to survive our repetition suppression control analyses (Fig. 7E). Right OFA manifested in every analysis and right FFA also appeared in the direct effects and repetition suppression analyses. Given that face spaces involve coordinate representations of faces over multiple “lower-level” basis dimensions, it is perhaps sensible that multiple forms of representation at different levels, including relatively domain general representations, might contribute to such multi-dimensional representation (Furl et al., 2012). Nevertheless, the computation of deviations from a norm, as observed in our direct effects analyses, implies a degree of category specific processing in the use of a category norm (an average of facial movements). Although the contributions of such a widespread swath of cortex might in principle square with the concept of a face space that has many diverse dimensions, they are more difficult to reconcile with the still-popular “Haxby model” and its more recent variants (Calder and Young, 2005; Haxby and Gobini, 2011). That is, if ventral regions (e.g., FFA) are indeed specialized for processing of invariant information and identity while dorsal areas (e.g., STS) process motion and expressions, one might expect more of a strictly dorsal occipitotemporal contribution here.

Dissimilarity spaces like those we studied here might relate to neural population coding. Populations of cells tuned to values of spatiotemporal dimensions, might give rise to fMRI responses that are organized according to dissimilarity. At this time, it is difficult to know how such neuron-level responses differentially contribute to univariate direct effects, repetition suppression or multivariate dissimilarity patterns. Given our (mainly) convergent results, possibly the same neural populations could contribute to the results across analyses. At the neural level, there is evidence that faces (Tsao and Freiwald, 2006; Freiwald and Tsao, 2009; Chang and Tsao, 2017), as well as other stimuli (Kayaert et al., 2005), might be coded as activity patterns over neurons tuned to component dimensions. Some cells are tuned to the spatial distinctiveness of static stimuli (Kayaert et al., 2005; Leopold et al., 2006), consistent with a norm-based space, but other types of dimensional coding also may be supported in primate face-selective patches (Freiwald and Tsao, 2009). These reports accord with our findings of both norm-deviated and magnitude direct responses to individual motion dimensions.

We attempted to capture relatively high-level spatiotemporal information, involved in tracking gross facial features (corners of mouth, eyebrows, etc.). We therefore cannot generalize our results to all types of spatiotemporal information that humans likely use when they perceive dynamic faces in the wild. Observers in real-world conditions may rely on a diverse array of low- and high-level spatiotemporal information. In future, we would be interested in exploring even more sophisticated quantifications of spatiotemporal information from video, including image-based quantifications such as configurations of motion energy patterns or any of the other myriad measures from computational models developed for computer vision. Our approach follows previous work on tracking motion information from natural movements in video and

Fig. 8. Representational similarity analysis. A, Mass univariate searchlight analysis. Bilateral regions in lateral occipitotemporal cortex and right superior temporal sulcus contained local multi-voxel patterns that reflected perceived dissimilarity between videos. The statistical parametric map shown here is thresholded at P < 0.05 corrected at the family wise error rate. LH = Left hemisphere; RH = Right hemisphere. B, Mean correlations between perceived dissimilarity and response pattern dissimilarity in localizer ROIs. Error bars represent 95% confidence intervals. Asterisks represent correlations that were significant at P < 0.005, two-tailed, Bonferroni-corrected for the seven ROIs.
relating this information to brain responses (Furl et al., 2017; Jabbi et al., 2015; Russ and Leopold, 2015). There have also been attempts to experimentally control high-level motion information by caricaturing tennis serve movements (Pollick et al., 2001), facial speech movements (Hill et al., 2005), point light displays of arm movements (Hill and Pollick, 2000) and point light displays of facial expressions (Pollick et al., 2003).

Researchers have previously studied static image caricatures and spatial dissimilarity space coding by measuring dissimilarity perception (Lee et al., 2000), direct fMRI effects and repetition suppression (Loffler et al., 2005) and RSA (Carlin and Kriegeskorte, 2017). However, studies to date did not apply these methods to study spatiotemporal dissimilarity representations (i.e., “face spaces”) for dynamic faces, as we have here. Although we report diverse new findings, including multiple neural manifestations of dissimilarity-based spatiotemporal spaces, there is no question that more work needs to be done. Studies of social and neural processing should use dynamic faces and improve methods for quantifying and manipulating motion information to better challenge visual cortex in ways that stimuli “in the wild” would.

Funding

This work was supported by The Research Councils of the United Kingdom Economic and Social Research Council (grant ES/101134X/2 to N.F.) and The BIAL Foundation (grant Application A-10293 to N.F.).

