eNeuro Accepted Manuscript

eNeuro

Research Article: Confirmation | Cognition and Behavior

Human face-selective cortex does not distinguish between members of a racial outgroup

https://doi.org/10.1523/ENEURO.0431-19.2020

Cite as: eNeuro 2020; 10.1523/ENEURO.0431-19.2020

Received: 16 October 2019
Revised: 4 April 2020
Accepted: 9 April 2020

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2020 Reggev et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

eNeuro Accepted Manuscript

Manuscript title: Human face-selective cortex does not distinguish between members of a racial outgroup

Abbreviated title: Face-selective cortex disregards outgroup identity

Niv Reggev*[1,2,3], Kirstan Brodie[1], Mina Cikara[1], & Jason P. Mitchell[1]

Author Contributions. N.R. and J.P.M. conceived the studies. All authors contributed to study design, analytic strategy, and interpretation of the data. N.R. collected and analyzed the data. N.R. and J.P.M drafted the manuscript, and K.B. and M.C provided critical revisions. All authors approved the final version of the manuscript for submission.

Correspondence should be addressed to Niv Reggev, reggevn@bgu.ac.il

Number of Figures: 4.

Number of words for abstract: 168.

Number of words for significance statement: 117.

Number of words for Introduction: 954.

Number of words for Discussion: 1613.

Acknowledgements. We thank J. Ontko for insights and assistance in data collection.

Declaration of Conflicting Interests. The authors declare no conflicts of interest.

Funding sources. This work was partially supported by the Israeli Science Foundation (No. 79/18, to N.R.) and was conducted at the Harvard University Center for Brain Science, which is supported by supported by the NIH Shared Instrumentation Grant Program (No. S10OD020039).

[1] Department of Psychology, Harvard University, Cambridge, MA, 02138, USA.
[2] Department of Psychology, Ben Gurion University of the Negev, P.O. Box 653, Be'er-Sheva, Israel, 84105
[3] Zlotowski Center for Neuroscience, Ben Gurion University of the Negev, P.O. Box 653, Be'er-Sheva, Israel, 84105

eNeuro Accepted Manuscript

Abstract

People often fail to individuate members of social outgroups, a phenomenon known as the outgroup homogeneity effect. Here, we used fMRI repetition suppression to investigate the neural representation underlying this effect. In a pre-registered study, White human perceivers ($N = 29$) responded to pairs of faces depicting White or Black targets. In each pair, the second face depicted either the same target as the first face, a different target from the same race, or a scrambled face outline. We localized face-selective neural regions via an independent task, and demonstrated that neural activity in the fusiform face area distinguished different faces only when targets belonged to the perceivers' racial ingroup (White). By contrast, face-selective cortex did not discriminate between other-race individuals. Moreover, across two studies (total $N = 67$) perceivers were slower to discriminate between different outgroup members and remembered them to a lesser extent. Together, these results suggest that the outgroup homogeneity effect arises when early-to-mid-level visual processing results in an erroneous overlap of representations of outgroup members.

eNeuro Accepted Manuscript

Significance statement

Researchers have repeatedly demonstrated that perceivers struggle to distinguish between different members of a racial outgroup. Here, we show in a pre-registered study that this failure arises when areas of the human brain that specifically process facial identity—most notably, the so-called "fusiform face area"—fail to detect differences between identities of members of a racial outgroup. When White perceivers viewed photos of two different Black men, the face area of their brains responded as if the two photos portrayed the same person. This effect was constrained to outgroup faces; the face area successfully distinguished faces of two different White individuals. Our results highlight the failure of basic representational mechanisms in processing individuals from other social groups.

eNeuro Accepted Manuscript

INTRODUCTION

The outgroup homogeneity effect (sometimes called the cross-race effect) describes the difficulty people often experience when trying to identify members of a racial outgroup. Although race is not a valid biological taxonomy, individuals in our society define separate human races by sociocultural experiences and use this social taxonomy as a basis for numerous social and cognitive processes (Wagner et al., 2017). For example, White perceivers typically remember faces of White targets better than faces of Black targets, and are more likely to say that a new, unfamiliar Black face is the same as one they have seen previously (Malpass and Kravitz, 1969; Meissner and Brigham, 2001). Perceivers also identify own-race faces faster and more accurately than cross-race faces (Marcon et al., 2010). This pattern of (mis)identification has demonstrable societal consequences—in an analysis of American police proceedings, White witnesses correctly identified 60% of perpetrators when a line-up comprised other White individuals, but identified only 45% of Black perpetrators; more than half the time, unrelated (i.e., innocent) Black individuals were identified as perpetrators (Behrman and Davey, 2001).

Most theories of the outgroup homogeneity effect suggest that perceivers create detailed, individuated representations of ingroup members, but view outgroup members as interchangeable instances of a category (Papesh and Goldinger, 2010; Hugenberg et al., 2013; Correll et al., 2017). Consequently, scholars hypothesize that this differential representation underlies perceivers' improved memory for members of their ingroup, as well as their heightened judgments of diversity and distinctiveness for ingroup targets (Park and Rothbart, 1982; Judd et al., 2005; Boldry et al., 2007). However, research to date has measured representations of racial in- and outgroup members mainly as a function of behavioral responses to in- and outgroup targets (e.g., reaction time differences between different targets; Papesh and Goldinger, 2010).

eNeuro Accepted Manuscript

To the best of our knowledge, no study has directly measured the target-specific representations of members of different groups, which we hypothesize underlie the previously mentioned behavioral results.

To bridge this gap, we conducted a functional magnetic resonance imaging (fMRI) study that made use of the phenomenon of *repetition suppression*, whereby neural responses to a repeated stimulus are reduced (or suppressed) relative to a sequence of two different stimuli (Gotts et al., 2012). For example, regions of the human brain that respond robustly to faces—such as the fusiform gyrus—will decrease their activity when participants view the same face repeatedly. By contrast, activation in this region will return to typically high levels of activation when a new face is presented (Gilaie-Dotan et al., 2010). Accordingly, this "release from suppression" can be used to measure the degree to which perceivers detect that they have viewed the faces of two different individuals. Thus, our approach extends previous neuroimaging studies that have, for the most part, characterized differences in mean levels of neural activity between social groups (e.g., Golby et al., 2001; Van Bavel et al., 2008; Mathur et al., 2012). Whereas these traditional univariate approaches can offer insights into the localization of neural activity differences in response to social groups, the release from suppression effect, by contrast, characterizes the uniqueness (or similarity) of representations of distinct Black and distinct White faces.

