Behavioral/Cognitive

A Neural Mechanism of Social Categorization

Ryan M. Stolier and Jonathan B. Freeman
Department of Psychology, New York University, New York, New York 10003

Humans readily sort one another into multiple social categories from mere facial features. However, the facial features used to do so are not always clear-cut because they can be associated with opponent categories (e.g., feminine male face). Recently, computational models and behavioral studies have provided indirect evidence that categorizing such faces is accomplished through dynamic competition between parallel, coactivated social categories that resolve into a stable categorical percept. Using a novel paradigm combining fMRI with real-time hand tracking, the present study examined how the brain translates diverse social cues into categorical percepts. Participants (male and female) categorized faces varying in gender and racial typicality. When categorizing atypical faces, participants' hand movements were simultaneously attracted toward the unselected category response, indexing the degree to which such faces activated the opposite category in parallel. Multivoxel pattern analyses (MVPAs) provided evidence that such social category coactivation manifested in neural patterns of the right fusiform cortex. The extent to which the hand was simultaneously attracted to the opposite gender or race category response option corresponded to increased neural pattern similarity with the average pattern associated with that category, which in turn associated with stronger engagement of the dorsal anterior cingulate cortex. The findings point to a model of social categorization in which occasionally conflicting facial features are resolved through competition between coactivated ventral–temporal cortical representations with the assistance of conflict-monitoring regions. More broadly, the results offer a promising multimodal paradigm to investigate the neural basis of "hidden", temporarily active representations in the service of a broad range of cognitive processes.

Key words: **categorization; conflict monitoring; dorsal anterior cingulate; dynamic competition; face perception; fusiform**

Significance Statement

Individuals readily sort one another into social categories (e.g., sex, race), which have important consequences for a variety of interpersonal behaviors. However, individuals routinely encounter faces that contain diverse features associated with multiple categories (e.g., feminine male face). Using a novel paradigm combining neuroimaging with hand tracking, the present research sought to address how the brain comes to arrive at stable social categorizations from multiple social cues. The results provide evidence that opponent social categories coactivate in face-processing regions, which compete and may resolve into an eventual stable categorization with the assistance of conflict-monitoring regions. Therefore, the findings provide a neural mechanism through which the brain may translate inherently diverse social cues into coherent categorizations of other people.

Introduction

Humans naturally sort the world into categories to "provide maximum information with the least cognitive effort" about its myriad contents (Rosch, 1978; p. 28). In the case of other people, such categorization occurs automatically (e.g., sex or race), in turn activating stereotypes and attitudes that influence interpersonal behavior (Macrae and Bodenhausen, 2000). Although seamlessly categorized, faces across the human population exhibit natural within-category variability and thus vary along relevant social category continua in a graded fashion. Therefore, we frequently encounter faces that vary in their prototypicality (e.g., a female face with masculine features). Once perceived, such within-category variability in facial features can affect the activation of stereotypes and attitudes (Blair et al., 2002) and bear real-world consequences (Galinsky et al., 2013).

Initial research acknowledging the possibility of such gradedness during social categorization argued against simplified binary assumptions (i.e., a category is either activated or not). According to this early account, not only do faces automatically activate a particular social category, but the strength of that activation can vary (Locke et al., 2005). More recent computational models,

Received Oct. 27, 2016; revised April 12, 2017; accepted May 3, 2017.

Author contributions: R.M.S. and J.B.F. designed research; R.M.S. and J.B.F. performed research; R.M.S. analyzed data; R.M.S. and J.B.F. wrote the paper.

This work was supported in part by the National Science Foundation (Grant BCS-1423708 to J.B.F.). We thank Zach Ingbretsen for technical assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Jonathan B. Freeman, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: jon.freeman@nyu.edu.

DOI:10.1523/JNEUROSCI.3334-16.2017

 0270-6474/17/375711-11$15.00/0

such as the dynamic interactive model, posit that multiple social categories are always activated in parallel to some extent, particularly when a given face's features are associated with different categories (Freeman and Ambady, 2011). For instance, a female face with certain masculine features may elicit partial activations of both male and female categories early on. This triggers a dynamic competition process that later stabilizes the category percept over time. Recent behavioral studies have provided evidence for this process through the use of computer mouse tracking, in which the attraction of hand trajectories to unselected response options (en route to a final response option) indexes partial activation of multiple categories during perception (e.g., in the case of gender and race; Freeman et al., 2008). However, it is currently unclear at what level of neural representation social category coactivation manifests and how the brain arrives at stable perceptions from multiple activated social categories.

The fusiform gyrus (FG) and surrounding ventral temporal cortex are involved in face (Haxby et al., 2001) and social category representation (Contreras et al., 2013). If multiple social categories are indeed coactivated by natural mixtures of facial features, then we may expect FG representational patterns of target faces to simultaneously approximate those of two distinct category representations in a graded manner. Indeed, gradations have been observed recently both within (Iordan et al., 2016) and between (Sha et al., 2015) semantic category representations in ventral–temporal cortex and linked to semantic categorizations behaviorally (Ritchie et al., 2015). However, research has yet to investigate how such graded neural representation may involve the competition and resolution of coactivated categories.

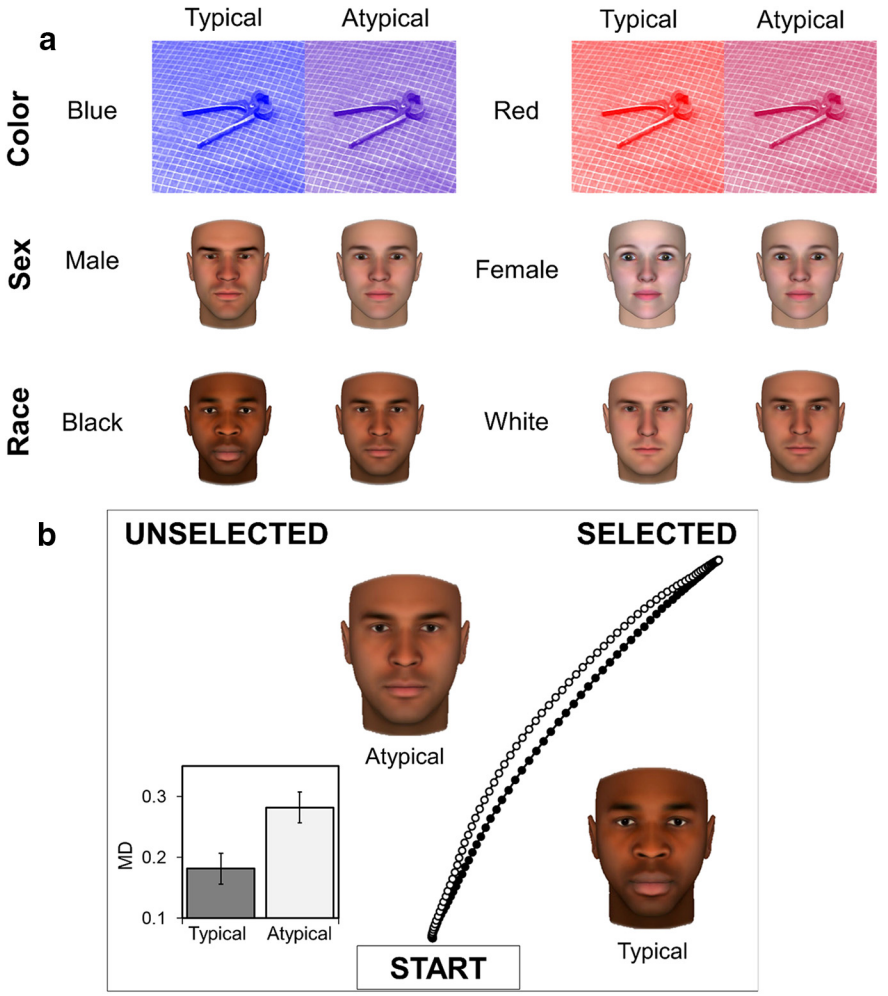

Figure 1. ***a***, Stimuli across all three tasks. Per categorization task (top: color; middle: sex; bottom: race), faces varied from one category to another, being either typical or atypical exemplars of their respective categories. ***b***, Behavioral results ($n = 16$) and example of mouse-tracking paradigm. Results showed significantly higher maximum perpendicular deviation (MD) toward atypical (white line) versus typical (black line) targets across tasks ($b = 0.05$, SE $= 0.006$, $t_{(15.058)} = 7.935$, $p < 0.0001$), demonstrating increased deviation toward the nonchosen response options (i.e., coactivation). Results in the bar plot are depicted to plot condition differences intuitively using within-subject error bars (39); however, the analysis was completed in a multilevel mixed model. In this paradigm, participants clicked a start button, after which the stimulus appeared and they selected response options in either top corners of the screen (e.g., male vs female; white vs black; red vs blue). Mouse trajectories were recorded continuously to observe the amount of deviation toward nonchosen responses and index category coactivation. Typical and atypical black faces are included as examples of these conditions.