CRediT authorship contribution statement

Nicholas Furl: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. Forda Begum: Conceptualization, Software, Investigation, Resources. Justin Sulik: Conceptualization, Software, Formal analysis, Investigation, Data curation, Writing - review & editing. Francesca Pizzorni Ferrarese: Conceptualization, Methodology, Software. Sarah Jans: Software, Resources. Caroline Woolley: Software, Resources.

Acknowledgements

Thanks to Iqra Arshad, Lindsay Dicdote, Alice Duffield, Will Dommet, Andrea-Ioana Filip, Ana Jancik, Sneha John, Alkistis Karagounis, Paayal Ladwa, Olivia Marshall, Kelly Palomino-Rodriguez, Santiago Torchiaro, Elsa Tuomi, Andrei Vizitue and Charlie Wilkinson. Thanks to Binghamton University for assistance with their BU-4DFE face set.

References

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cognit. Sci. 10, 14–23.
Lee, K., Byatt, G., Rhodes, G., 2000. Caricature effects, distinctiveness, and identification: testing the face-space framework. Psychol. Sci. 11, 379–85.
Zhang, X., Mondloch, C.J., Segalowitz, S.J., 2012. The timing of individual face recognition in the brain. Neuropsychologia 50, 14.
Aguirre, G.K., 2007. Continuous carry-over designs for fMRI. Neuroimage 35, 1480–1486.
Brett, M., Penny, W.D., Kiebel, S.J., 2003. Introduction to random effects. Trends Cognit. Sci. 10, 14–23.
Calder, A.J., Young, A.W., 2005. Understanding the recognition of facial identity and facial expression. Nat. Rev. Neurosci. 6, 641
Freiwald, W.A., Tsao, D.Y., Livingstone, M.S., 2009. A face space feature in the macaque superior temporal sulcus. J. Neurosci. 33, 17435–17443.
Furl, N., Hadi-Jouziane, F., Liu, N., Averbek, B.B., Unterberger, L.G., 2012. Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. J. Neurosci. 32, 15952–15962.
Furl, N., Henson, R.N., Friston, K.J., Alder, A.J., 2013. Top-down control of visual responses to fear by the amygdala. J. Neurosci. 33, 17435–17443.
Furl, N., Henson, R.N., Friston, K.J., Alder, A.J., 2015. Network interactions explain sensitivity to faces in the superior temporal sulcus. Cerebr. Cortex 25, 2876–2882.
Furl, N., Lobo, M., Pizzorni-Ferrarese, F., 2017. Low-frequency oscillations employ a general coding of the spatio-temporal similarity of dynamic faces. Neuroimage 157, 486–499.
Grosbras, M., Beaton, S., Eckhoff, S., 2012. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. Hum. Brain Mapp. 33 (2), 431–454. https://doi.org/10.1002/hbm.21222.
Harris, R.J., Young, A.W., Andrews, T.J., 2014. Dynamic stimuli stimulate a categorical representation of facial expression in the amygdala. Neuropsychologia 56, 47–52.
Haxby, J.V., Gobbini, M.I., 2011. Distributed neural systems for face perception. In: Calder, A.J., Rhodes, G., Johnson, M., Haxby, J.V. (Eds.), The Oxford Handbook of Face Perception. Oxford University Press, Oxford.
Hill, W., Poltrock, F., 2000. Exaggerating and reduced differences enhances recognition of individuals from point light displays. Psychol. Sci. 11 (3), 223–228. https://doi.org/10.1111/1467-9280.00245.
Hill, H.C.H., Roje, N.F., Johnston, 2005. A range and domain-specific exaggeration of facial speech. J. Vis. 5, 793–807.
Irons, J., McKenzie, E., Dumbleton, R., Barnes, N., He, X., Provis, J., Ivanovic, C., Kwa, A., 2014. A new theoretical approach to improving face recognition in disorders of central face: caricaturing. J. Vis. 14, 12.
Iz, M.L., Scheinberger, S.B., Kaufmann, J.M., 2017. Caricature generalization benefits for faces learned with enhanced idioms of shape or texture. Cognit. Affect Behav. Neurosci. 17, 185197.
Jabbi, M., Kohn, P.D., Nash, T., Jiani, A., Cootelee, C., Holroyd, T., Carver, F.W., Chen, Q., Crubb, P., Kippenhan, J.S., Robinson, S.E., Coppola, R., Berman, K.F., 2015. Convergent BOLD and beta band Activity in superior temporal sulcus and frontoinsulaic circuitry underpins human emotion cognition. Cerebr. Cortex 25, 1878–1888.
Jiang, F., Blanz, V., O’Toole, A.J., 2006. Probing the visual representations of faces with adaptation: a view from the other side of the mean. Psychol. Sci. 17, 493–500.
Juricic, I., Webster, M., 2012. Selectivity of face aftereffects for expressions and anti expressions. Front. Psychol. 3, 24.
Kaufmann, J.M., Scheinberger, S.B., 2008. Distortions in the brain: ERP effects of caricaturing familiar and unfamiliar faces. Brain Res. 1228, 177–188.
Kayeart, G., Biederer, I., De Beek, H., Vogels, R., 2005. Tuning for shape dimensions in macaque inferior temporal cortex. Eur. J. Neurosci. 22, 212–224.
Kriegeskorte, N., Kievit, R.A., 2013. Representationally geometric: integrating cognition, computation, and the brain. Trends Cognit. Sci. 17, 401–412.
Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U.S.A. 103, 3863–3868.
Lane, J., Robbins, R.A., Rohan, E.M.P., Crookes, K., Esse, R.W., Maddess, T., Sabeti, F., Mazlin, J.L., Irons, J., Gradden, T., Dawel, A., Barnes, N., He, X., Smithson, M., McKenzie, E., 2019. Caricaturing can improve facial expression recognition in low-resolution images and age-related macular degeneration. J. Vis. 19, 18.
Lee, K.J., Perrett, D.I., 2000. Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments. Perception 29, 1291-1312.