Here, we make use of the release-from-suppression logic to examine the representational basis of the outgroup homogeneity effect. Researchers have shown that the fusiform face area (FFA) shows repetition suppression even when participants see a single individual from different angles (Pourtois et al., 2005) or from different distances (Grill-Spector et al., 1999), suggesting that the FFA is sensitive to the identity of an individual rather than the similarity of their

eNeuro Accepted Manuscript

perceptual features. We capitalized on these characteristics to examine whether White perceivers will likewise represent two different Black individuals as more similar to each other than two White individuals. Specifically, we hypothesized that if perceivers individuate ingroup and outgroup faces equally, then we should observe similar release from suppression for two different face identities regardless of their race. However, to the extent that White perceivers are worse at "detecting a difference" between faces of different Black targets, then the FFA should show more suppression when a Black face is followed by a new, different Black face, despite the second face being different from the first (for related approaches, see Vizioli et al., 2010; Hughes et al., 2019). Importantly, only identity-sensitive regions such as the FFA should show the differential suppression for Black faces; face-specific identity-*insensitive* regions (e.g., occipital face area; OFA) should show equivalent suppression for different Black and White faces.

To test these predictions, we conducted a behavioral experiment and a pre-registered fMRI study using a repetition suppression paradigm. In each experiment, White participants sequentially viewed pairs of faces that varied in race (Black, White) and gender (woman, man). We matched face categories in perceptual and structural properties (see Materials and Methods). For each pair, participants indicated whether the faces were of the same or different individuals. In some trials, the two faces were identical; in an equal number of trials, the two faces depicted different individuals of the same race and gender. In addition, in one-third of trials in the fMRI experiment, the second face was replaced by a scrambled face-shaped patch; this condition allowed us to establish a difference in baseline neural processing of Black and White faces, and to directly replicate earlier studies (Golby et al., 2001; Van Bavel et al., 2011; Fig 1). Our behavioral pilot studies indicated that participants demonstrated a reliable race effect only for male targets. Therefore, our confirmatory analyses focused on male targets; we report the results of the

eNeuro Accepted Manuscript

exploratory analyses for female targets in the supporting information. Notably, most previous studies of the other-race homogeneity effect have included only male targets; here we provide an initial attempt to address this empirical lacuna.

MATERIALS AND METHODS

Human participants were recruited from the local community using the Harvard Department of Psychology Study Pool website. All participants provided their informed consent in a manner approved by the Committee on the Use of Human Subjects in Research at Harvard University.

Experiment 1. Thirty-eight self-identified White participants completed the Experiment (22 female, 16 male; mean age: 21.00, standard deviation: 2.71, range: 16-29). We excluded one additional participant who failed to respond on time to more than 20% of the trials in the face identification task. We collected the data on a rolling basis (multiple slots per day) with a target sample size of 32 participants. We identified this sample size to be sufficient to achieve a power of 0.8 to detect a hypothesized meaningful effect size estimate (Cohen's $d = 0.3$, approximately equivalent to $\eta^2_p = 0.03$; Cohen, 1988, pp. 276-281) for the interaction contrast of race (Black/White) by condition (repeated/different). Participants completed the Experiment for course credit or financial compensation (US$10).

Experiment 2. Based on an identical power calculation, we aimed to collect analyzable data from 32 participants. To achieve this goal, we collected data from 38 self-identified White participants. In line with our preregistered exclusion criteria, we excluded six participants prior to data inspection due to lack of response to more than 20% of the trials in the main task. In the main text, we report the results from 29 participants (15 female, 13 male, 1 non-binary; mean

eNeuro Accepted Manuscript

age: 21.97, standard deviation: 3.02, range: 18-28) because we excluded three additional participants whose reaction time or accuracy data were more than 2 standard deviations beyond the sample mean (we report analysis with the full sample in the supporting information—the results do not change). All 32 participants provided adequate data in terms of signal quality, as measured by a slice signal to noise ratio higher than 150 and by having no more than 3 discrete movements larger than 0.5mm. All participants were healthy, right-handed, native English speakers with normal or corrected-to-normal vision and no history of neurological or psychiatric conditions. Participants completed the Experiment for course credit or financial compensation (US$50).

Materials. We obtained the face images for this study from the Chicago Face Database (CFD) (Ma et al., 2015). We excluded faces that were identified as belonging to their respective racial group by fewer than 60% of the CFD independent raters, and faces that were identified as belonging to their racial group by between 60-75% of CFD raters were reviewed by two additional independent raters (N.R. and K.B.). Of these faces, 12 were excluded for having atypical features for their respective racial group, determined by interrater agreement between the additional raters. Lastly, all faces that were identified by 75% or more of CFD raters as belonging to their respective racial group were individually reviewed. Of these faces, 2 were excluded due to unique/distinctive facial features such as scars, and 3 were excluded due to noticeable artifacts in the image quality. This review process resulted in a final total of 94 Black female faces, 76 Black male faces, 77 White female faces, and 81 White male faces. This allowed us to select a total of 76 unique faces from each category for each participant (we used only 72 faces in Experiment 1 as it required fewer stimuli; see task description below). We converted all images to greyscale, matched them on luminance (separately for foreground and

eNeuro Accepted Manuscript

background) using the SHINE toolbox (Willenbockel et al., 2010) for MATLAB (Mathworks, Natick, MA, USA), cropped them to 1246 by 946 pixels, matched them on spatial frequency using the SHINE toolbox, and finally, resized all photos to 199 by 262 pixels. All face stimuli were presented in a rectangular box that included hair outline (see Figure 1).

To confirm that differences among the different categories were not confounded with image similarity, we measured the similarity between images by calculating the Structural Similarity Index (SSIM; Wang et al., 2004). This measure was computed after the experiment was completed. We computed the pairwise SSIM between all images in each category and averaged the SSIM score per image. This resulted in a vector of SSIM scores per category. Black and White men did not significantly differ in their average similarity (0.7169 versus 0.7148, respectively; $t_{(155)} = 0.75$, $p = 0.46$). Black women, however, were less similar to each other compared with White women (0.7012 versus 0.7146, respectively; $t_{(155)} = 5.31$, $p < 0.001$). This difference limits the potential interpretation of our exploratory analyses for female target faces; see Figure 3-5 and discussion below.

Procedure. We presented all tasks and stimuli via PsychoPy v1.84.2 (Peirce, 2007) running under Mac OS X 10.7 (Experiment 1) or Windows 7 (Experiment 2).

Face repetition suppression task. During the main task, we presented face images to participants in same-race, same-gender pairs. Experiment 1 included two conditions. In the repeated-face condition, the second face was identical to the first face. In the different-face condition, the second face was different from the first face. Experiment 2 included both the repeated-face condition and the different-face condition, as well as an additional single-face condition, where a single face was followed by a scrambled visual patch in the shape of a face; these pairs were used to establish a baseline. In Experiment 1, participants saw 24 face pairs per

eNeuro Accepted Manuscript

187 condition for each race (12 pairs for each gender). Experiment 2 included 38 face pairs per
188 condition for each race (19 for each gender).

189 Each trial in the face task began with a face presented for 600 milliseconds, followed by a
190 fixation crosshair presented in the center of the screen for 500 milliseconds. Then, a second face,
191 along with a response prompt, was presented for 600 milliseconds. Lastly, a final fixation
192 crosshair was presented for 300 milliseconds, for a total duration of 2 seconds per trial. A jittered
193 inter-trial interval (range 0-7, mean = 0.55, s.d. = 1.03) then followed (see below). Participants
194 used two fingers of their left hand to indicate, for each pair of faces, whether the second face was
195 the same as or different from the first. Participants provided their response while the second face
196 and response prompt were being presented on screen. The locations of both the first and second
197 faces on the screen included a randomized horizontal offset (within a predetermined range) to
198 minimize the interference of visual after-effects. For each participant, no single face image was
199 used in more than one pair. The specific faces assigned to each of the conditions, as well as their
200 pairings, were randomized between participants.

201 Before beginning the main task, participants practiced the task in order to become
202 acquainted with trial structure and speed. In Experiment 2, participants completed two rounds of
203 practice before entering the fMRI scanner, and then completed an additional practice round after
204 entering the scanner before they started the main task. Face images used in the practice rounds
205 were drawn from the face images that were excluded from the main task stimuli.

206 In Experiment 2, trials were divided into two runs with an equal distribution of conditions
207 between runs. To optimize estimation of the event-related fMRI response, conditions were
208 intermixed in a pseudo-random order and separated by a variable, algorithm-based inter-stimulus
209 interval consisting of a fixation crosshair. We used OptSeq2 (Dale, 1999) to generate sequences

eNeuro Accepted Manuscript

optimized for efficiency of the contrast (single>repeated) for a first-order counterbalanced event sequence. Of these sequences, we selected 4 sequences that contained no more than 6 consecutive trials of the same race. We randomly assigned (with replacement) an event sequence for each functional run to avoid spurious results attributable to differences between conditions in one specific event sequence (Mumford et al., 2014). Within condition, trials were presented in a unique random order for each participant. During the task, we measured behavioral task performance, including accuracy and response reaction time.

Face functional dynamic localizer task. After completing the main task, participants in Experiment 2 completed the dynamic localizer task (Pitcher et al., 2011) to localize brain regions associated with the processing of faces. Participants were informed about this task only upon its execution. The dynamic localizer task instructed participants to respond via a button press to dynamic stimuli – short (3 second) movie clips of various categories. The stimuli were grouped into 5 categories: faces, objects, bodies, landscape scenes, and scrambled objects. All faces and bodies belonged to White individuals. We chose these specific stimuli as they have been previously validated in a dynamic task (Pitcher et al., 2011). To the best of our knowledge, no previous studies have validated non-White dynamic face stimuli. Each run was presented in the following structure: first, a fixation crosshair was presented for 18 seconds; then, movie clips were presented back-to-back with no intertrial interval. Movie clips were blocked by category such that each block contained six video clips, of about 3 seconds each, all portraying the same category (e.g., block 1 contained only clips of faces, block 2 contained only clips of objects, etc.). Then, another fixation crosshair was presented for 18 seconds, followed by another series of different video clips organized like the first, but with the order of categories reversed. The run concluded with a final fixation crosshair presented for 18 seconds. Each run lasted 234 seconds,

eNeuro Accepted Manuscript

and there were four runs total. Participants were asked to press a button when they saw a repeated stimulus (a 1-back repetition detection); for each run, there was one repetition within each category. We implemented the task by adapting code written by Matteo Visconti di Oleggio Castello (Castello, 2017).

Post-task behavioral measures. Upon completion of the face task, participants in Experiment 1 completed a surprise recognition memory task. For this task, two faces were presented side-by-side on a screen, with one face having been presented previously, and one face being completely novel. Thirty-six pairs of faces per race per gender were presented (144 face pairs in total). Participants used 4 keys to indicate which was the previously presented face, given four choices: surely left, maybe left, maybe right, or surely right.

Participants in both experiments also completed the following questionnaires and behavioral measures: External and Internal Motivation to Control Prejudice (Plant and Devine, 1998), Social Dominance Orientation (Ho et al., 2015), and the Implicit Association Test (IAT; Greenwald et al., 1998) using Black and White faces as target stimuli with positive and negative categories (Nosek et al., 2007). Finally, in order to assess the degree of outgroup contact, participants responded to the following 3 items (all requiring open-ended responses, anchored at 0): “How many African-American friends do you have?”, “In a typical week, how many times do you meet with African-American friends?”, “How many of your close friends or family have African-American friends?”. Aside from the recognition memory task, all post-scan behavioral measures and questionnaires were included for the purposes of exploratory analysis, as noted in the preregistration. It is our intent to conduct further studies, and to eventually aggregate the results of these measures across studies once a suitable power is attained.

eNeuro Accepted Manuscript

fMRI acquisition and preprocessing. We collected all images with a 3T Siemens Prisma scanner system (Siemens Medical Systems, Erlangen, Germany) using a 64-channel radiofrequency head coil. First, we acquired high-resolution anatomical images using a T1-weighted 3D MPRAGE sequence (TR = 2200 ms, TI = 1100 ms, acquisition matrix = 256 × 256 × 176, flip angle = 7, voxel size = 1 × 1 × 1 mm^3). Second, we acquired a fieldmap in the same plane as the functional images (see below) to correct for inhomogeneities in the magnetic field (Cusack and Papadakis, 2002). Next, we collected whole brain functional images using a simultaneous multi-slice (multiband) T2*-weighted gradient echo sequence, sensitive to BOLD contrast, developed at the Center for Magnetic Resonance Research (CMRR) at the University of Minnesota (Feinberg et al., 2010; Moeller et al., 2010; Xu et al., 2013) (TR = 2000 ms, TE = 30 ms, voxel size = 2 × 2 × 2 mm^3, 75 slices auto-aligned to -25 degrees of the AC-PC line, image matrix = 104 × 104, FOV = 208 * 208 mm^2, flip angle = 75º, GRAPPA acceleration factor = 2, multiband factor = 3, phase encoding direction = A -> P). Following a short in-scanner practice scan, the face repetition suppression task included two runs consisting of 188 volumes each and was followed by the dynamic face localizer task including four runs, 120 volumes each; all runs were complemented by two additional dummy scans and an initial period of approximately 26 seconds dedicated to references for the GRAPPA procedure. The first three volumes from each run (i.e., in addition to dummy scans) were discarded to ensure T1 equilibrium. The last 5 volumes from the face repetition suppression runs always included a crosshair fixation to ensure the appropriate estimation of the hemodynamic function for the last events in each run.

We conducted rudimentary quality control using the recommendations for the quality control tool implemented at the scanner facility. We used SPM12 version 6225 (Wellcome Department of Cognitive Neurology, London, UK) on a 2015b MATLAB platform (Mathworks,

eNeuro Accepted Manuscript

Natick, MA, USA) to process and analyze the fMRI data. We corrected functional data for differences in acquisition time between slices, corrected for inhomogeneities in the magnetic field using the fieldmap (Cusack and Papadakis, 2002), realigned to the first image to correct for head movement using a 2nd degree B-spline interpolation, unwarped to account for residual movement related variance using a 4th degree B-spline interpolation, and co-registered with each participant's anatomical data. Then, we transformed the functional data into a standard anatomical space (2 mm isotropic voxels) based on the ICBM152 brain template (Montreal Neurological Institute). We then spatially smoothed (5 mm full-width at half-maximum, FWHM) normalized data using a Gaussian Kernel.

Statistical Analysis

Statistical modeling. Behavioral data. We analyzed reaction time and accuracy data with mixed models as implemented in the lme4 package version 1.1-14 (Bates et al., 2014) for R version 3.4.2 (R Core Team, 2017). To avoid transformation of raw reaction time data, we used generalized linear mixed models (gLMMs) with the inverse Gaussian identity link (Lo and Andrews, 2015). Memory performance in Experiment 1 was analyzed using logit generalized linear mixed models with the binomial link (Jaeger, 2008). We included random effects for the intercepts for participants in all analyses. We added random intercepts for faces and by-participant random slopes for the fixed effect of race if this addition did not result in a convergence failure. Trials that elicited no response (<1.5% of all trials; no difference between conditions) were excluded from reaction time analyses.

Neuroimaging data. We performed statistical analyses using the general linear model (GLM) that included boxcar functions of variable duration determined per trial by reaction time to target faces (i.e. variable epochs). We chose this analysis approach to control for effects of

eNeuro Accepted Manuscript

301 reaction time on the neural response (Grinband et al., 2008). We set the onset of the boxcar
302 function to the onset of prime face presentation on each trial. We deviated from the preregistered
303 protocol by modeling single-face trials as a boxcar function with a fixed duration of 600ms (the
304 duration of presentation for the first face, rather than reaction time to target face) to capture true
305 baseline activity for a single face. The reported results replicated when we conducted the
306 analyses without this deviation (see Fig. 3-3).

307 The model included six conditions per gender (2 races by 3 conditions: repeated faces,
308 different faces, single face). We modeled trials that elicited no response in a separate regressor,
309 and all regressors of interest were convolved with a canonical hemodynamic response function
310 and its temporal derivative. The final first-level GLM was high-pass filtered at 128 s and
311 included nuisance regressors specifying the six motion parameters calculated during the motion
312 correction procedures, their temporal derivative, and a session mean per run. Preregistered
313 validation analyses were conducted with an additional model that included an additional separate
314 regressor for trials in which participants erred (Fig. 3-2).

315 **Regions of interest (ROIs).** We defined ROIs independently from the task localizer data
316 by the Group-Constrained Subject-Specific method (Julian et al., 2012) as implemented in the
317 spm_ss toolbox (Nieto-Castañón and Fedorenko, 2012). Briefly, this method was designed to
318 discover regions that are systematically activated across participants and to define the borders
319 around and between each of these regions. This method identifies key "parcels" within which
320 most participants show activation for the contrast of interest. The selection of functional ROIs
321 for individual participants is then accomplished by intersecting each individual participant's
322 localizer activation map with each of the parcels, thus defining functional ROIs in each
323 individual participant in a fully algorithmic fashion. We applied this method to generate ROIs

eNeuro Accepted Manuscript

that responded to faces over all other categories (the face versus other contrasts; see Figure 3-6 for the full results of this procedure). We then extracted average parameter estimates across voxels from each participant-specific functional ROI using in-house scripts. We analyzed the data using a within-participant 2 (race) by 3 (condition) ANOVA as implemented by afex package (Singmann et al., 2018) for R, version 0.22-1, and plotted the results using the package ggstatsplot (Patil, 2018), version 0.2.0, and the package dabestr (Ho et al., 2019), version 0.2.2.

Statistical inference. For the main analysis of interest (hypotheses 5 and 6 in the preregistration), we focused on the race by condition interaction, with specific focus on two separate interaction contrasts. One interaction contrast tested activation differences between the single-face condition and the repeated-face condition as a function of race, and another interaction contrast tested activation differences between the single-face condition and the different-face condition as a function of race. Follow up simple effects models (one model per condition) tested differences between responses to repeated and different faces separately for Black and White faces. To demonstrate no difference between conditions, we performed an equivalence test using the equivalence package for R, version 0.7.2 (Robinson, 2016).

Open Practices. Experiment 1 was not formally preregistered. All data collection procedures and analytic choices for Experiment 2 were preregistered on the Open Science Framework (OSF; https://osf.io/cw4dj/). We explicitly report any deviations from preregistration in the manuscript and extended data. All de-identified data and code are freely available on the Open Science Framework (https://osf.io/6z5cj/).

RESULTS AND DISCUSSION

eNeuro Accepted Manuscript

345 **Experiment 1.** In Experiment 1, we used a combination of memory and reaction time
346 measures to validate our paradigm. Thirty-eight White participants first viewed 96 repeated or
347 different face pairs (24 per gender, per race). Subsequently, these participants completed a
348 surprise memory task in which they saw pairs of faces—with one previously seen face and one
349 new face in each pair— and indicated which face they had previously viewed. Because pilot
350 testing demonstrated that participants show the effect only for male targets, we report analyses
351 limited to these faces (see Fig. 1-1 for analysis of behavioral responses to female faces). In line
352 with previous studies, participants were more likely to accurately remember faces of White men
353 (61.6% correct) than faces of Black men (55.1% correct; odds ratio [OR] = 1.16, *Wald's z* =
354 3.005, $p = 0.003$; see Fig. 2b for full results). In addition, in the face repetition task, participants
355 were quicker to identify different White faces as different (mean: 451 ± S.E.: 7ms) than different
356 Black faces (475 ± 7ms; $t = 3.664$, $p < 0.001$), but they identified Black and White repeated faces
357 as identical equally fast (419 ± 7ms versus 424 ± 7ms for Black and White faces, respectively; t
358 $= 0.568$, $p > 0.5$; interaction model comparison: $\chi_{(1)} = 16.89$, $p < 0.001$; see Fig. 2a). In other
359 words, participants showed reduced memory and slower responses for male outgroup faces,
360 particularly when they viewed two different individuals from each group. This suggests that
361 although perceivers can successfully process a target from an outgroup if they see it repeatedly,
362 they treat different outgroup faces as more homogenous to one another relative to different
363 ingroup faces.

364 **Experiment 2.** Participants in Experiment 1 demonstrated a behavioral outgroup
365 homogeneity effect in the repetition paradigm. To examine neural representation differences
366 between the groups, Experiment 2 included 29 White participants who performed the task while
367 undergoing fMRI scanning. We used a separate face localizer task (Pitcher et al., 2011) to

eNeuro Accepted Manuscript

identify *a priori* regions-of-interest (ROIs) in right fusiform gyrus (FFA) and right occipital cortex (OFA; see Fig. 3a). We did not identify a robust cluster in a third hypothesized ROI, the anterior temporal lobe; see Materials and Methods for details of localization procedure (Julian et al., 2012).

We then examined patterns of repetition suppression for repeated and different faces. For each condition of interest (repeated and different faces), we computed the differences from baseline (single face) for Black and White targets. In line with our preregistration and Experiment 1, we report the results for male targets only (see Figure 3-5 for results for female targets). When a second face was identical to the first, activation in FFA was suppressed to an equivalent degree for both Black and White faces [repetition effect: $F_{(1,28)} = 160.20$, $MSE = 4.83$, $p < 0.001$, $\eta^2_p = 0.85$; interaction effect: $F_{(1,28)} = 1.34$, $MSE = 0.82$, $p = 0.26$, $\eta^2_p = 0.05$; equivalence test for the two repetition effects (mean difference: 0.39): $p = 0.04$, equivalence interval (-0.96, 0.18)]. In other words, perceivers showed similar levels of suppression in FFA when faces were repeated, regardless of race. By contrast, FFA was released from suppression when the second face differed from the first—but only for White faces. We observed more release from suppression for White faces than for Black faces, suggesting that participants perceived White faces—but not Black faces— as representing different individuals [repetition effect: $F_{(1,28)} = 13.16$, $MSE = 0.57$, $p = 0.001$, $\eta^2_p = 0.32$; interaction between repeating and different faces: $F_{(1,28)} = 7.64$, $MSE = 0.37$, $p = 0.01$, $\eta^2_p = 0.21$; see Fig. 3b]. Remarkably, we observed similar levels of repetition suppression for two different Black individuals as for two identical faces [equivalence test for suppression of different and repeated Black faces (mean difference: 0.24): $p < 0.001$, (-0.52, 0.04)], which suggests that participants did not consistently perceive the two photographs to represent distinct individuals.

eNeuro Accepted Manuscript

391 This pattern of results was selective for the FFA, which is thought to be the earliest visual
392 area that encodes the unique identity of faces, rather than just their distinct perceptual features
393 (Duchaine and Yovel, 2015). Patterns of repetition suppression did not vary for different Black
394 and White faces in either the OFA ($F_{(1,28)} = 0.46$, $MSE = 3.62$, $p = 0.50$, $\eta^2_p = 0.02$) or primary
395 visual cortex ($F_{(1,21)} < 0.01$, $MSE = 5.28$, $p > 0.99$, $\eta^2_p < 0.001$) (see Fig. 4). Finally, results were
396 robust across several pre-registered analytic variations, including analyses with excluded
397 participants as well as when excluding trials that were answered incorrectly (see Materials and
398 Methods and Figs. 3-1, 3-2 and 3-3). Together, these findings suggest that the FFA (but not
399 lower-level visual regions) processes racial outgroup individuals as more similar to each other
400 relative to ingroup individuals.

401 The present results expand on previous investigations in several important ways. First,
402 most studies of the other-race or the other-group effect have documented increased univariate
403 activity in the FFA in response to own-group faces compared to faces from a different social
404 group (Golby et al., 2001; Kim et al., 2006; Van Bavel et al., 2008, 2011; Feng et al., 2011; for a
405 review, see Molenberghs and Louis, 2018). However, mean activity level can be susceptible to
406 multiple moderators, including attention and motivation. Indeed, a recent study demonstrated
407 that participants who experience resource scarcity demonstrate reduced FFA activity for Black
408 faces and increased FFA activity for White faces (Krosch and Amodio, 2019). Thus, mean FFA
409 activity can reflect the influence of contextual factors on face perception, rather than measure the
410 representations underlying the perceived face. Here, instead of analyzing mean activity, we
411 utilize the robust phenomenon of repetition suppression to demonstrate that the FFA utilizes a
412 similar representation for different outgroup – but not ingroup – faces.

eNeuro Accepted Manuscript

Our approach also complements a recent demonstration of differential release from suppression for Black and White targets (Hughes et al., 2019). One outstanding issue concerns the demonstration of the repetition effect. In their study, Hughes and colleagues compared blocks of repeated faces to blocks of faces morphed to different degrees, which assumes equivalent magnitude of neural activity for repeated White and Black faces without explicitly accounting for it. By contrast, our paradigm allowed us to quantify neural activity for baseline, repeated, and different face trials for Black and White faces separately, providing a straightforward index of repetition suppression and release from suppression. Second, we observed behavioral and neural outgroup homogeneity effects within the same paradigm, whereas Hughes and colleagues demonstrated neural and behavioral effects in disparate paradigms. Third, we adopted a subject-specific ROI selection approach and included exploratory analyses of female faces. Thus, our pre-registered study provides a substantial extension of the existing literature by (i) conceptually replicating prior work (Hughes et al., 2019), thereby bolstering our confidence in the reliability of this effect, and (ii) providing further evidence for representation-based accounts of the other-race homogeneity effect (see also Yaros et al., 2019). Notably, neither of these findings speak to the developmental origins of the effect. Specifically, individuals continuously absorb information from their environment, and their social behaviors and representations undoubtedly update as a function of this input (Rule et al., 2013). Thus, our findings cannot indicate whether the differences we observed stem from innate processes or, alternatively, were acquired throughout participants' lifetimes.

Our findings have two potential limitations. First, we observed the effects only for male and not for female targets (see Fig. 3 and 3-5). One possibility is that women are less likely to be targets of the outgroup homogeneity effect. These findings accord with a broader literature

eNeuro Accepted Manuscript

436 documenting that outgroup men are more likely targets of intergroup discrimination and harm
437 than are women (e.g., Social Dominance Theory (Sidanius and Pratto, 2001) and the Theory of
438 Gendered Prejudice (Sidanius et al., 2019), both highlighting the importance of gender as a
439 moderator in intergroup relations). These findings could also indicate that representations in the
440 FFA reflect multiple categorical geometries, including race, gender, and their interaction
441 (Freeman et al., 2018). That said, we found that images of Black female targets in our task were
442 more different from each other (with regard to image properties) than the other conditions. Thus,
443 a third possibility is that increased variability made different Black female faces more
444 distinguishable. Rather than omit female faces from the paradigm like so many other studies of
445 this phenomenon, the current investigation included female targets to help start building a
446 knowledge base to adjudicate among these competing explanations. A second limitation of our
447 study is that it was constrained to White perceivers; nevertheless, given the robustness of the
448 outgroup homogeneity effect across groups and cultures (Wan et al., 2015; Kokje et al., 2018),
449 these results are likely to generalize to additional groups and targets (e.g., Asian perceivers
450 viewing White faces). Testing this hypothesis remains a goal for future studies.

451 Many theories of human sociality begin with the assumption that perceivers can keep
452 track of others' reputation by correctly identifying and later remembering what they did, and to
453 whom they did it. Nevertheless, humans routinely fail to engage in such basic social cognition
454 for outgroup members, in large part because perceivers do not consistently distinguish among
455 individual members of such outgroups (especially those delineated by race and ethnicity). This
456 outgroup homogeneity effect undermines one of the basic starting conditions of human
457 (pro)sociality, in that it forestalls the ability to identify individuals with a unique set of past
458 actions and behavioral tendencies. Here, we suggest that the potential origins of the outgroup

eNeuro Accepted Manuscript

459 homogeneity effect lie in failures of visual processing to form distinct representations of
460 individual members of outgroups—something that it nevertheless accomplishes exquisitely for
461 members of one's own social groups.

462

463 **References**

464 Bates DM, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using
465 Eigen and S4. Available at: http://cran.r-project.org/package=lme4.

466 Behrman BW, Davey SL (2001) Eyewitness identification in actual criminal cases: An archival
467 analysis. Law Hum Behav 25:475–491.

468 Boldry JG, Gaertner L, Quinn J (2007) Measuring the measures: A meta-analytic investigation of
469 the measures of outgroup homogeneity. Gr Process Intergr Relations 10:157–178.

470 Castello MV di O (2017) Dynamic Face/Body/Scene localizer. Available at:
471 https://github.com/mvdoc/pitcher_localizer.

472 Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences, 2nd ed. New Jersey:
473 Lawrence Erlbaum Associates, Inc.

474 Correll J, Hudson SM, Guillermo S, Earls HA (2017) Of kith and kin: Perceptual enrichment,
475 expectancy, and reciprocity in face perception. Personal Soc Psychol Rev 21:336–360.

476 Cusack R, Papadakis N (2002) New robust 3-D phase unwrapping algorithms: Application to
477 magnetic field mapping and undistorting echoplanar images. Neuroimage 16:754–764.

478 Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–
479 114.

eNeuro Accepted Manuscript

480 Duchaine B, Yovel G (2015) A revised neural framework for face processing. Annu Rev Vis Sci
481 1:393–416 Available at: http://www.annualreviews.org/doi/10.1146/annurev-vision-
482 082114-035518.

483 Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Glasser MF, Miller KL, Ugurbil
484 K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain FMRI
485 and fast diffusion imaging. PLoS One 5:e15710.

486 Feng L, Liu J, Wang Z, Li J, Li L, Ge L, Tian J, Lee K (2011) The other face of the other-race
487 effect: An fMRI investigation of the other-race face categorization advantage.
488 Neuropsychologia 49:3739–3749 Available at:
489 http://dx.doi.org/10.1016/j.neuropsychologia.2011.09.031.

490 Gilaie-Dotan S, Gelbard-Sagiv H, Malach R (2010) Perceptual shape sensitivity to upright and
491 inverted faces is reflected in neuronal adaptation. Neuroimage 50:383–395 Available at:
492 http://dx.doi.org/10.1016/j.neuroimage.2009.12.077.

493 Golby AJ, Gabrieli JDE, Chiao JY, Eberhardt JL (2001) Differential responses in the fusiform
494 region to same-race and other-race faces. Nat Neurosci 4:845–850.

495 Gotts SJ, Chow CC, Martin A (2012) Repetition priming and repetition suppression: A case for
496 enhanced efficiency through neural synchronization. Cogn Neurosci 3:227–237.

497 Greenwald AG, Mcghee DE, Schwartz JLK (1998) Measuring individual differences in implicit
498 cognition: The implicit association test. J Pers Soc Psychol 74:1464–1480.

499 Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential
500 processing of objects under various viewing conditions in the human lateral occipital

eNeuro Accepted Manuscript

complex. Neuron 24:187–203.

Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J (2008) Detection of time-varying signals in event-related fMRI designs. Neuroimage 43:509–520.

Ho AK, Sidanius J, Kteily N, Sheehy-Skeffington J, Pratto F, Henkel KE, Foels R, Stewart AL (2015) The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO_7 scale. J Pers Soc Psychol 109:1003–1028 Available at: http://doi.apa.org/getdoi.cfm?doi=10.1037/pspi0000033.

Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A (2019) Moving beyond P values: data analysis with estimation graphics. Nat Methods 16:565–566 Available at: http://dx.doi.org/10.1038/s41592-019-0470-3.

Hugenberg K, Wilson JP, See PE, Young SG (2013) Towards a synthetic model of own group biases in face memory. Vis cogn 21:1392–1417 Available at: http://dx.doi.org/10.1080/13506285.2013.821429.

Hughes BL, Camp NP, Gomez J, Natu VS, Grill-spector K, Eberhardt JL (2019) Neural adaptation to faces reveals racial outgroup homogeneity effects in early perception. Proc Natl Acad Sci 116:14532–14537.

Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. J Mem Lang 59:434–446.

Judd CM, Park B, Yzerbyt V, Gordijn EH, Muller D (2005) Attributions of intergroup bias and outgroup homogeneity to ingroup and outgroup others. Eur J Soc Psychol 35:677–704.

Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally

eNeuro Accepted Manuscript

522 defining regions of interest in the ventral visual pathway. Neuroimage 60:2357–2364
523 Available at: http://dx.doi.org/10.1016/j.neuroimage.2012.02.055.

524 Kim JS, Yoon HW, Kim BS, Jeun SS, Jung SL, Choe BY (2006) Racial distinction of the
525 unknown facial identity recognition mechanism by event-related fMRI. Neurosci Lett
526 397:279–284.

527 Kokje E, Bindemann M, Megreya AM (2018) Cross-race correlations in the abilities to match
528 unfamiliar faces. Acta Psychol (Amst) 185:13–21 Available at:
529 https://doi.org/10.1016/j.actpsy.2018.01.006.

530 Krosch AR, Amodio DM (2019) Scarcity disrupts the neural encoding of Black faces: A
531 aocioperceptual pathway to discrimination. J Pers Soc Psychol 117:859–875.

532 Lo S, Andrews S (2015) To transform or not to transform : Using Generalized Linear Mixed
533 Models to analyse reaction time data. Front Psychol 6:1–16.

534 Ma DS, Correll J, Wittenbrink B (2015) The Chicago face database: A free stimulus set of faces
535 and norming data. Behav Res Methods 47:1122–1135.

536 Malpass RS, Kravitz J (1969) Recognition for faces of own and other race. J Pers Soc Psychol
537 13:330–334.

538 Marcon JL, Meissner CA, Frueh M, Susa KJ, MacLin OH (2010) Perceptual identification and
539 the cross-race effect. Vis cogn 18:767–779.

540 Mathur VA, Harada T, Chiao JY (2012) Racial identification modulates default network activity
541 for same and other races. Hum Brain Mapp 33:1883–1893.

542 Meissner CA, Brigham JC (2001) Thirty years of investigating the own-race bias in memory for

eNeuro Accepted Manuscript

543 faces: A meta-analytic review. Psychol Public Policy, Law 7:3–35.

544 Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Uğurbil K (2010) Multiband
545 multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with
546 application to high spatial and temporal whole-brain fMRI. Magn Reson Med 63:1144–
547 1153.

548 Molenberghs P, Louis WR (2018) Insights From fMRI Studies Into Ingroup Bias. Front Psychol
549 9:1–12 Available at: https://www.frontiersin.org/article/10.3389/fpsyg.2018.01868/full.

550 Mumford JA, Davis T, Poldrack RA (2014) The impact of study design on pattern estimation for
551 single-trial multivariate pattern analysis. Neuroimage 103:130–138 Available at:
552 http://www.ncbi.nlm.nih.gov/pubmed/25241907.

553 Nieto-Castañón A, Fedorenko E (2012) Subject-specific functional localizers increase sensitivity
554 and functional resolution of multi-subject analyses. Neuroimage 63:1646–1669.

555 Nosek BA, Smyth FL, Hansen JJ, Devos T, Lindner NM, Ranganath KA, Smith CT, Olson KR,
556 Chugh D, Greenwald AG, Banaji MR (2007) Pervasiveness and correlates of implicit
557 attitudes and stereotypes. Eur Rev Soc Psychol 18:36–88.

558 Papesh MH, Goldinger SD (2010) A multidimensional scaling analysis of own- and cross-race
559 face spaces. Cognition 116:283–288 Available at:
560 http://dx.doi.org/10.1016/j.cognition.2010.05.001.

561 Park B, Rothbart M (1982) Perception of out-group homogeneity and levels of social
562 categorization: Memory for the subordinate attributes of in-group and out-group members. J
563 Pers Soc Psychol 42:1051–1068.

eNeuro Accepted Manuscript

564 Patil I (2018) ggstatsplot: "ggplot2" Based Plots with Statistical Details. Available at:
565 https://cran.r-project.org/package=ggstatsplot.