Once social categories are coactivated, conflict-monitoring mechanisms are likely important to detect the conflict and help resolve competition between multiply activated representations (Botvinick et al., 2001). Indeed, the resolution of competing perceptual representations is integral to categorization responses (Carlson et al., 2014). A large body of research suggests that such functions may be performed by the cingulo-opercular network (Dosenbach et al., 2006), including the anterior insula/frontal operculum (aI/fO) and, centrally, the presupplementary motor area and dorsal anterior cingulate cortex (pre-SMA/dACC). Neuroimaging studies have suggested that more dorsal components of the extended pre-SMA/dACC region hold a conflict monitoring signal over and above other often confounding processes in nearby regions, namely task difficulty (Neta et al., 2014), arousal (Nachev et al., 2005), and prediction error (Jahn et al., 2016). It has also been proposed that the nature of pre-SMA processing is more cognitive than the motor processing associated with the nearby SMA (Nachev et al., 2008). Relevant to the current research, conflict-monitoring regions have been shown to respond to similar instances of conflictual social category activations, such as cases in which bottom-up facial features are inconsistent with top-down expectations (Hehman et al., 2014). Therefore, here, we hypothesized the engagement of the pre-SMA/dACC by instances of conflicting social categories.

Materials and Methods

The present work sought to test a model of social categorization in which social categories are represented in a sensitive, graded fashion in the FG. In cases of natural inconsistencies frequently encountered in the social world (e.g., feminine male, a white face with black-related features), we predict a corepresentation of conflictual social categories in the FG, which may in turn trigger conflict-related processes in the pre-SMA/dACC and other cingulo-opercular regions that may help to resolve multiply activated social categories into stable perceptions. To test this, we developed a novel paradigm to synchronize fMRI with real-time categorization dynamics assessed by computer mouse tracking.

In the scanner, participants ($n = 16$) made speeded categorizations of the sex (male vs female) and race (white vs black) of typical and atypical face targets and, to assess the domain generality of the effects, of the color (blue vs red) of object targets as well (Fig. 1*a*). Atypical targets were those still reliably perceived as belonging to the correct category yet exhibiting a slight featural resemblance to the opponent category (e.g., male face with feminine features; Fig. 1*a*). We hypothesized that right FG (Kanwisher et al., 1997; Haxby et al., 2001; specializing in face perception) multivoxel patterns should show evidence of social category representation, including the corepresentation of conflictual categories, whereas the pre-SMA/dACC and cingulo-opercular regions should exhibit a stronger engagement during such conflicts.

To investigate the coactivation of opponent categories during perception, we measured online mouse trajectories and blood oxygenation-level-dependent (BOLD) responses from subjects performing a categorization task during fMRI. We used a fiber-optic computer mouse system (NataTech) to allow subjects to operate the mouse in the scanner environment. Pretesting ensured this introduced negligible motor artifacts. The experiment involved three separate 2 × 2 within-subjects design categorization tasks: gender (gender: female vs male × typicality: typical vs atypical), race (race: black vs white × typicality: typical vs atypical), and color (color: blue vs red × typicality: typical vs atypical). Participants also completed a standard demographic survey.

Subjects

Sixteen adult subjects were recruited from the Dartmouth College undergraduate student community (62% male; $M_{age} = 19.37$; 6 white, 6 Asian, 2 black, 2 other; all right-handed; number of subjects based upon recent studies applying similar methods to study face processing; Ratner et al., 2013; Watson et al., 2014; Stolier and Freeman, 2016). Subjects were financially compensated or received partial course credit for participation. Before the study, subjects underwent an informed consent process and screening for fMRI scanning, which was approved by the Committee for the Protection of Human Subjects at Dartmouth College.

Materials

Task. Evidence for social category coactivation has mostly been obtained via a computer mouse-tracking paradigm. In the mouse-tracking paradigm, hand movements en route to response options are recorded such that, in addition to a final categorical response, the hand's attraction toward each response option indexes the extent of its activation. For instance, when categorizing a female face with subtle masculine features, although participants ultimately select the female response, their hand trajectory simultaneously exhibits a partial attraction to select the male response on the opposite side of the screen (Freeman et al., 2008). To date, such parallel attraction effects in mouse-tracking paradigms have successfully provided evidence in favor of category coactivation during the categorization of a face's sex (Freeman et al., 2008), race (Freeman et al., 2010), age (Cloutier et al., 2014), or emotion (Mattek et al., 2016).

The use of hand movements as a continuous index of evolving categorization dynamics is widely supported by neurophysiological research (Cisek and Kalaska, 2005). For instance, in perceptual decision-making tasks in which a monkey commits a response by reaching in one of two potential directions, premotor cortical populations initially tune toward the two response directions simultaneously. As evidence accumulates over time, gradually, the population for the to-be-selected response is amplified, whereas that for the unselected response is suppressed, demonstrating that information about a perceptual decision is made immediately available to the premotor cortex as it accrues, rather than once it has finalized (Cisek and Kalaska, 2005). In humans, event-related potential studies show that ongoing processing results during categorization (e.g., evidence for a male vs. female perceptual target and according response in a categorization task) are immediately and continuously shared with the motor cortices to steer a hand-guided categorical response over time (Freeman et al., 2011b). Such work suggests that a participant's hand motion, as recorded in mouse-tracking paradigms, can reflect dynamic updates of a decision process as it evolves over fractions of a second (Freeman and Ambady, 2010; Hehman et al., 2015). Nevertheless, hand movements in these paradigms cannot definitively rule out the possibility that a more indirect trajectory reflects merely a less decisive movement (i.e., weaker activation of the selected category) rather than genuine parallel attraction (i.e., coactivation; van der Wel et al., 2009) and thus would benefit from converging evidence. Although other methods of detecting competition between response options are also prevalent (e.g., eye tracking), hand tracking is best suited to the current context due to its high temporal resolution (providing an index of response competition ~70 times/s) and its continuous manual data are preferable over discrete oculomotor data for measuring genuine simultaneous activation of multiple response options (Freeman and Ambady, 2010; Freeman et al., 2011).

Participants completed a set of two-choice categorization tasks within the fMRI scanner. The task was designed with MouseTracker software (Freeman and Ambady, 2010). This allowed us to collect online mouse trajectory data in addition to response decision and timing data (Spivey et al., 2005; Freeman et al., 2008; Wojnowicz et al., 2009). Mouse-tracking trials were implemented in a standard two-choice categorization task. Participants were required to make a speeded two-choice categorization decision once the target stimulus appeared. Subjects first clicked a start button at the bottom-center of the screen and then used the mouse to click response options at the top-right and top-left corners of the screen (e.g., male vs female). During this motion online *x*- and *y*-coordinates of the mouse were continuously recorded as the participants responded at ~70 Hz. These data were used later to estimate curvature toward opposing responses (e.g., when categorizing an atypical female face, deviation toward the "male" response en route to a final "female" response). These deviations were used as an index of category coactivation.

Stimuli. Face stimuli were generated with FaceGen Modeler. This software uses a 3D morphing algorithm based on anthropometric parameters of the human population, in which various social category cues can be precisely manipulated while holding other extraneous cues constant. Forty unique face identities were generated for both face categorization tasks (gender and race; Fig. 1*a*). These identities were then morphed to appear as typical and atypical category members (e.g., male and female; white and black). Atypical category members were still reliably recognized as members of their specified category but displayed facial features of the opposing category (e.g., female with slight masculine features). This resulted in a total of 160 face stimuli per task, made up of 4 conditions: category (2: male and female or white and black) × typicality (2: typical vs atypical).

Color stimuli were 40 household object photographs fully tinted to different colors. Object photographs were used so as to have an equal number of exemplars of each category as used in the face tasks. Each object photograph was colored as typical and atypical colors. Specifically, each photograph was tinted as typical red and blue, as well as two colors on the spectrum between red and blue still recognized as their respective color category condition. Therefore, a total of 160 color stimuli were generated, made up of 4 conditions: category (2: blue vs red) × typicality (2: typical vs atypical).

Stimulus condition validation. Given the *a priori* typicality condition labeling of stimuli as typical or atypical based upon parameters in their generation, we collected additional, independent data to validate the assigned typicality of each stimulus. In an independent online sample, we had three groups of participants rate the typicality of each stimulus used in the main imaging study for the color task ($n_{color} = 25$; $M_{age} = 37.84$, $SD_{age} = 13.79$, 40% male, all white), race task ($n_{race} = 25$; $M_{age} = 39.96$, $SD_{age} = 14.63$, 72% male, all white), and sex task ($n_{sex} = 25$; $M_{age} = 40.38$, $SD_{age} = 12.88$, 40% male, all white). Participants were recruited online through Amazon Mechanical Turk and received monetary compensation for their participation. They gave informed consent in a manner approved by the University Committee on Activities Involving Human Subjects at New York University. Participants were presented with each stimulus within their assigned stimulus set from the scanner tasks (160 stimuli per task, as described above) and asked to indicate how "typical" the stimulus appeared of its respective category (e.g., "How typical of the color BLUE is this image?"; 7-point Likert items, spanning 1 "Not at all typical" to 7 "Very typical"). Typicality ratings were median centered within each participant and average typicality scores were computed per stimulus (160 stimuli per task, 480 stimuli in total). To validate

the condition labeling, we regressed the independent sample typicality ratings on the condition labels per stimulus (contrast coded: −1 = typical, 1 = atypical). We found that atypical stimuli were rated as significantly less typical than typical stimuli ($F_{(1,478)} = 956.112, p < 0.00001$). This analysis validates the generation of stimuli varying along the typicality condition.

Procedure

Before beginning the experiment, participants completed a practice shape categorization task (triangles vs squares) in the scanner to familiarize them with the scanner mouse and task. Participants then completed three two-choice categorization tasks during fMRI: race, gender, and color. The task order was pseudorandomized per participant; however, the color task never occurred first. Tasks were completed one at a time, each comprising four sequential functional runs. Each run included 40 total trials with a trial order pseudorandomized optimally for event-related BOLD signal estimation using optseq (Dale, 1999), presenting each task condition 10 times within the run. Therefore, participants completed a total of 40 trials per condition over the course of the experiment. Another 10 trials were null events including a fixation cross to estimate baseline. Trials were 4000 ms in duration, in which participants had up to 2500 ms to provide a response. The stimulus was replaced by a fixation cross after any response or if participants did not respond on time, which remained on screen until the beginning of the next trial. During this period, participants were required to return the mouse to click the "start" button at the bottom of the screen and await the next trial. The next trial was not presented if participants failed to return to the start button on time. After the scan, participants completed a general demographic survey.

Experimental design and statistical analyses

Mouse-trajectory preprocessing. Standard mouse-tracking preprocessing was used (Freeman and Ambady, 2010). All response trajectories were rescaled into a standard coordinate space (top left: [−1, 1.5]; bottom right: [1, 0]) and normalized into 100 time bins using linear interpolation to permit averaging of their full length across multiple trials. For comparison, all trajectories were remapped rightward. To obtain a by-trial index of category coactivation, we calculated the maximum perpendicular deviation (MD) of each mouse trajectory toward the opposite response option. During two-choice mouse-tracking categorization tasks (e.g., male vs female), deviation in a subject's mouse trajectory toward an opposite category response (indexed by MD) is a well validated measure of the degree to which that other category was also activated during the perceptual process (Fig. 2*b*; Spivey and Dale, 2006; Freeman and Ambady, 2010; Freeman et al., 2011a).

Image acquisition. Subjects were scanned using a 3 T Philips Intera Achieva Scanner equipped with a SENSE birdcage head coil in the Dartmouth Brain Imaging Center. All stimuli were back projected onto a screen visible via a mirror mounted on the MRI head coil (visual angle ~13.5 × 13.5°). Anatomical images were acquired using a T1-weighted protocol (256 × 256 matrix, 128 1.33 mm transverse slices). Functional images were acquired using a single-shot gradient echo EPI sequence (TR = 2000 ms, TE = 35 ms). Thirty-five interleaved oblique–axial slices (3 mm × 3 mm × 4 mm voxels; no slice gap) parallel to the AC–PC line were obtained.

Data preprocessing and pattern estimation. Preprocessing of the imaging data was conducted using AFNI software (version 16.0.09; Cox, 1996). Functional imaging data preprocessing included high-pass filtering of frequencies, slice timing correction, 3D motion correction, voxelwise detrending, spatial smoothing using a 3D Gaussian filter (4 mm FWHM for pattern analyses; 8 mm FWHM for univariate ANOVA analyses), and time-series *z*-normalization. Structural and functional data of each subject were transformed to standard MNI space. We estimated the average hemodynamic response per voxel for each condition (using the 3dDeconvolve procedure in AFNI). The design matrix included a total of eight predictors: the four stimulus conditions within each task (typical and atypical conditions per each of the two categories) and several predictors of no interest were modeled as well (incorrect responses, no responses, failed starts, null trials). All predictors were modeled as boxcar functions across the first 2 s of each event (during which the face stimuli were presented) and convolved with a gamma variate function (GAM in AFNI). Trial-by-trial neural response estimates were also performed with 3dDeconvolve and the same response function (GAM) and onset specifications, however fitting a unique regressor per stimulus presentation timepoint (via the stim_times_IM method). For pattern analysis, we used the resulting voxelwise *t*-values (comparing condition responses with baseline) to comprise the whole-brain patterns of activation per stimulus condition (either per run for classification analyses or per trial within each run for trial-by-trial analyses). *t* statistics were used for multivariate pattern analyses because they have been found advantageous in analyses decoding fMRI data and these features were not normalized (Misaki et al., 2010). For univariate ANOVA analyses, we followed standard procedures and used the resulting voxelwise β values (comparing condition-responses to baseline) per condition (either per run for ANOVA analyses or per trial per run for trial-by-trial analyses).

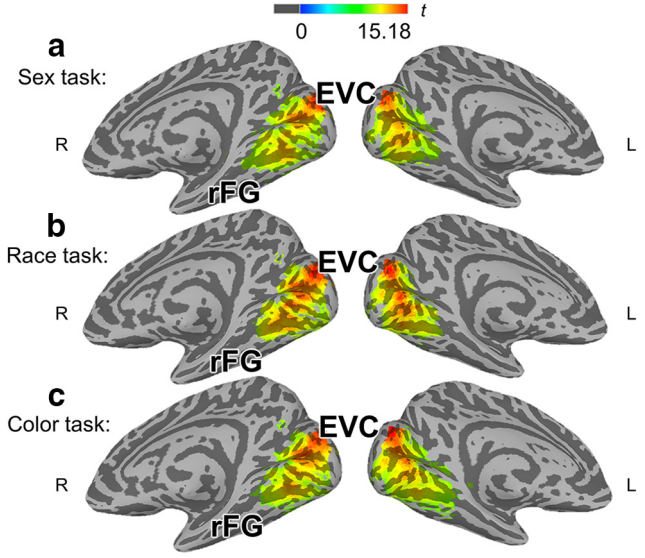

Figure 2. Results from category searchlight classification per task ($n = 16$). *n*-fold cross-validated classification results (using support vector machines) from a searchlight analysis (radius = 3 voxels) were analyzed at the group level, indicating regions where general target category (e.g., female, Black, blue) could be decoded accurately (above chance, i.e., 50% of the time in a 2-way classification analysis). A swath of cortex spanning earlier ventral and dorsal streams from early visual cortex (EVC) through the bilateral fusiform gyrus (FG) was found to hold information about target categories and significantly decode sex, race, and color category membership. Result maps were significance tested and corrected for multiple comparisons using a cluster-wise nonparametric permutation scheme (voxelwise $p < 0.005$; FWE rate of 0.01). All task result maps are depicted on cortical surfaces: ***a***, Sex task. ***b***, Race task. ***c***, Color task.

Multivoxel pattern analyses (MVPAs). All MVPAs were performed using PyMVPA (Hanke et al., 2009). Per task, we performed a two-way classification between each general category condition (i.e., sex: male vs female; race: white vs black; color: blue vs red). Classification was executed with a support vector machine algorithm. All classification analyses were cross-validated in a leave-one-run-out cross-validation scheme ($n - 1$ cross-validation scheme with 4 runs; 2 observations per condition within each run, therefore 6 observations per condition in the training data). These analyses were performed whole brain through a searchlight algorithm (Kriegeskorte et al., 2006). Specifically, cross-validated classification was performed within a 123 voxel sphere (radius = 3 voxels) surrounding each voxel in the brain, with average performance of the classifier mapped back to the center voxel of the sphere. This resulted in a whole-brain map of average classification performance in each task per subject to be submitted to group-level analysis.

Group-level analyses and multiple-comparisons corrections. Whole-brain group-level classification results reported were significance tested and corrected for multiple comparisons using a cluster-wise nonparametric permutation scheme appropriate to MVPA results acquired through a searchlight procedure (Hanke et al., 2009; Stelzer et al., 2013;

Table 1. Occipitotemporal regions of activation (extending bilaterally) elicited by the whole-brain searchlight SVM classification analysis of categories

Task	*x*	*y*	*z*	Mean *t*	Mean accuracy	Voxels
Race	3	−90	18	4.24	64.5%	3853
Sex	3	−87	21	4.37	65.0%	4187
Color	0	−87	18	4.54	65.8%	4977

$p < 0.01$ corrected (voxelwise threshold of $p < 0.005$; FWE rate of 0.01 corrected with a nonparametric permutation and clustering procedure; Stelzer et al., 2013) per task (color: blue vs red, race: black vs white, sex: female vs male). Coordinates reported are cluster peak *t* statistics. Mean accuracy is classification accuracy on average across all voxels in each cluster.

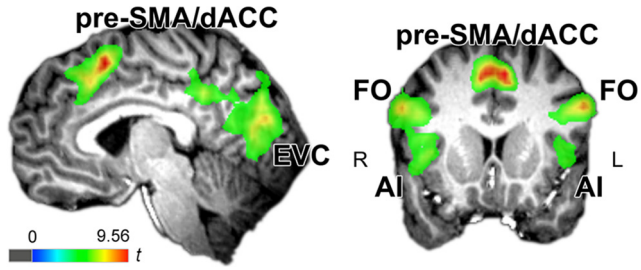

Figure 3. Results from atypical versus typical contrast from whole-brain typicality × task group-level mixed effects repeated-measures ANOVA. Results indicate regions where responses were greater to atypical versus typical category targets on average across tasks (there were no interactions with task across the brain). Interestingly, we found that the cingulo-opercular network, the pre-SMA/dACC and FO/IA, responded more strongly to atypical than typical category members, suggesting potential involvement of conflict-monitoring processes in response to increased category coactivation and competition during atypical target perception. The pre-SMA/dACC cluster was used to construct an ROI for trial-by-trial analyses.

Table 2. Regions of activation elicited by the main effect of typicality in the typicality × task ANOVA

Region	Side	*x*	*y*	*z*	Mean *t*	Voxels
Early visual cortex	M	−3	−84	27	4.39	1038
Frontal operculum/anterior insula	R	36	27	6	5.01	532
Frontal operculum/anterior insula	L	−51	9	36	4.97	470
Pre-SMA/dACC	M	3	15	54	5.29	417
Supramarginal gyrus	R	57	−63	33	4.62	367
Supramarginal gyrus	L	−54	−72	27	4.22	291
Superior frontal gyrus	L	−30	21	54	4.47	183

$p < 0.01$ corrected (voxelwise $p < 0.001$ uncorrected with a minimum cluster extent of 83). There was no evidence of any interactions across the brain between typicality and task. All regions showed greater activation to atypical relative to typical targets.

L, Left; R, right; M, medial.

GroupClusterThreshold in PyMVPA). This algorithm first performed within-subject classification accuracy permutation analyses by generating 100 maps per subject (shuffling classifier labels) and using an identical classifier and cross-validation method as the nonpermuted analyses. Next, a voxelwise cluster-forming threshold for the accuracy map was formed from permutation testing of a group-level map of voxelwise null distributions (feature-wise threshold of $p < 0.005$; 100,000 permuted group-level maps formed via a stratified random sampling bootstrapping process, averaging maps between participants). These thresholded bootstrap samples were then used to derive an empirical probability of various cluster sizes in searchlight classification accuracy maps under the null hypothesis [familywise error (FWE) rate of 0.01]. Results reported are searchlight classification clusters surviving this significance test (Fig. 2, Table 1).

To compare univariate differences in BOLD responses across regions of the brain, a 2 (typicality: typical vs atypical) × 3 (task: sex, race, color) whole-brain mixed-effects ANOVA was conducted ($p < 0.01$, corrected; participant included as a random effect; 3dANOVA2 in AFNI). Furthermore, to further explore fundamental task differences, we contrast coded the main effect of task [1 sex, 1 race, −2 color] in a whole-brain analysis. This analysis provided a whole-brain map of univariate effects per subject to be submitted to group-level analyses (*b*-value maps per typicality, task, their interaction, as well as the contrast coded effect of face vs color tasks). For such univariate analyses, we corrected for multiple comparisons using Monte Carlo simulations (3dClustSim in AFNI; smoothness estimated by a spatial autocorrelation function). We maintained an experiment-wide $\alpha < 0.01$ by using a voxelwise threshold of $p < 0.001$ with a minimum cluster extent of 83. Minimum threshold and cluster extents were those provided by the output of 3dClustSim.

Trial-by-trial analyses. To conduct analyses investigating the close trial-by-trial relationship of mouse trajectories (MDs) and neural responses (multivoxel pattern effects and univariate activation effects), we extracted trial-by-trial estimates of independent neural effects within ROIs identified from each of our whole-brain analyses. From the results of our classification analysis (Fig. 2, Table 1), we segmented an ROI of the right FG (rFG) from result maps per task (sex rFG ROI = 198 voxels, race rFG ROI = 130 voxels; color rFG ROI = 84 voxels). This mask was created as the portion of the result clusters spanning visual regions that was upon the ventral temporal cortex, being an intersection mask of the result map cluster with a ventral temporal lobe mask as defined by the Harvard–Oxford atlas (Jenkinson et al., 2012). From the results of our whole-brain univariate ANOVA analysis (Fig. 3, Table 2), we created a single pre-SMA/dACC ROI from the pre-SMA/dACC cluster responding significantly stronger toward atypical than typical trials (pre-SMA/dACC ROI = 295 voxels). Due to the robustness of ANOVA results, the pre-SMA/dACC cluster that survived significance testing was notably larger than the rFG ROIs at voxelwise $p < 0.005$ used to correct classification analyses (Table 2). Therefore, the pre-SMA/dACC mask was created as the surviving pre-SMA/dACC cluster at a conservative voxelwise $p < 0.0001$ to make the cluster a more reasonably comparable size.

To match neural effects to our trial-by-trial behavioral measures, we extracted them from independent trial-by-trial estimates of neural responding. Specifically, to acquire a trial-by-trial estimate of category coactivation in the rFG, within the rFG ROI per task, we extracted the correlation distance (Kriegeskorte et al., 2008) neural pattern similarity of each atypical trial to the average neural pattern of its opponent category. For instance, on a given atypical male trial, we would estimate the neural pattern similarity of that trial (e.g., during perception of the atypical male) to the average neural pattern of its opponent category (e.g., average of neural pattern to typical female), therefore indexing the degree to which an atypical male elicits a neural response similar to female. To acquire trial-by-trial estimates of potential conflict monitoring activation, for each atypical trial we extracted the pre-SMA/dACC BOLD response estimated in the trial-by-trial general linear model (GLM) (against baseline), indexing the extent of responsiveness trial-by-trial in that region to atypical exemplars.

This provided a dataset in which, for each participant, stimulus, trial, and task, we matched the MD, reaction time (RT), rFG neural pattern similarity to opponent category, and pre-SMA/dACC atypical activation for that trial. These datasets were analyzed in R software (http://www.R-project.org/). We used the multilevel mixed linear model (lmer) from the R package lme4 (Bates et al., 2014). Unstandardized regression coefficients are reported. All variables were normalized to optimize performance of the random slopes algorithm (lmer performs best with similarly scaled variables; Bates et al., 2014; normalized across subjects, each variable mean set to 0 and SD to 1). This same normalization transformation was used in later mediation analyses to maintain a similar metric. Each of these models allowed for random slopes and intercept with specific stimulus identity nested within each participant.

Category competition versus general indecision. Another alternative explanation of increased trajectory deviations is that they merely reflect a less decisive movement toward the selected response rather than a genuine parallel attraction toward the opposite response. For instance, during perception of a less typical white face, participant trajectories may take a less direct route to the white response due to mere indecision (e.g., slower accumulation of evidence in favor of the white category) as opposed to genuine attraction to the alternate (black) category reflecting category coactivation and competition. We conducted an additional behavioral experiment to rule out this potential alternative, in which we included target faces bearing partial cues of both irrelevant and relevant

categories. For instance, in a white versus black categorization task, partial Asian cues on a white face would be task irrelevant, whereas partial black cues on a white face would be task relevant. If participants' trajectory effects merely reflected less decisive movements because both irrelevant and relevant conditions feature the same degree of increased ambiguity and noise with respect to category cues, then they should elicit similar MD effects. If trajectory effects can reflect genuine coactivation and parallel attraction toward the opposite response, then only when facial cues partially specify the opposite category response will trajectories deviate toward that response.

Participants. An additional behavioral experiment was performed in which participants ($n = 49$) were recruited to perform a computer mouse tracking task online ($M_{age} = 33.06$, $SD_{age} = 9.1$; 61.22% male; 71.43% white, 10.2% black, 18.37% other; one participant omitted due to trackpad use). Participants were recruited online through Amazon Mechanical Turk and received monetary compensation for their participation. Subjects gave informed consent in a manner approved by the University Committee on Activities Involving Human Subjects at New York University.

Task. Participants completed three sets of two-choice categorization tasks. The task was designed with MouseTracker (see Materials and Methods; Freeman and Ambady, 2010). Mouse-tracking trials reflect standard two-choice categorization trials. Participants were required to make a speeded two-choice categorization decision once the target stimulus appeared. However, participants first clicked a start button at the bottom of the screen and then used the mouse to click response options at the top-right and top-left corners of the screen (e.g., "black" vs "white"). During this motion, online *x*- and *y*-coordinates of the mouse were recorded continuously as the participants responded. These data were used later to estimate curvature toward opposing responses (e.g., when categorizing an atypical black face, deviation toward the "white" response en route to a final "black" response).

Stimuli. Face stimuli were generated with FaceGen software. Ten unique face identities were generated. For each identity, three base race faces were generated (Asian, black, and white). In addition, for each base race, three race cue conditions were generated: no partial cues (base race) and partial cues of the two other races. This generated a total of 90 (10 identity × 3 base race × 3 race cues) face stimuli. Each face stimulus was placed on a gray background (RGB: 175, 175, 175). Relevance of face partial race cues was determined by task.

Procedure. Participants were instructed to categorize each face according to its perceived race. Each participant completed all three tasks (Asian vs black, Asian vs white, black vs white). Task order was randomly assigned for each participant. Within each task, participants categorized a total of 60 stimuli: per relevant race faces (2), 10 base race faces, 10 relevant cue faces, and 10 irrelevant cue faces. Stimulus presentation was randomly ordered per task.

Analysis preparation. One subject was removed for not following task instructions. Consistent with prior face categorization mouse-tracking work, we limited analysis to correct trials with quick RTs (<2000 ms). The average error rate across subjects was low (2.44% of trials) and nearly all RTs were within the target range (98.8% of trials, <2000 ms). To perform analyses with subject as the unit of analysis, we estimated our measure of category coactivation (MD) for three conditions within each subject: base race, relevant partial cues, and irrelevant partial cues. This analysis was collapsed across task, base race, and partial cues race for parsimony.

Results

Sex, race, and color categorization proceeded in separate runs. During neuroimaging, participants categorized targets by moving a fiber-optic computer mouse. On each trial, a start button appeared at the bottom center of the screen. Once clicked, the target face or object appeared at the bottom center of the screen and participants were asked to click a response at the top-left and top-right corners of the screen as quickly and accurately as possible. The movement trajectory recorded during each trial, including the MD toward the unselected category response (on the opposite side of the screen), indexed coactivation of that category (Fig. 1*b*; Freeman and Ambady, 2010).

Behavioral results

First, we analyzed the relationship between MD and target typicality to assess behavioral measures of coactivation and competition between categories due to partial cues of the opponent category. We used a multilevel random-slopes model to regress MD upon typicality (typical vs atypical; coded −1 and 1, respectively) on a trial-by-trial basis across tasks (allowing for random intercepts and slopes, with face identities nested within subjects; this model structure was used for all subsequent analyses). Consistent with prior mouse-tracking studies involving sex, race, and color categorization (Freeman et al., 2008, 2010; Freeman and Ambady, 2010), there was significantly higher MD during atypical compared with typical trials [$b = 0.122$, SE = 0.015, $t_{(15.058)} = 7.935$, $p < 0.0001$, 95% confidence interval (CI) = 0.091–0.153; Fig. 1*b*], suggesting that mouse trajectories were partially attracted to the opposite category response due to atypical cues related to that category.

Rather than providing evidence for a parallel competition between coactivated categories, increased MD could also be spuriously produced from sequential shifts in movements. Specifically, our prediction is that, throughout the response trajectory, a participant's movement in the atypical condition should always reflect a dynamically weighted combination of movement toward both categories due to parallel activation (e.g., both male and female; both white and black). If true, then the average trajectory in the atypical condition should exhibit graded, partial attraction toward the opposite category en route to the selected category. However, a higher average MD could also be caused by nongraded activations with several discrete-like errors in which, on some trials, one category activates ~100% (straight movement to "female"), followed by a subsequent correction and the other category activating ~100% (straight movement toward "male"). If all trials in the atypical condition exhibited such discrete shifts, then the average trajectory would clearly be shaped as such, which was not the case (Fig. 1*b*). However, if only a subpopulation of trials in the atypical condition exhibited such shifts, it is possible the average trajectory would spuriously exhibit graded, partial attraction, but the amount of attraction (MD) would be bimodally distributed. This is because some trials would involve a shift (i.e., extreme attraction), whereas others would proceed normally with a direct movement (i.e., zero attraction). The modality of the MD distribution was tested with Hartigan's dip statistic, a method found to most reliably distinguish between such discrete-shift versus parallel-attraction trajectory profiles in mouse-tracking experiments (Freeman and Dale, 2013). There was no evidence of multimodality in the MD distribution of the atypical condition ($D_{atypical} = 0.0043$, $p_{atypical} = 0.9831$, n.s.), nor in that of the typical condition ($D_{typical} = 0.0031$, $p_{typical} = 0.9977$, n.s.; $D_{all} = 0.0019$, $p_{all} = 0.9999$, n.s.).

Together, these findings cement the evidence that the trajectory attraction effects observed in the fMRI experiment reflect genuine coactivation of both social categories in parallel, rather than a subpopulation of discrete-like error responses or a mere weaker representation of the selected category and less decisive movement.

Neuroimaging results

Category representational analyses

We first sought to identify regions involved in representing faces' social categories. Rather than assessing response differences be-

tween conditions averaged across voxels within a region, MVPAs may identify regions where conditions reliably elicit distributed multivoxel patterns of local activity, which is often the case in perceptual representation (Haxby et al., 2001). We performed classification to identify regions that could discriminate between overall category conditions above chance (male vs female; white vs black; blue vs red). A searchlight procedure was used, in which classification analyses (via support vector machines) were conducted iteratively in local searchlight spheres throughout cortex (Kriegeskorte et al., 2006). We limited these analyses to occipital and ventral–temporal cortex given their well established role in perceptual representation of such stimuli (Kanwisher et al., 1997; Haxby et al., 2001). Searchlight analysis (FWE rate of 0.01, corrected) revealed a single broad swath spanning occipital and early ventral–temporal cortex, including the FG, which exhibited above-chance classification accuracies in all three tasks (Fig. 2, Table 1). These findings are consistent with prior research observing categorical representation of faces and colors in these regions (Brouwer and Heeger, 2013; Contreras et al., 2013; Freeman and Johnson, 2016; Stolier and Freeman, 2016).

Having confirmed that faces' sex and race categories were indeed reflected in multivoxel patterns of the FG and other ventral–temporal regions involved in perceptual representation, we next tested whether the multivoxel pattern elicited by a given face approximates that associated with the face's opponent category to the extent that the face bears cues associated with that category. Specifically, we were interested in whether the extent of coactivation of an opponent category (e.g., female for a male face), as measured by MD, is associated with a stronger approximation in right FG multivoxel patterns toward those associated with that opponent category (e.g., female). To do so, we first estimated the neural pattern representational similarity (Kriegeskorte et al., 2008) of each trial category representation to its opponent category's. This was calculated as the similarity of each atypical trial voxel pattern to the average voxel pattern of the opponent category (Pearson correlational distance; $1 - r$). For instance, we calculated the representational similarity of each atypical male trial voxel pattern to the average voxel pattern of the typical female condition. These trial-by-trial data were calculated within the rFG region elicited by our searchlight classification analysis (see Materials and Methods). Whereas the whole-brain searchlight demonstrated the discriminability of categories collapsing across typicality conditions (e.g., male vs female regardless of typicality), here, we performed an independent analysis within the atypical condition. Specifically, we assessed the trial-by-trial relation of each atypical exemplar's neural-pattern similarity with that trial's associated MD, an analysis that is statistically independent from the initial whole-brain analysis (overall discriminability of categories within each subject). Using multilevel regression that can incorporate trial-by-trial data (performed in R with lme4), rFG neural category similarity values were regressed upon MD, finding a significantly positive relationship between rFG pattern similarity and MD ($b = 0.015$, SE $= 0.003$, $t_{(17.284)} = 4.427$, $p = 0.0004$, 95% CI $= 0.008–0.023$). Therefore, to the extent that participants were partially attracted to the alternate category response behaviorally (e.g., toward "male" for a female face), rFG patterns exhibited a degree of greater similarity to that alternate category (e.g., male).

Category competition analyses

To identify neural regions responsive to the extent of coactivation and competition, we first conducted a 2 (typicality: typical vs atypical) × 3 (task: sex, race, color) whole-brain mixed-effects ANOVA ($p < 0.01$, corrected; participant included as a random effect; 3dANOVA3 in AFNI). This revealed a significant main effect of typicality, with stronger BOLD responses to atypical versus typical target faces in the cingulo–opercular network, including the pre-SMA/dACC and aI/fO (Fig. 3, Table 2). No regions were elicited by the typicality × task interaction effect that survived correction ($p < 0.01$, corrected).

More importantly, we sought to further characterize the nature of the pre-SMA/dACC's stronger responses to atypical exemplars. Although suggestive that the pre-SMA/dACC may be involved in monitoring for conflicting coactivations that are more present in the atypical condition, stronger evidence in support of this hypothesis is that pre-SMA/dACC response in this context is specifically responsive to category coactivation; if so, then pre-SMA/dACC activation should correlate with MD on a trial-by-trial basis. Using the pre-SMA/dACC region elicited by the previous whole-brain analysis (see Materials and Methods), trial-level mean responses within this ROI were calculated to examine the relationship of pre-SMA/dACC activity with trial-by-trial behavioral indices of category competition (MD). A separate GLM was constructed to estimate mean pre-SMA/dACC response for each trial (see Materials and Methods). Within this ROI of the pre-SMA/dACC, we performed an analysis wherein trial-by-trial neural responses for only atypical targets were extracted and their relationship with trial-by-trial MD was tested. Note that this represents an independent analysis from the initial whole-brain contrast of atypical > typical.

Consistent with our hypothesis, regressing pre-SMA/dACC activation on MD in a trial-by-trial fashion indicated a significantly positive relationship ($b = 0.120$, SE $= 0.025$, $t_{(14.552)} = 4.76$, $p = 0.0003$, 95% CI $= 0.081–0.196$). To more stringently assess the nature of the pre-SMA/dACC response, we also performed this analysis while additionally controlling for an alternative explanation of pre-SMA/dACC responses, namely mere task difficulty (i.e., RTs) and motor effort (i.e., total hand motion and force; summated velocity and absolute acceleration across all time points in the mouse-trajectory data per trial). If pre-SMA/dACC responsiveness reflected task difficulty alone rather than genuine coactivation, then controlling for trial-by-trial RT should eliminate the correlation with MD. In addition, if pre-SMA/dACC responsiveness reflected mere motor effort due to task demands of more atypical trials (given the putative recruitment of regions surrounding the pre-SMA in motor preparation as well; Nachev et al., 2008), then controlling for trial-by-trial motor effort indices should eliminate the relationship with MD. Indeed, prior work has identified a signal for conflict processes separable from other pre-SMA/dACC signals such as task difficulty through such covariate analyses (Neta et al., 2014). We regressed pre-SMA/dACC activity upon MD controlling for RT (task difficulty), summated *x*-/*y*-axis velocity (total motion), and summated absolute acceleration (total force), again finding a significant positive relationship between pre-SMA/dACC activation and MD ($b = 0.083$, SE $= 0.029$, $t_{(23.2)} = 2.884$, $p = 0.0083$, 95% CI $= 0.027–0.143$). These findings suggest that the pre-SMA/dACC is specifically responsive to competition between opposing social category responses above and beyond the mere difficulty and motor effort of the categorization (Nachev et al., 2008; Jahn et al., 2016), which is consistent with certain prominent accounts of this region.

Mediation analysis

Last, if competition between coactivated categories in the rFG elicits conflict-monitoring processes in the pre-SMA/dACC, then we may expect increased pre-SMA/dACC activity on trials with greater rFG category coactivation. Specifically, we tested a mediation model in which MD (behavioral index of competition) mediates the relationship between rFG pattern similarity (category coactivation) and pre-SMA/dACC activity. We tested this model with the multilevel approach put forth by Bauer et al. (2006), which uses a Monte Carlo simulation (10,000 iterations) to estimate 95% CIs for the total and indirect effect. Consistent with the previous analyses, there was a significant total effect ($p < 0.0001$, 95% CI = 0.24235–0.36201), with a positive relationship between rFG category coactivation and pre-SMA/dACC activation. More importantly, there was a significant indirect effect of MD ($p = 0.005$, 95% CI = 0.0031–0.01563), supporting our hypothesis that the positive relationship between rFG pattern similarity and pre-SMA/dACC responses may partly be accounted for by the competition between category activations (Fig. 4; $b_{\text{indirect effect}} = 0.00893$, $SE_{\text{indirect effect}} = 0.00321$, $p_{\text{indirect effect}} = 0.005$, 95% $CI_{\text{indirect effect}} = 0.00263–0.01523$; $b_{\text{path a}} = 0.08029$, $SE_{\text{path a}} = 0.01745$; $b_{\text{path b}} = 0.1043$, $SE_{\text{path b}} = 0.02374$; $b_{\text{path c}} = 0.30237$, $SE_{\text{path c}} = 0.03056$, $p_{\text{path c}} < 0.0001$, 95% $CI_{\text{path c}} = 0.24246–0.36227$]; $b_{\text{path c'}} = 0.2934$, $SE_{\text{path c'}} = 0.0301$). Together, this suggests that competition between category coactivations in the rFG may trigger stronger responses in the pre-SMA/dACC that is then engaged to help resolve such conflict.

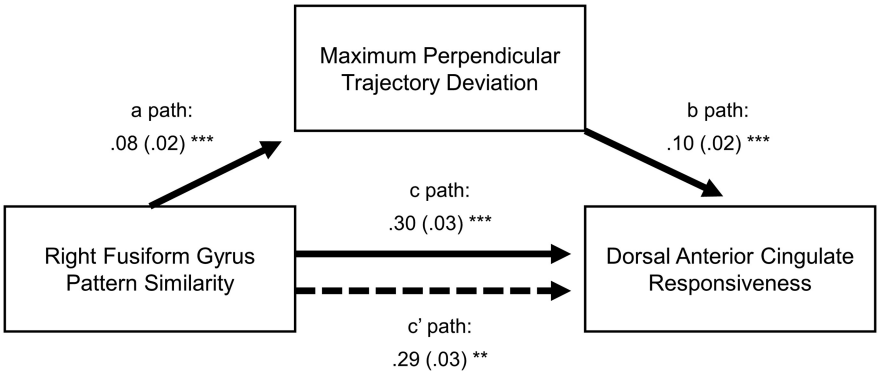

Figure 4. Results from multilevel mediation analysis. A significant indirect effect of category coactivation was observed ($p = 0.005$, 95% CI = 0.0031–0.01563), where the relationship between rFG pattern similarity and pre-SMA/dACC activation (total effect: path c; $b_{\text{path c}} = 0.30237$, $SE_{\text{path c}} = 0.03056$, $p_{\text{path c}} < 0.0001$, 95% $CI_{\text{path c}} = 0.24246–0.36227$) was partly accounted for by the extent of category competition as measured behaviorally (MD; reduced effect: path c'; $b_{\text{path c'}} = 0.2934$, $SE_{\text{path c'}} = 0.0301$). This result suggests that visual representations approximating the target category (e.g., male) and its competitor (e.g., female) simultaneously may lead to increased category competition, which in turn leads to stronger engagement of the pre-SMA/dACC. Results were significance tested with a Monte Carlo simulation (10,000 iterations) to estimate confidence intervals for the total and indirect effect (for other effects, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.0001$). Unstandardized betas and their SEs are reported per path.

Category competition versus general indecision

Last, we analyzed our additional data to rule out the alternative explanation that increased trajectory deviations are merely due to less decisive movement toward the response. We used a repeated-measures ANOVA with Helmert coding to make two primary comparisons: (1) relevant cues versus both irrelevant cues and base race and (2) irrelevant cues versus base race. We found relevant cues to elicit significantly greater MD (M = 0.509) than other conditions on average (M = 0.475; $F_{(1,48)} = 7.006$, $p = 0.01$). Pairwise comparisons revealed that, whereas relevant cues (M = 0.509) elicited significantly higher MD than both irrelevant cues (M = 0.478; $t_{(48)} = 2.214$, $p = 0.034$) and base race (M = 0.472; $t_{(48)} = 2.567$, $p = 0.014$), there was no evidence of a difference between irrelevant cues and base race ($t_{(48)} = 0.486$, $p = 0.629$).

These results show that relevant partial other race cues elicited increases in MD relative to no partial cues and irrelevant partial cues, whereas irrelevant partial cues did not elicit MDs relative to no partial cues. Therefore, only when a face bore cues specifying the alternate category did participants exhibit a parallel attraction to that category; mere ambiguity or uncertainty was not sufficient to elicit MD effects. These findings confirm the specific sensitivity of MD to task-relevant cues and category coactivation, rather than mere indecision.

Discussion

Through the integration of mouse-tracking and neuroimaging, we found convergent evidence that multiple social categories are simultaneously and partially activated during the perception of faces whose features partly overlap with other categories (e.g., feminine male face). Specifically, while categorizing faces' gender or race, or objects' color, participants' hand trajectories partially deviated toward the opposite category response when a target's features resembled that category. Such results are evidence for coactivation between social categories that compete over time, consistent with previous behavioral studies (Freeman and Johnson, 2016), and additional analyses ruled out alternative explanations such as discrete-like errors or mere indecisiveness. Neuroimaging results demonstrated that such social category coactivation was reflected in the similarity of multivoxel patterns in the rFG and ventral visual stream. To the extent that participants' response trajectory exhibited a parallel attraction to an opponent social category (e.g., male for a feminine male face), the elicited rFG pattern was correspondingly more similar to the average pattern associated with that opponent category. In turn, this increased pattern similarity effect predicted stronger overall pre-SMA/dACC activation, likely reflecting the pre-SMA/dACC's role in detecting and resolving the category inconsistency. Moreover, such results were not limited to the social domain (sex, race), instead generalizing across nonsocial categorization as well (color).

The present findings bolster previous behavioral studies suggesting that perceivers translate a natural spectrum of facial cues into stable categorizations of other people through a competition process involving multiply activated categories (Freeman and Johnson, 2016), as proposed by recent computational models of social categorization (Freeman and Ambady, 2011). Earlier accounts proposing activation of a single category representation of variable strength (Locke et al., 2005) are not well accommodated by the findings. Comparisons of neural responses across social and nonsocial categorization strongly converged, supporting accounts that social categorization draws on domain-general mechanisms suited to other forms of perceptual categorization (Freeman and Ambady, 2011). Indeed, dynamic competition between coactivated representations is prevalent across domains,

such as motor decision making (Cisek and Kalaska, 2010), object recognition (O'Reilly et al., 2013), and language processing (Spivey et al., 2005), and may therefore reflect a common computational property of cortical representation (Cisek and Kalaska, 2005, 2010; Rolls, 2000). Therefore, we believe that it is important to highlight that the results here likely reflect domain-general properties of perception and cognition. Indeed, prior work has found ventral-temporal cortical representation to cluster category members conceptually (Connolly et al., 2012) and centrally around intracategory norms (Leopold et al., 2006). Our findings extend this work, suggesting a process by which such representations may approximate other categories during competition and the role of conflict-monitoring and top-down modulation of the pre-SMA/dACC in assisting this competition.

The current results have several implications for understanding ventral–temporal representation. The ventral visual cortex has been long known to house distinct information about visual categories (Goodale and Milner, 1992; Haxby et al., 2001). Recent research has uncovered much about this process, such as selectivity for prominent categories (faces, places, and bodies; Vul et al., 2012), conceptual organization by animacy (Connolly et al., 2012), and predominance of basic-level category structure (Iordan et al., 2015). However, research has yet to probe directly how category representation unfolds and is influenced by other categories inherently in competition with the target. One recent study found representational distance of visual objects from a representational decision boundary (animate vs inanimate) to predict behavioral animacy categorization RTs (Carlson et al., 2014). This finding shows how representational space relevant to a more abstract category (animacy) may underlie perceptual decision making. This speaks largely to the viability of perceptual decision making having a basis in earlier perceptual cortices (cf. Freedman and Assad, 2011). However, the current findings provide a novel demonstration of how categories are represented in relation to one another, specifically in relation to perceptually competing categories. This speaks beyond explicit perceptual decisions and shows that competitive components of behavioral responses over and above explicit decisions are manifest in ventral–temporal neural pattern similarity.

The dACC has been shown to play an important role in the detection and signaling of information-processing conflicts (Botvinick et al., 1999). The pre-SMA/dACC is considered part of a wider ranged cingulo-opercular network entailing the pre-SMA/dACC and aI/fO (Neta et al., 2014). In the case of categorization, prior research has found responsiveness of these regions to categorization uncertainty (Grinband et al., 2006) and the presence of stereotypically incongruent social categories (Hehman et al., 2014; Cassidy et al., 2017). However, pre-SMA/dACC function has been a large topic of debate due to competing explanations. A recent study assessed the contribution of multiple accounts to pre-SMA/dACC responsiveness: ambiguity (i.e., conflict), accuracy, and RT (Neta et al., 2014). The investigators found that each separately contributed to the regional response. In our analyses, we focused on correct categorization trials and statistically controlled for RT. In addition, we used a unique behavioral index of category competition (or conflict), MD, directly measuring the tentative commitment to a response that was considered but not ultimately selected. The results therefore suggest that pre-SMA/dACC response in this context is more indicative of conflict-monitoring functions triggered by multiple social category coactivation rather than mere difficulty or ambiguity (i.e., RT). Therefore, the data bolster findings and theory of unique conflict-monitoring functions in this region (Botvinick et al., 1999; Botvinick et al., 2001; Neta et al., 2014) and suggest this a response to category competition.

More specifically, one perspective is that, when considerable conflict between representations is detected, the pre-SMA/dACC may be recruited to assist the category competition performed in ventral–temporal regions by directing more cognitive resources (Narayanan et al., 2013; Cavanagh and Shackman, 2015) or increasing attention toward relevant stimulus properties and away from irrelevant stimulus properties (Sheth et al., 2012; Oehrn et al., 2014; Ullsperger et al., 2014; Tang et al., 2016). This is consistent with recent theories on this region's function in the expected value of control more generally (Shenhav et al., 2013). Computational models such as attractor neural network models suggest that ventral–temporal regions such as the fusiform cortex alone could force partially active social category representations to compete via lateral inhibition and nonlinear dynamics that do not require any outside mechanism (Rolls, 2000; Usher and McClelland, 2001; Freeman and Ambady, 2011; Wyatte et al., 2012). Accordingly, the pre-SMA/dACC may not be involved in intervening directly on ventral–temporal competitive dynamics, but may serve an important role in directing critical cognitive and attentional resources needed to more rapidly resolve the competition and adapt ventral–temporal regions to the most diagnostic perceptual cues for the task at hand. Such bidirectional interaction between these regions could be supported by their well documented structural and functional connectivity (Dosenbach et al., 2007; Shenhav et al., 2013). Although speculative, future research could test these issues directly.

The novel paradigm of synchronized neuroimaging and mouse tracking has numerous implications for understanding the neural basis of "hidden" response activation that may not manifest in an explicit behavioral response, whether in social categorization or otherwise. Mouse tracking without neuroimaging has now been leveraged in numerous domains across the cognitive sciences, including moral decision making (Koop, 2013), numerical cognition (Faulkenberry, 2016), perceptual decision making (Lepora and Pezzulo, 2015), memory encoding (Papesh and Goldinger, 2012), emotional processing (Schneider et al., 2015), and self-control (Ha et al., 2016). In the present work, the trajectory data allowed us to disambiguate parallel competitive from discrete responses, measure a more direct index of category competition (over and above RT), and collect data along the time course of each response. Furthermore, through the integration of recent MVPA approaches (RSA; Kriegeskorte et al., 2008), we were able to predict similarity of neural patterns on a trial-by-trial basis (Carlson et al., 2014). This technique could also potentially lock neural responses to different temporal components of motion trajectories, allowing a degree of temporal precision unprecedented in fMRI research due to the temporal resolution of the method. We believe the collection of dynamic behavioral data during neuroimaging could be of great promise to a number of research areas.

Nevertheless, there are several limitations of the current research. Our ability to make inferences about neural responses from correlational data limits our interpretations of these findings (Poldrack, 2006). Our understanding of these complex networks will rely on the accumulation of convergent evidence from multiple measurement methodologies, especially along with experimental designs (e.g., TMS; cognitive control manipulations). Moreover, although we are interested in a dynamic process that unfolds rapidly and uses mouse tracking to gain insight into this process, we cannot make strong inferences about the temporal dynamics of the neural response involved due to limited tempo-

ral resolution of fMRI. Therefore, in modeling the interplay of the rFG and the pre-SMA/dACC, the statistical mediation is suggestive of a possible causal chain of events, but the correlational nature limits any strong inference. Despite this, the role of rFG pattern-similarity in representing social categories, including coactivated categories, and the pre-SMA/dACC and cingulo-opercular response to such pattern similarity effects in the presence of coactivated categories is clear.

In summary, the members of any social category that we encounter are rarely a perfect prototype of that given category. Not only do they deviate in the degree of their membership, but also to the extent that their cues relate to alternative, opponent categories. Although recent computational models and behavioral studies have suggested that this leads to the coactivation and competition of potential categories (Freeman and Ambady, 2011; Freeman and Johnson, 2016), the neural basis of how the brain resolves occasionally conflicting cues into social categorical percepts has remained unclear. The present results provide evidence that, in processing the gender or race of a face, opponent social categories coactivate in the rFG, which compete and may resolve into an eventual stable categorization with the assistance of the pre-SMA/dACC. Therefore, the findings provide a neural mechanism through which the brain may translate inherently diverse social cues into coherent categorizations of other people.

References

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Software 67:1–48. CrossRef

Bauer DJ, Preacher KJ, Gil KM (2006) Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. Psychol Methods 11:142–163. CrossRef Medline

Blair IV, Judd CM, Sadler MS, Jenkins C (2002) The role of Afrocentric features in person perception: Judging by features and categories. J Pers Soc Psychol 83:5–25. CrossRef Medline

Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999) Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402:179–181. CrossRef Medline

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624–652. CrossRef Medline

Brouwer GJ, Heeger DJ (2013) Categorical clustering of the neural representation of color. J Neurosci 33:15454–15465. CrossRef Medline

Carlson TA, Ritchie JB, Kriegeskorte N, Durvasula S, Ma J (2014) Reaction time for object categorization is predicted by representational distance. J Cogn Neurosci 26:132–142. CrossRef Medline

Cassidy B, Sprout G, Freeman J, Krendl A (2017) Looking the part (to me): effects of racial prototypicality on race perception vary by prejudice. Soc Cogn Affect Neurosci. 12:685. Medline

Cavanagh JF, Shackman AJ (2015) Frontal midline theta reflects anxiety and cognitive control: meta-analytic evidence. J Physiol Paris 109:3–15. CrossRef Medline

Cisek P, Kalaska JF (2005) Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. Neuron 45:801–814. CrossRef Medline

Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. Annu Rev Neurosci 33:269–298. CrossRef Medline

Cloutier J, Freeman JB, Ambady N (2014) Investigating the early stages of person perception: the asymmetry of social categorization by sex vs age. PLoS One 9:e84677. CrossRef Medline

Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu YC, Abdi H, Haxby JV (2012) The representation of biological classes in the human brain. J Neurosci 32:2608–2618. CrossRef Medline

Contreras JM, Banaji MR, Mitchell JP (2013) Multivoxel patterns in fusiform face area differentiate faces by sex and race. PLoS One 8:e69684. CrossRef Medline

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173. CrossRef Medline

Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114. Medline

Dosenbach NU, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE (2006) A core system for the implementation of task sets. Neuron 50:799–812. CrossRef Medline

Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME. Schlaggar BL, Petersen SE (2007) Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci U S A 104:11073–11078. CrossRef Medline

Faulkenberry TJ (2016) Testing a direct mapping versus competition account of response dynamics in number comparison. Journal of Cognitive Psychology 28:825–842.

Freedman DJ, Assad JA (2011) A proposed common neural mechanism for categorization and perceptual decisions. Nat Neurosci 14:143–146. CrossRef Medline

Freeman JB, Ambady N (2010) MouseTracker: Software for studying real-time mental processing using a computer mouse tracking method. Behav Res Methods 42:226–241. CrossRef Medline

Freeman JB, Ambady N (2011) A dynamic interactive theory of person construal. Psychol Rev 118:247–279. CrossRef Medline

Freeman JB, Dale R (2013) Assessing bimodality to detect the presence of a dual cognitive process. Behav Res Methods 45:83–97. CrossRef Medline

Freeman JB, Johnson KL (2016) More than meets the eye: split-second social perception. Trends Cogn Sci 20:362–374. CrossRef Medline

Freeman JB, Ambady N, Rule NO, Johnson KL (2008) Will a category cue attract you? Motor output reveals dynamic competition across person construal. J Exp Psychol Gen 137:673–690. CrossRef Medline

Freeman JB, Pauker K, Apfelbaum EP, Ambady N (2010) Continuous dynamics in the real-time perception of race. Journal of Experimental Social Psychology 46:179–185. CrossRef

Freeman JB, Dale R, Farmer TA (2011a) Hand in motion reveals mind in motion. Front Psychol 2:59. CrossRef Medline

Freeman JB, Ambady N, Midgley KJ, Holcomb PJ (2011b) The real-time link between person perception and action: Brain potential evidence for dynamic continuity. Social Neuroscience 6:139–155. CrossRef Medline

Galinsky AD, Hall EV, Cuddy AJ (2013) Gendered races: implications for interracial marriage, leadership selection, and athletic participation. Psychol Sci 24:498–506. CrossRef Medline

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20–25. CrossRef Medline

Grinband J, Hirsch J, Ferrera VP (2006) A neural representation of categorization uncertainty in the human brain. Neuron 49:757–763. CrossRef Medline

Ha OR, Bruce AS, Pruitt SW, Cherry JB, Smith TR, Burkart D, Bruce JM, Lim SL (2016) Healthy eating decisions require efficient dietary self-control in children: a mouse tracking food decision study. Appetite 105:575–581. CrossRef Medline

Hanke M, Halchenko YO, Sederberg PB, Olivetti E, Fründ I, Rieger JW, Herrmann CS, Haxby JV, Hanson SJ, Pollmann S (2009) PyMVPA: a unifying approach to the analysis of neuroscientific data. Front Neuroinform 3:3. Medline

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

Hehman E, Ingbretsen ZA, Freeman JB (2014) The neural basis of stereotypic impact on multiple social categorization. Neuroimage 101:704–711. CrossRef Medline

Hehman E, Stolier RM, Freeman JB (2015) Advanced mouse tracking analytic techniques for enhancing psychological science. Group Processes and Intergroup Relations 18:384–401. CrossRef

Iordan MC, Greene MR, Beck DM, Fei-Fei L (2015) Basic level category structure emerges gradually across human ventral visual cortex. J Cogn Neurosci 27:1427–1446. CrossRef Medline

Iordan MC, Greene MR, Beck DM, Fei-Fei L (2016) Typicality sharpens category representations in object-selective cortex. Neuroimage 134:170–179. CrossRef Medline

Jahn A, Nee DE, Alexander WH, Brown JW (2016) Distinct regions within medial prefrontal cortex process pain and cognition. J Neurosci 36:12385–12392. CrossRef Medline

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) Fsl. Neuroimage 62:782–790. CrossRef Medline

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311. Medline

Koop GJ (2013) An assessment of the temporal dynamics of moral decisions. Judgment and Decision Making 8:527.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868. CrossRef Medline

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2:4. CrossRef Medline

Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442:572–575. CrossRef Medline

Lepora NF, Pezzulo G (2015) Embodied choice: how action influences perceptual decision making. PLoS Comput Biol 11:e1004110. CrossRef Medline

Locke V, Macrae CN, Eaton JL (2005) Is person categorization modulated by exemplar typicality? Social Cognition 23:417–428. CrossRef

Macrae CN, Bodenhausen GV (2000) Social cognition: Thinking categorically about others. Annu Rev Psychol 51:93–120. CrossRef Medline

Mattek AM, Whalen PJ, Berkowitz JL, Freeman JB (2016) Differential effects of cognitive load on subjective versus motor responses to ambiguously valenced facial expressions. Emotion 16:929–936. CrossRef Medline

Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage 53:103–118. CrossRef Medline

Nachev P, Rees G, Parton A, Kennard C, Husain M (2005) Volition and conflict in human medial frontal cortex. Curr Biol 15:122–128. CrossRef Medline

Nachev P, Kennard C, Husain M (2008) Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci 9:856–869. CrossRef Medline

Narayanan NS, Cavanagh JF, Frank MJ, Laubach M (2013) Common medial frontal mechanisms of adaptive control in humans and rodents. Nat Neurosci 16:1888–1895. CrossRef Medline

Neta M, Schlaggar BL, Petersen SE (2014) Separable responses to error, ambiguity, and reaction time in cingulo-opercular task control regions. Neuroimage 99:59–68. CrossRef Medline

Oehrn CR, Hanslmayr S, Fell J, Deuker L, Kremers NA, Do Lam AT, Elger CE, Axmacher N (2014) Neural communication patterns underlying conflict detection, resolution, and adaptation. J Neurosci 34:10438–10452. CrossRef Medline

O'Reilly RC, Wyatte D, Herd S, Mingus B, Jilk DJ (2013) Recurrent processing during object recognition. Front Psychol 4:124. CrossRef Medline

Papesh MH, Goldinger SD (2012) Memory in motion: Movement dynamics reveal memory strength. Psychonomic Bulletin and Review 19:906–913. CrossRef Medline

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10:59–63. CrossRef Medline

Ratner KG, Kaul C, Van Bavel JJ (2013) Is race erased? Decoding race from patterns of neural activity when skin color is not diagnostic of group boundaries. Soc Cogn Affect Neurosci 8:750–755. CrossRef Medline

Ritchie JB, Tovar DA, Carlson TA (2015) Emerging object representations in the visual system predict reaction times for categorization. PLoS Comput Biol 11:e1004316. CrossRef Medline

Rosch E (1978) Principles of categorization. In: Cognition and categorization (Rosch E, Lloyd BB, eds), pp 27–48. Hillsdale, NJ: Erlbaum.

Rolls ET (2000) Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron 27:205–218. CrossRef Medline

Schneider IK, van Harreveld F, Rotteveel M, Topolinski S, van der Pligt J, Schwarz N, Koole SL (2015) The path of ambivalence: tracing the pull of opposing evaluations using mouse trajectories. Front Psychol 6.

Sha L, Haxby JV, Abdi H, Guntupalli JS, Oosterhof NN, Halchenko YO, Connolly AC (2015) The animacy continuum in the human ventral vision pathway. J Cogn Neurosci 27:665–678. CrossRef Medline

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240. CrossRef Medline

Sheth SA, Mian MK, Patel SR, Asaad WF, Williams ZM, Dougherty DD, Bush G, Eskandar EN (2012) Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature 488:218–221. CrossRef Medline

Spivey MJ, Dale R (2006) Continuous dynamics in real-time cognition. Curr Dir Psychol Sci 15:207–211. CrossRef

Spivey MJ, Grosjean M, Knoblich G (2005) Continuous attraction toward phonological competitors. Proc Natl Acad Sci U S A 102:10393–10398. CrossRef Medline

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

Stolier RM, Freeman JB (2016) Neural pattern similarity reveals the inherent intersection of social categories. Nat Neurosci 19:795–797. CrossRef Medline

Tang H, Yu H-Y, Chou C-C, Crone NE, Madsen JR, Anderson WS, Kreiman G (2016) Cascade of neural processing orchestrates cognitive control in human frontal cortex. eLife 5: pii: e12352. CrossRef Medline

Ullsperger M, Danielmeier C, Jocham G (2014) Neurophysiology of performance monitoring and adaptive behavior. Physiol Rev 94:35–79. CrossRef Medline

Usher M, McClelland JL (2001) The time course of perceptual choice: The leaky, competing accumulator model. Psychol Rev 108:550–592. CrossRef Medline

van der Wel RP, Eder JR, Mitchel AD, Walsh MM, Rosenbaum DA (2009) Trajectories emerging from discrete versus continuous processing models in phonological competitor tasks: A commentary on Spivey, Grosjean, and Knoblich (2005). Journal of Experimental Psychology: Human Perception and Performance 35:588–594. CrossRef Medline

Vul E, Lashkari D, Hsieh PJ, Golland P, Kanwisher N (2012) Data-driven functional clustering reveals dominance of face, place, and body selectivity in the ventral visual pathway. J Neurophysiol 108:2306–2322. CrossRef Medline

Watson R, Latinus M, Noguchi T, Garrod O, Crabbe F, Belin P (2014) Crossmodal adaptation in right posterior superior temporal sulcus during face–voice emotional integration. J Neurosci 34:6813–6821. CrossRef Medline

Wojnowicz MT, Ferguson MJ, Dale R, Spivey MJ (2009) The self-organization of explicit attitudes. Psychol Sci 20:1428–1435. CrossRef Medline

Wyatte D, Herd S, Mingus B, O'Reilly R (2012) The role of competitive inhibition and top-down feedback in binding during object recognition. Front Psychol 3:182. CrossRef Medline