Leopold, D.A., O’Toole, A.J., Vetter, T., Blanz, V., 2001. Prototype-referenced shape encoding revealed by high-level aftereffects. Nat. Neurosci. 4, 89-94.

Leopold, D.A., Bondar, I.V., Giese, M.A., 2006. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442, 572-575.

Litman, L., Robinson, J., Abberbock, T., 2017. TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. Behav. Res. Methods 49, 423-442.

Loffler, G., Yourganov, G., Wilkinson, F., Wilson, H.R., 2005. fMRI evidence for the neural representation of faces. Nat. Neurosci. 8, 1386-1390.

McKone, E., Robbins, R.A., He, X., Barnes, N., 2018. Caricaturing faces to improve identity recognition in low vision simulation: how effective is current-generation automatic assignment of landmark points? PloS One 13, e0204361.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. PLoS Comput. Biol. 10, e1003553.

Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. GoSmMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Front. Neuroinf. 10, 27.

Panis, S., Wagemans, J., Op de Beeck, H.P., 2011. Dynamic norm-based encoding for unfamiliar shapes in human visual cortex. J. Cognit. Neurosci. 23, 1829-1843.

Pollick, F.E., Fidopiastis, C., Braden, V., 2001. Recognising the style of spatially exaggerated tennis serves. Perception 30, 323-338.

Pollick, F.E., Hill, H., Calder, A., Paterson, H., 2003. Recognising facial expression from spatially and temporally modified movements. Perception 32, 813-826.

Russ, B., Leopold, D., 2015. Functional MRI mapping of dynamic visual features during natural viewing in the macaque. Neuroimage 109, 84-94.

Said, C.F., Dotsch, R., Todorov, A., 2010. The amygdala and FFA track both social and nonsocial face dimensions. Neuropsychologia 48, 3596-3605.

Skinner, A., Benton, C., 2010. Antip-expression aftereffects reveal prototype-referenced coding of facial expressions. Psychol. Sci. 21, 1248-1253.

Tao, D.Y., Frieswal, W.A., 2006. What’s so special about the average face? Trends Cognit. Sci.

Valentine, T., 1991. A unified account of the effects of distinctiveness, inversion, and race in face recognition. Q. J. Exp. Psychol. 43, 161-204.

Valentine, T., 1996. Face-space models of face recognition. J. Math. Psychol. 42, 495.

Van der Schalk, J., Hawk, S.T., Fischer, A.H., Doosje, B.J., 2011. Moving faces, looking places: the Amsterdam dynamic facial expressions set (ADFES). Emotion 11, 907-929.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., Diedrichsen, J., 2016. Reliability of dissimilarity measures for multivoxel pattern analysis. Neuroimage 137, 188200.

Winston, J.S., Henson, R.N., Fine-Goulden, M.R., Dolan, R.J., 2004. fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J. Neurophysiol. 92, 1830-1839.

Woo, C.W., Krishnan, A., Wagner, T.D., 2014. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. Neuroimage 91, 412-419.

Wuttke, S.J., Schweinberger, S.R., 2019. The F200 predominantly reflects distance-to-norm in face space whereas the N250 reflects activation of identity-specific representations of known faces. Biol. Psychol. 140, 86-95.

Xu, X., Biederman, I., 2010. Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. J. Vis. 10, 36.

Yin, L., Chen, X., Sun, Y., Worm, T., Reale, M., 2008. A high-resolution 3D dynamic facial expression database. In: Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition.