Categorical perception of race is mediated by distributed patterns of activity in the brain

Lily Tsoi[1], Yune S. Lee[2], Kevin Jiang[3], Liane Young[3]

[1] Department of Psychology, Princeton University

[2] Department of Speech and Hearing Science, The Ohio State University

[3] Department of Psychology and Neuroscience, Boston College

Contact Information:

Lily Tsoi
Peretsman Scully Hall
Princeton University
Princeton, NJ 08544
Email: ltsoi@princeton.edu

Abstract

Race is an abstract social category that is often associated with perceptual cues. We used functional magnetic resonance imaging (fMRI) to investigate the cognitive and neural processes that support categorical perception of race: the warping of perceptual facial information that vary along a continuum into discrete racial groups. Participants viewed morphed faces along continua from White to Black in the scanner and had their categorical boundaries measured outside the scanner. The warping of face stimuli along a continuum into discrete race categories appears to be mediated by distributed patterns of activity in regions associated with the attention network. Additionally, ROI (region-of-interest)-based multivariate pattern analyses revealed contributions of face-processing regions (i.e., right FFA) and social cognition (i.e., left temporoparietal junction) in categorical race perception. We also reveal a different role for regions implicated in visual processing and top-down processing of visual stimuli; activity in these regions tracks with changes in facial features associated with racial ambiguity. Together, these findings point to different types of information afforded by pattern-based and activation-based analyses regarding processes involved in race perception.

Keywords: MVPA, race perception, social categorization

1. Introduction

When meeting someone in person for the first time, one might observe numerous visual cues, such as face shape, skin color, pupil color, wrinkles around the eye, and arrive at quick conclusions about the person's membership in specific social categories, for instance, as middle-aged or Black. The current study investigates the processes by which people take in visual cues that vary along a continuum (e.g., skin color, which varies from light to dark) and by which people perceive categorical features (e.g., as a White person). This phenomenon, known as categorical perception, has been identified for socially complex features, including race (Levin & Angelone, 2002). The current study investigates how categorical perception of race is represented in the human brain, using the contrast between Black and White as a case study. By understanding the processes that people use to categorize others, we can gain greater insight into why people categorize others the way they do, especially with regard to socially complex and often constructed categories.

One hypothesis is that categorical perception of a complex, socially-constructed feature like race relies on the same mechanisms as categorical perception of less complex non-social features. Research on categorical perception has primarily focused on the auditory domain, revealing neural substrates that support categorical perception of speech (Bidelman et al., 2013; Chang et al., 2010; Desai et al., 2008; Lee et al., 2012; Liberman et al., 1957; Myers & Swan, 2012; Prather et al., 2009; Raizada & Poldrack, 2007) and music (Klein & Zatorre, 2015; Klein & Zatorre, 2011; Lee et al., 2011). In the visual domain, researchers have gained a greater understanding of categorical perception of color (Franklin et al., 2008; Özgen & Davies, 2002; Roberson et al., 2008), shapes/patterns (Goldstone, 1994; Newell & Bülthoff, 2002; Notman et al., 2005), and face identity and facial expression (Beale & Keil, 1995; Etcoff & Magee, 1992;

Rotshtein et al., 2004). Categorical perception of a feature like race may rely on similar mechanisms as the aforementioned features. Alternatively, because of the complexity and abstract notion of the social category, categorical perception of race may be supported instead by different mechanisms and/or by different brain networks.

Research closest to our investigation has shown that categorical perception of complex visual features such as face identity and facial expression is supported by brain regions further along the visual processing stream and higher-level cognitive regions. For example, the right fusiform gyrus and fusiform face area (FFA) appear sensitive to perceived changes in face identity (e.g., from Marilyn Monroe to Margaret Thatcher) (Rotshtein et al., 2004). Similarly, a study using an adaptation paradigm shows sensitivity of the FFA and posterior superior temporal sulcus (pSTS) to perceived changes in identity (i.e., different individuals) as well as expression (i.e., angry, afraid, disgusted, happy) (Fox et al., 2009). Regions outside of this core face network shows further specificity: the middle STS shows release from adaptation when participants perceive a change in expression but not identity, whereas the precuneus shows release from adaptation for the opposite change. Relatedly, categorical perception of gender from faces appears to be represented by the orbitofrontal cortex (Freeman et al., 2010). Overall, this body of work suggests that regions further along the visual processing stream and/or higher-level brain regions outside of the core face network support categorical perception of facial features like face identity and facial expression. These same regions may also support categorical perception of other complex features such as race. If, instead, other regions support categorical perception of race, this finding would suggest that either different or additional mechanisms are involved.

To identify regions that support categorical race perception, we used multivariate pattern analysis (MVPA), an approach that has been successfully used for discriminating neural patterns that map continuously morphed stimuli into discrete categories (Lee et al., 2012). This pattern-based approach, unlike an activation-based approach, allows us to identify subtle differences across conditions that share many physical features (Kriegeskorte et al., 2006). Along with examining categorical race perception at the whole-brain level, we used functional ROI-based MVPA to examine whether categorical perception of race is supported by patterns of local activity in specific regions. One candidate region is the FFA, an important region for categorical perception of facial features discussed above. Another set of candidate regions include regions involved in theory of mind (ToM), such as bilateral temporoparietal junction (TPJ), precuneus, and medial prefrontal cortex (mPFC) (Fletcher et al., 1995; Gallagher et al., 2000; Gobbini et al., 2007; Saxe & Kanwisher, 2003). People infer socially meaningful information such as people's internal states (e.g., traits, mental states) rapidly and reflexively from perceptual information, including facial appearance (for reviews, see Olivola et al., 2014; Todorov et al., 2015). Some evidence points to the recruitment of regions implicated in ToM even when people are merely viewing natural scenes containing people (Wagner et al., 2011). We investigate whether categorization of people into social categories such as race relies, to some extent, on regions implicated in social processing. In short, the present study investigates the cognitive and neural processes supporting categorical race perception, examining possible contributions of late visual processing (e.g., faces) and higher-level cognitive processing (e.g., social cognition).

It is important to note how our research question departs from other related questions already addressed in the social neuroscience literature. For instance, earlier work has examined whether face-processing regions such as the fusiform gyrus, and more specifically the FFA, respond differently to Black vs. White faces (Van Bavel et al., 2008; 2011). While these studies reveal that, overall, activity in these regions do not differ for Black and White faces, other work has revealed that the fusiform gyrus does nevertheless encode information about these two racial categories when examining neural patterns at the multivariate level (Contreras et al., 2013), even when racial category information is irrelevant to the task (Ratner et al., 2012). Other relevant work reveals greater fusiform responses for faces corresponding to people's own race versus other races (Golby et al., 2001; Lieberman et al., 2005). Theoretical and empirical work also provides support for the notion that facial features, when viewed, activate multiple social categories, emotions, and traits that then activate semantic associations (stereotypes), which, in turn, influence subsequent processing, behavior, and perception (Freeman et al., 2018, 2020). By contrast, our work focuses on understanding the singular process by which people take in facial features relevant to race along a continuum as inputs and outputs a racial category label.

2. Methods

2.1. Participants

Twenty-nine right-handed participants between the ages of 18 and 40 (mean ± standard deviation = 26.5 ± 6.46; 13 females) were recruited from the Boston community. All participants were native English speakers, had normal or corrected-to-normal vision, and reported having no history of psychiatric or neurological disorders. One participant fell asleep early in the scan session and was removed from the study. The remaining 28 participants consisted of 14 Caucasians (7 females), 12 African-Americans (4 female), 1 Asian-American (female), and 1 male of unspecified race. Participants gave written informed consent and were paid $25/hour for their participation. The study was approved by the Boston College Institutional Review Board.

2.2. Stimuli and Procedures

FaceGen Modeller 3.5 (Singular Inversions) was used to generate 3D faces that were morphed by race (from White to Black). The European to African morphing slider was used, which contained 41 steps: faces every four steps along the slider were used as stimuli, starting with the third step and ending with the third-to-last step. The race morphing control combined 95 different features, such as skin shade (dark/light), brow ridge (e.g., high/low), cheeks (e.g., round/gaunt), and eye sockets (dark/light); because the various feature controls were correlated, adjusting the race morphing slider affected all of these features.

Ten face morphs varying along a continuum from White to Black were produced for each gender (Figure 1a). Each face was presented at a 20.05° yaw angle. An elderly version of each face was created as an oddball stimulus; there were 20 oddball stimuli in total. Face images (faces on a black background) and fixation crosses were centered on a gray background (Figure 1b). During fMRI scanning, these images were projected onto a screen (1024 x 768 pixel resolution) at the end of the magnet bore with an InFocus IN5542 projector. Each face image horizontally subtended a visual angle of approximately 11°.

Participants were instructed to indicate the appearance of any oddball stimuli (i.e., elderly-looking faces) via button box press with their right hand. There were 10 runs; each run lasted 4.2 minutes and consisted of 40 trials: 38 typical trials and 2 oddball trials. In a typical trial, the same face morph was presented three times (each for 0.5 s) with a blank screen (0.1 s) between face morphs. In an oddball trial, the same face morph was presented two times, and an elderly-looking version of the face morph was presented once. The temporal position of the oddball was random (i.e., first, second, or third face in the trial) but roughly matched: across 20 oddball trials, the average participant was presented with an oddball 6.83 times on the first position, 7.08 times on the second, and 5.92 times on the third. Due to technical issues, we were able to collect only 22 out of 28 participants' oddball responses in the scanner. A fixation cross (4.5 s) was presented between all trials. Two sets of randomized stimulus presentation order sequences were created using the sequence optimizer optseq2 (Dale, 1999), and the order sequence was counterbalanced across participants. The experiment was controlled by MATLAB 2008b and the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) on a MacBook Pro.

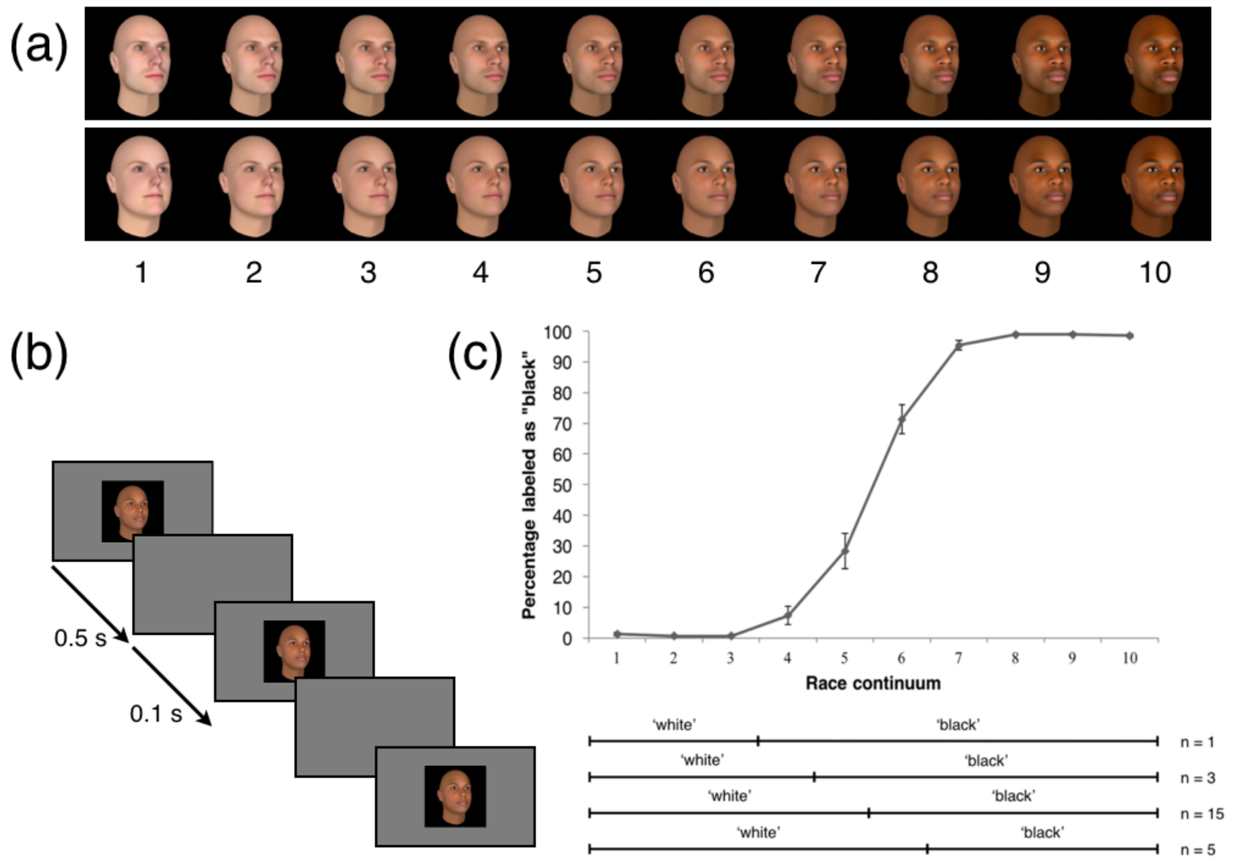

Figure 1. Stimuli for the fMRI experiment and behavioral results. (a) Two sets of 10 faces morphed along a White-Black continuum: one male (top), one female (bottom). (b) Typical trial: the same face is presented three times (depicted in figure). Oddball trial: the same face is presented two times and the oddball (an elderly-looking version of the face) randomly appears in one of the three positions (not depicted). Participants are instructed to make a button-box response upon seeing an oddball. (c) Perceptions of faces along White-Black continua from 24 participants. *Top*: the average psychometric curve. The y-axis shows the percentage of responses labeled as "Black" for each face token (i.e. morph-level) in the continuum (1=White; 10=Black). Error bars indicate SEM. *Bottom*: Participant count by categorical boundary.

2.3. Functional localizer tasks

Participants completed two functional localizer tasks. One functional localizer task (Epstein & Kanwisher, 1998) was used to define bilateral fusiform face area. A second localizer task (Dodell-Feder et al., 2011) was used to functionally define regions involved in ToM or social cognition more generally. Information about the two tasks can be found in Supplementary Materials.

2.4. Post-scan behavioral tasks

After the fMRI experiment, participants completed a behavioral task outside the scanner that measured their subjective categorical boundary for the face morphs during the experimental task in the scanner. Each face was presented ten times, and participants were instructed to indicate whether they viewed the face as Black or White by pressing one of two buttons. The boundary was defined as the 50% crossover point between viewing the face as white or black on each participant's psychometric curve. This boundary was used to define binary classes (Black vs. White) for subsequent labeling of each participant's neural data. Stimuli were presented using PsychoPy (Peirce, 2007) on a MacBook Pro. Finally, for exploratory purposes, a subset of participants ($N = 14$) completed a race Implicit Association Test (IAT), a measure of implicit racial bias.

2.5. fMRI data acquisition

Twenty of the participants were scanned in a Siemens 3T Tim Trio MRI scanner at the Center for Brain Science (Harvard University). Because the Center for Brain Science switched to a 3T Siemens Prisma scanner and because we wanted to continue scanning with the same protocols and the same type of scanner, we scanned the final nine participants in a Siemens 3T

Tim Trio MRI scanner at the Athinoula A. Martinos Imaging Center (Massachusetts Institute of Technology).

Participants were scanned using a 12-channel head coil. Thirty-six axial slices (3-mm isotropic voxels, 0.54-mm gap) were acquired using the following gradient-echo planar imaging (EPI) sequence parameters: repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle (FA) = 90°; field of view (FOV): 216 x 216; interleaved acquisition. Anatomical data were collected with T1-weighted multi-echo magnetization prepared rapid acquisition gradient echo image sequences using the following parameters: TR = 2530 ms; TE = 1.64 ms; FA = 7°; 1-mm isotropic voxels; 0.5 mm gap between slices; FOV = 256 x 256.

2.6. fMRI data analyses

Data preprocessing and analyses were performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm) and custom software with MATLAB 2014a. All images were slice-time corrected, realigned to the first EPI of the first run, and spatially normalized into Montreal Neurological Institute (MNI) standard stereotactic space (EPI template) with preserved original voxel size.

*2.6.1 Searchlight analysis***.** For the searchlight analysis, fMRI time courses for all voxels were extracted from unsmoothed images and high-pass filtered with a 128 s cutoff. The signals were also mean-centered to normalize intensity differences among runs. Trials during which an oddball appeared were removed from all analyses. Time points were labeled as belonging to “Black” or “White” categories based on participants’ subjective categorical boundaries. The preprocessed time courses were mapped to each race category based on a GLM framework: first, a regressor for each race category per run was constructed by convolving the onset of each trial associated with the race category with the canonical HRF (hemodynamic response function).

Next, the mean height of the regressor was calculated. Lastly, if the height of the regressor at each time point was greater than the mean height, this time point was assigned to a particular race category.

We moved a 3-voxel radius sphere throughout the brain, centering on each voxel. In each searchlight sphere, a binary classification ("Black" or "White") was performed using a Gaussian Naïve Bayes classifier (Pereira et al., 2009; Raizada & Lee, 2013). For validating the classification, we used leave-one-run-out procedure, wherein data from one of the runs were reserved for testing, and remaining data were used for training (a total of 10 cross-validations). The output searchlight images for all participants were used to perform a group-level analysis. The group maps were thresholded using a voxel-wise threshold of $p < .001$ (uncorrected), after which we corrected for multiple comparisons ($p < .05$) across the whole brain based on cluster extent and Gaussian random field theory (Friston et al., 1994; Worsley et al., 1992). Anatomical labels were retrieved using a combination of SPM Anatomy toolbox (v 2.2) (Eickhoff et al., 2005), xjView (https://www.alivelearn.net/xjview), and SPM Neuromorphometrics.

2.6.2. ROI-based MVPA. We examined the spatial pattern of neural activity for Black and White faces within regions involved in face and race processing (i.e., bilateral FFA) and regions involved in social cognition (i.e., right and left TPJ, precuneus, and dmPFC). Analyses were performed in MNI space. To define the FFA, beta values were estimated in each voxel for blocks corresponding to each of four types of stimuli (scenes, faces, objects, and scrambled objects) using a boxcar function. A contrast map was obtained for [faces > objects] in every participant. The FFA was defined as all voxels in a 9-mm radius of the peak voxel that passed threshold in the contrast image ($p < 0.001$, uncorrected, extent cluster size > 10). To define regions involved in social cognition, beta values were estimated in each voxel for stories describing mental states (e.g., belief) or physical representations (e.g., photo). A contrast map was obtained for [belief > photo] in every participant. ROIs were defined as all voxels in a 9-mm radius of the peak voxel that passed threshold in the contrast image ($p < 0.001$, uncorrected, extent cluster size > 10). Information about the number of participants whose ROIs we were able to define and the MNI coordinates for those ROIs can be found in Table S1 (Supplementary Material).

The classification procedure of ROI-based MVPA was identical to that used for searchlight. Leave-one-run-out cross-validation was used for all analyses. An accuracy score averaged across training/testing set combinations was computed for each ROI and every individual.

2.6.3. Representational similarity analysis (RSA). For each RSA, a linear model was constructed with matrix regressors in the lower triangle of 10 x 10 token space, in which the (i, j) entry represents the cross-correlation of voxel vectors for the ith and jth face token. Regressor entries indicate the hypothesized similarity between each pair of face tokens; e.g., entry (2,6) corresponds to the similarity between face token 2 and 6. Similarity scores ranged from 0 to 1, such that 0 represents maximal racial dissimilarity and 1 represents maximal racial similarity. Two regressors, continuous and categorical, were generated based on competing hypotheses for race perception. For the continuous regressor, face tokens were mapped at equidistant points along a White-Black continuum (1=White; 10=Black), and entries of the regressor were calculated based on the normalized Euclidean distance between token pairs along this continuum. For the categorical regressor, face tokens were labeled as either “White” or “Black” based on post-scan measurements of each participant’s subjective categorical boundary, and entries of the regressor were either coded as 0 (maximal racial dissimilarity) for Black-White pairs, or 1 (maximal racial similarity) for Black-Black and White-White pairs.

Within each ROI, voxel patterns for individual face tokens were correlated to produce a 10 x 10 empirical similarity matrix of pairwise neural similarities, which was then modeled with our RSA matrix regressors using CoSMoMVPA (OosterHof et al., 2016). Individual voxel vector entries for each face token consisted of the average of all time course data mapped to that face token based on a GLM framework (data mapping as described in section 2.6.1). Voxel vectors were demeaned before correlation (Diedrichsen & Kriegeskorte, 2017), and both the empirical similarity matrix and hypothesized regressors were Z-scored before estimating regression coefficients.

3. Results

3.1.1 Behavioral results

Participants' responses in the post-scan behavioral task confirmed that participants overall perceived a sharp categorical shift for faces near the middle of the continuum (Figure 1c). Four participants perceived the continua in a non-categorical manner; because we were interested in categorical race perception, these participants' data were removed from all further analyses (though we provide a brief mention of these participants in the Discussion). The analyses below therefore reflect the data of the remaining 24 participants. Analyses examining the relationship between people's responses in the post-scan task and implicit racial bias as assessed by a race Implicit Association Test (IAT) revealed no meaningful relationship (see Supplementary Material).

Additionally, we assessed participants' performance on the oddball-detection task as a measure of participants' attentiveness. Due to a coding error, we were able to consistently collect responses to oddballs on only the first and second positions within the trial. On average, participants responded to 5.44 (*SD* = 2.64) out of 6.83 oddballs and 5.50 (*SD* = 2.64) out of 7.08

oddballs in the first position and second position, respectively. Overall, these results suggest that participants were attending to the task.

3.1.2. fMRI results

Searchlight analyses revealed several regions whose patterns of activity distinguished between participants' subjective boundary of who they perceive as Black or White. These regions include the right superior parietal lobule, right middle frontal gyrus, left insula, right superior frontal gyrus, right fusiform gyrus, left middle temporal gyrus, and left postcentral gyrus (Figure 2; Table 2). Due to concerns that White and Black participants would show different patterns of results, we compared the searchlight results for these two groups of participants; no significant differences were found between the two groups.

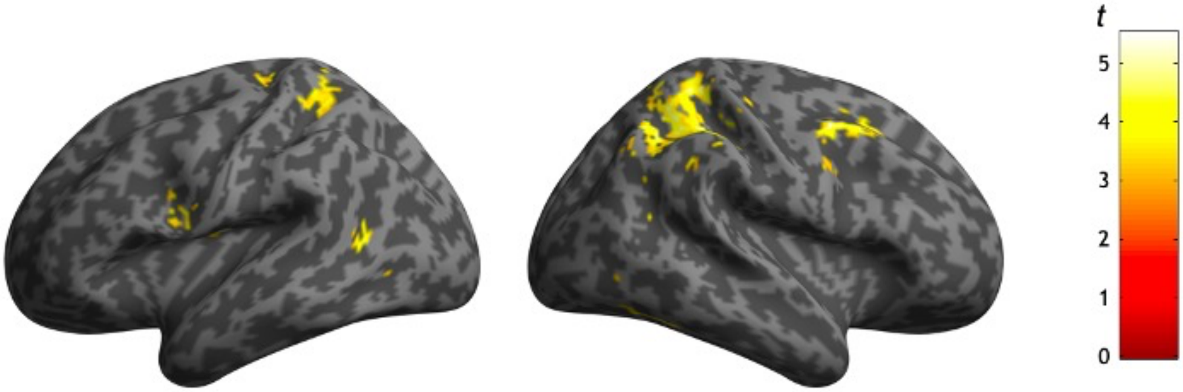

Figure 2. Group-level searchlight analysis based on participants' subjective categorical boundaries reveal clusters with peaks in the right superior parietal lobule, right middle frontal gyrus, left insula, right superior frontal gyrus, right fusiform gyrus, left middle temporal gyrus, and left postcentral gyrus (cluster-level corrected at $p < 0.05$, FWE).

Table 1. Searchlight results. Regions mediating categorical race perception based on participants' subjective categorical boundaries (cluster-level corrected at $p < 0.05$, FWE).

	MNI coordinates				
Region name	*x*	*y*	*z*	*t value*	# of voxels
R superior parietal lobule	**42**	**-34**	**37**	**5.53**	**554**
R inferior parietal lobule	39	-40	46	5.28	
R inferior parietal lobule	39	-49	49	5.24	
R middle frontal gyrus	**45**	**2**	**55**	**5.36**	**175**
R middle frontal gyrus	39	-4	64	4.19	
R precentral gyrus	51	5	43	4.06	
L insula	**-39**	**-13**	**22**	**5.36**	**93**
White matter	-30	-7	25	4.46	
L rolandic operculum	-45	-4	19	4.24	
R superior frontal gyrus	**21**	**8**	**40**	**5.17**	**78**
R cingulate gyrus	18	-4	40	4.86	
R superior frontal gyrus	21	2	49	4.05	
R fusiform gyrus	**36**	**-58**	**-5**	**5.14**	**88**
R inferior temporal gyrus	48	-52	-17	4.85	
R fusiform gyrus	42	-43	-8	4.79	
L middle temporal gyrus	**-33**	**-49**	**16**	**5.1**	**83**
L supramarginal gyrus	-33	-49	25	4.45	
L middle temporal gyrus	-42	-58	7	4.41	
L postcentral gyrus	**-18**	**-40**	**55**	**4.38**	**114**
L postcentral gyrus	-24	-46	55	4.37	
L precentral gyrus	-18	-28	49	4.3	

Note: Regions in bold are peak voxels; indented regions indicate sub-peak voxels.

We also performed a searchlight analysis based on an 'objective' boundary (i.e., half of the tokens were labeled as "White", whereas the other half were labeled as "Black") to address an alternative possibility in which racial features could be categorized. This analysis yielded no regions after cluster-wise correction ($p < 0.05$). This finding, coupled with the one above using subjective boundaries, suggests that spatial patterns of activity in the brain can be discriminated when using participants' subjective but not objective categorizations of faces by race.

We also tested whether the spatial patterns of neural activity for Black and White faces as defined subjectively are distinct from each other within (1) regions involved in face processing, i.e., bilateral fusiform face area (rFFA and lFFA), and (2) regions involved in social cognition, i.e., right and left temporoparietal junction (TPJ), precuneus, and dmPFC. We were able to define the left and right fusiform area, right temporoparietal junction, and precuneus in 21 out of 24 participants, left temporoparietal junction in 20 out of 24 participants, and dorsomedial prefrontal cortex in 15 out of 24 participants (Table S1). ROI-based MVPA showed that the spatial patterns of neural activity for Black and White faces were distinguishable from each other in the left FFA and right FFA; classification accuracy was above chance (0.5) for both the left FFA (*mean accuracy* = 0.53, *SD of accuracy* = 0.04, $t(20) = 4.40$, $p < 0.001$, one-tailed) and right FFA (*mean accuracy* = 0.52, *SD of accuracy* = 0.04, $t(20) = 3.274$, $p = 0.002$, one-tailed). Classification accuracy was at chance level for the rTPJ, precuneus, and dmPFC (*p*s > 0.05) but above chance for lTPJ ($M = 0.51$, $SD = 0.03$, $t(19) = 2.311$, $p = 0.017$).

As a further test of the categorical nature of race perception, we performed representational similarity analyses (RSA). With this analysis, we can test different hypotheses about the nature of the representation. Our primary hypothesis is that spatial patterns for face tokens along the continuum would be more similar within the same category than between categories, based on participants' subjective categorical boundary. An alternative hypothesis is that spatial patterns for face tokens along the continuum would be presented in a continuous manner; that is, spatial patterns for face token 1 would be more similar to face token 2 than face token 3, along the continuum of face tokens ranging from 1 to 10. We found evidence supporting our primary hypothesis in the right FFA: within right FFA, beta weights for the *categorical* regressor (similarity regressor corresponding to our primary hypothesis) was significantly above zero across participants (*M* = 0.014, *t*(20) = 2.323, *p* = .031), while beta weights for the *continuous* regressor (similarity regressor corresponding to our alternative hypothesis) did not significantly differ from zero (*M* = -0.004, *t*(20) = -1.180, *p* = .252). In all other ROIs, beta weights did not significantly differ from zero for both *categorical* regressors (lFFA: *M* = 0.001, *t*(20) = 0.228, *p* = 0.822; lTPJ: *M* = 0.009, *t*(19) = 1.642; *p* = 0.117; rTPJ: *M* = 0.009, *t*(20) = 1.485, *p* = 0.153; precuneus: *M* = -0.006, *t*(20) = -1.310; *p* = 0.205; dmPFC: *M* = -0.004, *t*(14) = -0.680, *p* = 0.508) and *continuous* regressors (lFFA: *M* = 0.007, *t*(20) = 1.224, *p* = 0.235; lTPJ: *M* = 0.007, *t*(19) = 1.424; *p* = 0.171; rTPJ: *M* = 0.000, *t*(20) = -0.027, *p* = 0.979; precuneus: *M* = 0.012, t(20) = 2.083; *p* = 0.0503; dmPFC: *M* = -0.003, *t*(14) = -0.661, *p* = 0.520).

We note that when we examined overall levels of neural responses across Black vs. White faces based on participants' subjective and categorical boundaries to search for regions that responded more to one category over the other, no significant clusters emerged. However, parametric activation-based analyses (Table 2) revealed that activity in the following regions increased with decreases in racial ambiguity (or increases in racial unambiguity) as defined experimentally: the left lingual gyrus, left fusiform gyrus, left inferior frontal gyrus, and right posterior medial frontal cortex. In contrast, activity in bilateral mid orbital gyrus increased with increases in racial ambiguity.

These findings together corroborate the idea that people's subjective perception of faces into racial categories cannot be captured using conventional activation-based analyses but can be well-captured using pattern-based analyses. However, activation-based analyses reveal regions that track changes in facial features that lead to objective racial unambiguity.

Table 2. Regions involved in the parametric effect of racial ambiguity

	MNI coordinates				
Region name	*x*	*y*	*z*	*t* value	# of voxels
Regions whose activity increased as faces became more racially defined					
L lingual gyrus	**-27**	**-94**	**-14**	**5.31**	**51**
L lingual gyrus	-33	-88	-14	4.70	
L fusiform gyrus	**-33**	**-67**	**-14**	**4.84**	**55**
L fusiform gyrus	-33	-55	-20	4.82	
L cerebellum	-33	-73	-20	4.38	
L inferior frontal gyrus	**-48**	**17**	**22**	**4.73**	**46**
L inferior frontal gyrus	-42	11	28	4.66	
L inferior frontal gyrus	-45	29	19	3.82	
R posterior medial frontal cortex	**3**	**23**	**49**	**4.33**	**44**
Regions whose activity increased as faces became more racially ambiguous					
L mid orbital gyrus	**-6**	**38**	**-11**	**5.23**	**67**
R mid orbital gyrus	3	44	-8	5.03	
Anterior cingulate cortex	0	50	-2	4.10	

Note: Regions in bold are peak voxels; indented regions indicate sub-peak voxels. Regions listed here were cluster-level corrected at $p < 0.05$ (FWE).

4. Discussion

The current study investigated neural mechanisms that support categorical perception of race. Different analyses reveal different aspects of race perception: pattern-based analyses reveal that the warping of face stimuli along a continuum into discrete race categories appears to be mediated by distributed patterns of activity across the brain, including the following regions: right superior parietal lobule, right middle frontal gyrus, left insula, right superior frontal gyrus, right fusiform gyrus, left middle temporal gyrus, and left postcentral gyrus. Meanwhile, activation-based analyses show that mean levels of activity within regions for visual processing and top-down processing of visual stimuli linearly tracked with racial ambiguity. Together, these findings point to different types of information afforded by pattern-based and activation-based analyses regarding processes involved in race perception.

Categorical race perception is mediated by distributed patterns of activity in regions typically involved in top-down control of visual attention (e.g., superior parietal lobule) (Corbetta et al., 2008; Parlatini et al., 2017) and gatekeeping between top-down cognitive control and bottom-up sensory-driven attention (e.g., middle frontal gyrus) (Japee et al., 2015). Some researchers have proposed that the frontoparietal network provides a 'priority map', integrating factors related to bottom-up and top-down attention (Bisley & Goldberg, 2010; Katsuki & Constantinidis, 2014). Indeed, one study has revealed frontoparietal contributions to working memory by representing feature-specific information about relevant stimuli *and* by mediating top-down cognitive control (Ester et al., 2015). This frontoparietal network may be encoding stimulus features related to race and/or encoding race information based on top-down factors like prior knowledge (e.g., past experiences with people of different races).

We also focused on specific regions of interest based on our *a priori* hypotheses regarding face-processing regions (i.e., FFA) and regions involved in social cognition (i.e., bilateral TPJ, precuneus). Research in social neuroscience has directly investigated how the FFA responds to faces that differ by race. Studies using conventional activation-based analyses have found preferential activation of the fusiform gyrus to own-race versus other-race faces (Feng et al., 2011; Golby et al., 2001; Lieberman et al., 2005; Natu et al., 2011); however, evidence points to the idea that the fusiform gyrus may not be responding to race specifically, but to motivationally relevant social categories (Van Bavel et al., 2008, 2011). Indeed, when race is made orthogonal to group membership, the fusiform gyrus responds similarly to White and Black faces but differentially to own-group versus other-group faces regardless of race (Van

Bavel et al., 2008). Although overall activity in the fusiform gyrus may be driven by motivationally relevant social categories, recent work using MVPA reveals that the spatial patterns of activity for Black and White faces within the fusiform gyrus and specifically the FFA are distinguishable from each other above chance level (Contreras et al., 2013; Ratner et al., 2013). This work suggests that race is in fact encoded in the fusiform gyrus. Our study provides evidence that the right FFA also supports *categorical perception* of visual features related to race. Indeed, prior work on the neural basis of categorical perception of visual features such as facial identity or expression showed that the right FFA is sensitive to perceived (categorical) changes but not physical changes in facial identity and expression (Fox et al., 2009; Rotshtein et al., 2004). Our results show that the FFA mediates categorical perception of race. This finding provides converging evidence for the role of the FFA in categorical perception of visually accessible facial features, including race. Separately, regarding the ToM network, our findings reveal a specific role for the left TPJ but no other regions in the network. Prior work contrasting the left TPJ with the right TPJ suggests an involvement of the left TPJ in computing differences in perspectives (Aichhorn et al., 2009; Aichhorn et al., 2006; Perner et al., 2006). Perhaps people, upon viewing faces, generate possible perspectives behind those faces as they categorize them; while the current study cannot provide evidence to support this, future work could assess the scope of perspectives that people generate when viewing faces.

Activation-based analyses suggest that increased activity in the left lingual gyrus, left fusiform gyrus, left IFG, and right posterior medial frontal cortex is linked to increases in racial unambiguity. Prior work has linked the lingual gyrus, a region along the ventral visual stream (Ungerleider & Haxby, 1994), to processing intermediate visual forms (e.g., curvature, radial and concentric patterns; (Gallant et al., 2000) and more visually complex images (Machielsen et al.,

2000) as well as encoding visual memories (Bogousslavsky et al., 1987; Nenert et al., 2014; Rombouts et al., 1999; Ueno et al., 2007). There is also some evidence that the lingual gyrus, along with the fusiform gyrus, is important for object color processing (Chao & Martin, 1999; Hsu et al., 2012; Miceli et al., 2001; Simmons et al., 2007; Ueno et al., 2007b; Wang et al., 2013; but see Bogousslavsky et al., 1987; McKeefry & Zeki, 1997), though these two regions appear to contribute to color processing in different ways. According to some research, the lingual gyrus responds to any color stimuli regardless of task (e.g., as a sensory response), whereas the fusiform gyrus responds to color stimuli in tasks that require attention to color information (Beauchamp et al., 1999; Hsu et al., 2012; Simmons et al., 2007). In short, color processing may be important for processing increasingly racially defined faces, which converges with behavioral evidence showing that in children and adults, skin color plays a prominent role in race categorization (Dunham et al., 2015). Additionally, other work has shown involvement of the IFG and right medial frontal cortex in processing own-race vs. other-race faces, with these frontal regions suggested to support top-down processing of visual stimuli like faces (Feng et al., 2011). Why top-down processing of visual stimuli would increase for more racially defined faces is unclear. Prior work has proposed that activation of the left IFG is found when people have to select among competing sources of information to guide a response (Thompson-Schill et al., 1997); perhaps, top-down processing of more racially defined faces leads people to shift their attention to the many features that might help one disambiguate the race of a face. Our results also indicate that increased activity in the mid orbital gyrus is linked to increased racial ambiguity. This finding aligns well with work showing that the orbitofrontal cortex (OFC) is recruited for decisions under uncertainty or ambiguity (Elliott et al., 2000; Hsu, 2005; Krain et al., 2006). Intriguingly, this region appears to be involved even when participants in the scanner were not actively making decisions that depended on assessing the racial ambiguity of faces.

Even though other work has probed the neural bases of race perception, these studies typically focus on how Caucasians perceive Black and White faces. We show that both Black and White people perceive faces in a categorical manner with no evidence of group differences in the neural basis of categorical perception of race, though we acknowledge the limited sample sizes for this group comparison. Moreover, within one paradigm, we reveal different components of race perception, from the lower-level linear tracking of race-related facial features to the perception of discrete race categories. We also show that, while participants may use different thresholds for their perception of faces as Black or White, the categorical nature of their perception elicits similar patterns of activity.

4.1. Other considerations

The focus on categorical perception of race was inspired by social psychological and sociological work that reveal categorization of racially ambiguous people to reflect the principle of hypodescent (i.e., categorization of mixed-race individuals to the minority group) (Davis, 1991). That is, people tend to categorize Black-White biracial individuals as Black than as White (Halberstadt et al., 2011; Ho et al., 2011; Peery & Bodenhausen, 2008). However, we acknowledge that, while race is, by and large, perceived in a categorical manner, race sometimes can also be perceived in a continuous manner (indeed, a small minority of our participants showed this pattern). Social psychology research reveals that categorization of racially ambiguous faces is subject to factors such as political orientation (Krosch et al., 2013) and the availability of labeling options (Chen & Hamilton, 2012). When people have the explicit option to use the Multiracial label for Black-White biracial individuals, the majority of people use the

Multiracial label; however, the frequency with which it is used is less than the use of the label White for White individuals and Black for Black individuals, and moreover, the likelihood of using the Multiracial label for Black-White biracial individuals decreases under cognitive load or time pressure (Chen & Hamilton, 2012). Some researchers therefore suggest that, while people are able to categorize Black-White biracial individuals as Multiracial when making more deliberate or reflective categorizations, they nevertheless categorize biracial individuals as Black when making rapid or reflexive categorizations (Peery & Bodenhausen, 2008). Our aim in the current work was to probe this rapid, reflexive categorization of race, given its contribution to the formation and expression of implicit racial attitudes and prejudice–the focus of abundant work in social psychology (Dovidio et al., 2002; Dovidio et al., 1997; Goff et al., 2008; Greenwald et al., 1998; Lai et al., 2014; McConnell & Leibold, 2001; Wittenbrink et al., 1997).

We want to end by recognizing that race, along with many other features, is subject to cultural norms and changes in public perception. For example, in the United States, each census, since the first one in 1790, has contained at least one question about racial identity. However, the ways in which questions have been posed and answers have been coded have changed dramatically over time (Pew Research Center, 2015). For example, in 1980 and 1990, if a respondent marked more than one race category (e.g., White; Black; American Indian, Eskimo, and Aleut; Asian and Pacific Islander), the Census Bureau re-categorized that person to a single race, typically the race of the respondent's mother. By contrast, in the year 2000, the U.S. Census Bureau introduced the option to be classified as more than one race, acknowledging the ever-increasing number of multiracial individuals in the United States. Perhaps with greater exposure to multiracial individuals, people may transition to think about race in a less categorical manner. Other constructed features such as gender, which has traditionally been viewed as categorial and binary (Campanella et al., 2001), can now be conceived differently: individuals can experience their gender outside the binary or view gender as a continuum (Galupo et al., 2017; Hyde et al., 2019). How people conceive of race, too, could shift in that direction over time.

4.2 Conclusion

Researchers have consistently focused on select areas when studying the neural basis of race processing: the visual cortex, fusiform gyrus, FFA, and the amygdala. In the current study on categorical race perception, we extend and complement the previous evidence by showing that categorical perception is mediated not only by regions along the late visual processing stream but also by higher-level cortical regions that fall outside these typical regions of interest. Our use of the whole brain searchlight approach combined with ROI-based MVPA afforded an investigation of the entire cortical network involving perceptual, attentional, and conceptual processing systems, contributing to greater insight into the neural and cognitive correlates of categorical perception of race.

Acknowledgments

We thank the staff at the Center for Brain Science at Harvard and the Martinos Center at MIT for their support and members of the Boston College Morality Lab for useful discussions. This work was supported by the National Science Foundation (L.Y.; grant number 1627157) and a National Science Foundation Graduate Research Fellowship (L.T.; grant number 1258923).

References

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, *30*(3), 1059–1068. https://doi.org/10.1016/j.neuroimage.2005.10.026

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: Falseness, beliefs, or attention. *Journal of Cognitive Neuroscience*, *21*(6), 1179–1192. https://doi.org/10.1162/jocn.2009.21082

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217–239.

Beauchamp, M. S., Haxby, J. V., Jennings, J. E., & DeYoe, E. A. (1999). An fMRI version of the Farnsworth-Munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *9*(3), 257–263.

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, *79*, 201–212. https://doi.org/10.1016/j.neuroimage.2013.04.093

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*(1), 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823

Bogousslavsky, J., Miklossy, J., Deruaz, J. P., Assal, G., & Regli, F. (1987). Lingual and fusiform gyri in visual processing: A clinico-pathologic study of superior altitudinal hemianopia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *50*(5), 607–614.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition*, *8*(2), 237–262. https://doi.org/10.1080/13506280042000072

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*(11), 1428–1432. https://doi.org/10.1038/nn.2641

Chao, L. L., & Martin, A. (1999). Cortical regions associated with perceiving, naming, and knowing about colors. *Cognitive Neuroscience, Journal Of*, *11*(1), 25–35.

Chen, J. M., & Hamilton, D. L. (2012). Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology*, *48*(1), 152–164. https://doi.org/10.1016/j.jesp.2011.10.005

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel patterns in fusiform face area differentiate faces by sex and race. *PLoS ONE*, *8*(7), e69684. https://doi.org/10.1371/journal.pone.0069684

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*(2–3), 109–114.

Davis, F. J. (1991). *Who is black?: One nation's definition*. Pennsylvania State University Press.

Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience*, *20*(7), 1174–1188. https://doi.org/10.1162/jocn.2008.20081

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS computational biology, 13(4), e1005508.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). FMRI item analysis in a theory of mind task. *NeuroImage*, *55*(2), 705–712. https://doi.org/10.1016/j.neuroimage.2010.12.040

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68. https://doi.org/10.1037//0022-3514.82.1.62

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540. https://doi.org/10.1006/jesp.1997.1331

Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2015). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science*, *18*(3), 469–483. https://doi.org/10.1111/desc.12228

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex*, *10*(3), 308–317.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227–240.

Feng, L., Liu, J., Wang, Z., Li, J., Li, L., Ge, L., Tian, J., & Lee, K. (2011). The other face of the other-race effect: An fMRI investigation of the other-race face categorization advantage. *Neuropsychologia*, *49*(13), 3739–3749. https://doi.org/10.1016/j.neuropsychologia.2011.09.031

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*(2), 109–128.

Fox, C., Moon, S., Iaria, G., & Barton, J. (2009). The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *NeuroImage*, *44*(2), 569–580. https://doi.org/10.1016/j.neuroimage.2008.09.011

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences*, *105*(47), 18221–18225. https://doi.org/10.1073/pnas.0809952105

Freeman, J. B., Rule, N. O., Adams, R. B., & Ambady, N. (2010). The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. *Cerebral Cortex*, *20*(6), 1314–1322. https://doi.org/10.1093/cercor/bhp195

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. In *Advances in Experimental Social Psychology* (Vol. 61, pp. 237–287). Elsevier. https://doi.org/10.1016/bs.aesp.2019.09.005

Freeman, J. B., Stolier, R. M., Brooks, J. A., & Stillerman, B. S. (2018). The neural representational geometry of social perception. *Current Opinion in Psychology*, *24*, 83–91. https://doi.org/10.1016/j.copsyc.2018.10.003

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., & Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent: Assessing focal activations by spatial extent. *Human Brain Mapping*, *1*(3), 210–220. https://doi.org/10.1002/hbm.460010306

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of “theory of mind” in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11–21.

Gallant, J. L., Shoup, R. E., & Mazer, J. A. (2000). A human extrastriate area functionally homologous to macaque V4. *Neuron*, *27*(2), 227–235.

Galupo, M. P., Pulice-Farrow, L., & Ramirez, J. L. (2017). “Like a Constantly Flowing River”: Gender Identity Flexibility Among Nonbinary Transgender Individuals. In J. D. Sinnott (Ed.), *Identity Flexibility During Adulthood* (pp. 163–177). Springer International Publishing. https://doi.org/10.1007/978-3-319-55658-1_10

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803–1814. https://doi.org/10.1162/jocn.2007.19.11.1803

Goff, P. A., Eberhardt, J. L., Williams, M. J., & Jackson, M. C. (2008). Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences. *Journal of Personality and Social Psychology*, *94*(2), 292–306. https://doi.org/10.1037/0022-3514.94.2.292

Golby, A. J., Gabrieli, J. D., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience*, *4*(8), 845–850.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology. General*, *123*(2), 178–200.

Greenwald AG, McGhee DE, Schwartz JL. 1998. Measuring individual differences in implicit cognition: the implicit association test. J Pers Soc Psychol. 74:1464–1480.

Halberstadt, J., Sherman, S. J., & Sherman, J. W. (2011). Why Barack Obama is Black: A cognitive account of hypodescent. *Psychological Science*, *22*(1), 29–33. https://doi.org/10.1177/0956797610390383

Ho, A. K., Sidanius, J., Levin, D. T., & Banaji, M. R. (2011). Evidence for hypodescent and racial hierarchy in the categorization and perception of biracial individuals. *Journal of Personality and Social Psychology*, *100*(3), 492–506. https://doi.org/10.1037/a0021562

Hsu, M. (2005). Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making. *Science*, *310*(5754), 1680–1683. https://doi.org/10.1126/science.1115327

Hsu, N. S., Frankland, S. M., & Thompson-Schill, S. L. (2012). Chromaticity of color perception and object color knowledge. *Neuropsychologia*, *50*(2), 327–333. https://doi.org/10.1016/j.neuropsychologia.2011.12.003

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, *74*(2), 171–193. https://doi.org/10.1037/amp0000307

Japee, S., Holiday, K., Satyshur, M. D., Mukai, I., & Ungerleider, L. G. (2015). A role of right middle frontal gyrus in reorienting of attention: A case study. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00023

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. *The Neuroscientist*, *20*(5), 509–521. https://doi.org/10.1177/1073858413514136

Klein, M. E., & Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, *49*(5), 878–887. https://doi.org/10.1016/j.neuropsychologia.2011.01.008

Klein, M. E., & Zatorre, R. J. (2015). Representations of invariant musical categories are decodable by pattern analysis of locally distributed BOLD responses in superior temporal and intraparietal sulci. *Cerebral Cortex*, *25*(7), 1947–1957. https://doi.org/10.1093/cercor/bhu003

Krain, A. L., Wilson, A. M., Arbuckle, R., Castellanos, F. X., & Milham, M. P. (2006). Distinct neural mechanisms of risk and ambiguity: A meta-analysis of decision-making. *NeuroImage*, *32*(1), 477–484. https://doi.org/10.1016/j.neuroimage.2006.02.047

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Krosch, A. R., Berntsen, L., Amodio, D. M., Jost, J. T., & Van Bavel, J. J. (2013). On the ideology of hypodescent: Political conservatism predicts categorization of racially ambiguous faces as Black. *Journal of Experimental Social Psychology*, *49*(6), 1196–1203. https://doi.org/10.1016/j.jesp.2013.05.009

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., … Nosek, B. A. (2014). Reducing implicit racial preferences: A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785. https://doi.org/10.1037/a0036260

Lee, Y.-S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *NeuroImage*, *57*(1), 293–300. https://doi.org/10.1016/j.neuroimage.2011.02.006

Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *Journal of Neuroscience*, *32*(11), 3942–3948. https://doi.org/10.1523/JNEUROSCI.3814-11.2012

Levin, D. T., & Angelone, B. L. (2002). Categorical perception of race. *Perception*, *31*(5), 567–578.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and

Caucasian-American individuals. *Nature Neuroscience*, *8*(6), 720–722. https://doi.org/10.1038/nn1465

Machielsen, W. C., Rombouts, S. A., Barkhof, F., Scheltens, P., & Witter, M. P. (2000). FMRI of visual encoding: Reproducibility of activation. *Human Brain Mapping*, *9*(3), 156–164.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442. https://doi.org/10.1006/jesp.2000.1470

McKeefry, D. J., & Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*, *120*(12), 2229–2242.

Miceli, G., Fouch, E., Capasso, R., Shelton, J. R., Tomaiuolo, F., & Caramazza, A. (2001). The dissociation of color from form and function knowledge. *Nature Neuroscience*, *4*(6), 662–667.

Myers, E. B., & Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *Journal of Cognitive Neuroscience*, *24*(8), 1695–1708. https://doi.org/10.1162/jocn_a_00243

Natu, V., Raboy, D., & O'Toole, A. J. (2011). Neural correlates of own- and other-race face perception: Spatial and temporal response differences. *NeuroImage*, *54*(3), 2547–2555. https://doi.org/10.1016/j.neuroimage.2010.10.006

Nenert, R., Allendorfer, J. B., & Szaflarski, J. P. (2014). A model for visual memory encoding. *PLoS ONE*, *9*(10), e107761. https://doi.org/10.1371/journal.pone.0107761

Newell, F. N., & Bülthoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, *85*(2), 113–143.

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition*, *95*(2), B1–B14. https://doi.org/10.1016/j.cognition.2004.07.002

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 566–570. https://doi.org/10.1016/j.tics.2014.09.007

Oosterhof, N. N., Connolly, A. C., and Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab / GNU Octave. Frontiers in Neuroinformatics, doi:10.3389/fninf.2016.00027.

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*(4), 477–493. https://doi.org/10.1037//0096-3445.131.4.477

Parlatini, V., Radua, J., Dell'Acqua, F., Leslie, A., Simmons, A., Murphy, D. G., Catani, M., & Thiebaut de Schotten, M. (2017). Functional segregation and integration within fronto-parietal networks. *NeuroImage*, *146*, 367–375. https://doi.org/10.1016/j.neuroimage.2016.08.031

Peery, D., & Bodenhausen, G. V. (2008). Black + White = Black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science*, *19*(10), 973–977. https://doi.org/10.1111/j.1467-9280.2008.02185.x

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*, *45*(1), S199–S209. https://doi.org/10.1016/j.neuroimage.2008.11.007

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*(3–4), 245–258. https://doi.org/10.1080/17470910600989896

Pew Research Center. (2015). *Multiracial in America: Proud, diverse and growing in numbers*. http://www.pewsocialtrends.org/2015/06/11/multiracial-in-america/

Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience*, *12*(2), 221–228. https://doi.org/10.1038/nn.2246

Raizada, R. D. S., & Lee, Y.-S. (2013). Smoothness without smoothing: Why Gaussian Naive Bayes is not naive for multi-subject searchlight studies. *PLoS ONE*, *8*(7), e69566. https://doi.org/10.1371/journal.pone.0069566

Raizada, R. D. S., & Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*(4), 726–740. https://doi.org/10.1016/j.neuron.2007.11.001

Ratner, K. G., Kaul, C., & Van Bavel, J. J. (2013). Is race erased? Decoding race from patterns of neural activity when skin color is not diagnostic of group boundaries. *Social Cognitive and Affective Neuroscience*, *8*(7), 750–755. https://doi.org/10.1093/scan/nss063

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*(2), 752–762. https://doi.org/10.1016/j.cognition.2007.09.001

Rombouts, S. A., Scheltens, P., Machielson, W. C., Barkhof, F., Hoogenraad, F. G., Veltman, D. J., Valk, J., & Witter, M. P. (1999). Parametric fMRI analysis of visual encoding in the human medial temporal lobe. *Hippocampus*, *9*(6), 637–643. https://doi.org/10.1002/(SICI)1098-1063(1999)9:6<637::AID-HIPO4>3.0.CO;2-V

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2004). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience*, *8*(1), 107–113. https://doi.org/10.1038/nn1370

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind.” *NeuroImage*, *19*, 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810. https://doi.org/10.1016/j.neuropsychologia.2007.05.002

Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Ueno, A., Abe, N., Suzuki, M., Hirayama, K., Mori, E., Tashiro, M., Itoh, M., & Fujii, T. (2007). Reactivation of medial temporal lobe and occipital lobe during the retrieval of color information: A positron emission tomography study. *NeuroImage*, *34*(3), 1292–1298. https://doi.org/10.1016/j.neuroimage.2006.10.022

Ungerleider, L. G., & Haxby, J. V. (1994). 'What'and 'where'in the human brain. *Current Opinion in Neurobiology*, *4*(2), 157–165.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias a functional magnetic resonance imaging investigation. *Psychological Science*, *19*(11), 1131–1139.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: Evidence of in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience*, *23*(11), 3343–3354.

Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cerebral Cortex*, *21*(12), 2788–2796. https://doi.org/10.1093/cercor/bhr074

Wang, X., Han, Z., He, Y., Caramazza, A., Song, L., & Bi, Y. (2013). Where color rests: Spontaneous brain activity of bilateral fusiform and lingual regions predicts object color knowledge performance. *NeuroImage*, *76*, 252–263. https://doi.org/10.1016/j.neuroimage.2013.03.010

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274.

Worsley, K. J., Evans, A. C., Marrett, S., & Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, *12*(6), 900–918. https://doi.org/10.1038/jcbfm.1992.127