566 Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011) Differential selectivity for
567 dynamic versus static information in face-selective cortical regions. Neuroimage 56:2356–
568 2363 Available at: http://dx.doi.org/10.1016/j.neuroimage.2011.03.067.

569 Plant EA, Devine PG (1998) Internal and external motivation to respond without prejudice. J
570 Pers Soc Psychol 75:811–832 Available at:
571 http://journals.sagepub.com/doi/10.1177/0146167205275304.

572 Pourtois G, Schwartz S, Seghier ML, Lazeyras F, Vuilleumier P (2005) View-independent
573 coding of face identity in frontal and temporal cortices is modulated by familiarity: an
574 event-related fMRI study. Neuroimage 24:1214–1224 Available at:
575 http://www.ncbi.nlm.nih.gov/pubmed/15670699 [Accessed March 31, 2012].

576 Robinson A (2016) Package "equivalence." Available at: https://cran.r-
577 project.org/package=equivalence.

578 Rule NO, Freeman JB, Ambady N (2013) Culture in social neuroscience: A review. Soc
579 Neurosci 8:3–10.

580 Sidanius J, Hudson S-KTJ, Davis G, Bergh R (2019) The theory of gendered prejudice: A social
581 dominance and intersectionalist perspective. In: The Oxford Handbook of Behavioral
582 Political Science (Mintz A, Terris L, eds). Oxford University Press.

583 Sidanius J, Pratto F (2001) Social dominance: An intergroup theory of social hierarchy and
584 oppression. Cambridge University Press.

eNeuro Accepted Manuscript

585 Singmann H, Bolker B, Westfall J, Aust F (2018) afex: Analysis of factorial experiments.
586 Available at: https://cran.r-project.org/package=afex.

587 Van Bavel JJ, Packer DJ, Cunningham WA (2008) The neural substrates of in-group bias.
588 Psychol Sci 19:1131–1139 Available at: http://journals.sagepub.com/doi/10.1111/j.1467-
589 9280.2008.02214.x.

590 Van Bavel JJ, Packer DJ, Cunningham WA (2011) Modulation of the Fusiform face area
591 following minimal exposure to motivationally relevant faces: Evidence of in-group
592 enhancement (not out-group disregard). J Cogn Neurosci 23:3343–3354.

593 Vizioli L, Rousselet GA, Caldara R (2010) Neural repetition suppression to identity is abolished
594 by other-race faces. Proc Natl Acad Sci 107:20081–20086 Available at:
595 http://www.pnas.org/cgi/doi/10.1073/pnas.1005751107.

596 Wagner JK, Yu JH, Ifekwunigwe JO, Harrell TM, Bamshad MJ, Royal CD (2017)
597 Anthropologists' views on race, ancestry, and genetics. Am J Phys Anthropol 162:318–327.

598 Wan L, Crookes K, Reynolds KJ, Irons JL, Mckone E (2015) A cultural setting where the other-
599 race effect on face recognition has no social – motivational component and derives entirely
600 from lifetime perceptual experience. Cognition 144:91–115 Available at:
601 http://dx.doi.org/10.1016/j.cognition.2015.07.011.

602 Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: From error
603 visibility to structural similarity. IEEE Trans Image Process 13:600–612.

604 Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-
605 level image properties: The SHINE toolbox. Behav Res Methods 42:671–684.

eNeuro Accepted Manuscript

606 Xu J, Moeller S, Auerbach EJ, Strupp J, Smith SM, Feinberg DA, Yacoub E, Uğurbil K (2013)
607 Evaluation of slice accelerations using multiband echo planar imaging at 3 T. Neuroimage
608 83:991–1001.

609 Yaros JL, Salama DA, Delisle D, Larson MS, Miranda BA, Yassa MA (2019) A memory
610 computational basis for the other-race effect. Sci Rep 9:19399 Available at:
611 http://dx.doi.org/10.1038/s41598-019-55350-0.

612

613

eNeuro Accepted Manuscript

FIGURE LEGENDS

Figure 1. Task design. In the fMRI scanner, participants viewed face pairs from different races (Black, White) and genders (women, men) and decided if the two faces belonged to the same individual. Participants saw each face for 600ms and responded when the second face was on screen. Faces were presented with a random horizontal jitter to prevent low-level suppression effects. The design included three face pair conditions for each combination of race and gender. Participants saw (a) a face followed by a scrambled visual patch in the shape of a face in the single condition (i.e., baseline trials; we included this condition only in Experiment 2); (b) a face followed by the same face in the repeated-face condition; and (c) a face followed by a different face (matched in gender and race) in the different-face condition. We report the behavioral results in Figures 2 (male targets) and 1-1 (female targets).

Figure 2. Reaction time and memory performance for male faces in Experiments 1 and 2. All panels depict estimates computed by generalized linear mixed models, as detailed in the Methods section. (a) In the face repetition task in Experiment 1, participants responded faster to repeated faces ($t = 11.33$, $p < 0.001$). This effect was qualified by an interaction with Race, such that participants responded faster to different White individuals compared to different Black individuals (see main text for inferential statistics). (b) In a 2-alternative forced-choice test, participants in Experiment 1 indicated their memory by selecting the individual they thought was presented in the face repetition task. All trials included one previously presented individual ('old') and one never-before-seen individual ('new'). Participants remembered old White targets better than old Black targets (odds ratio [OR] = 1.16, *Wald's z* = 3.005, $p = 0.003$). We did not observe a significant effect of type of repetition on memory (OR = 1.07, *Wald's z* = 1.48, $p = 0.14$) or of the interaction of repetition with race (OR = 1.05, *Wald's z* = 1.09, $p = 0.28$). (c) Experiment 2 included an additional condition—single faces—to facilitate comparison of repetition suppression to baseline neural activity for each race. Race did not affect reaction time in the single-face condition ($t = 1.04$, $p = 0.300$). When analyzing only face-pair trials (omitting the single face trials), participants responded faster to repeated faces ($t = 2.46$, $p = 0.014$), an effect that was qualified by an interaction, replicating the results of Experiment 1 (interaction model comparison: $\chi_{(1)} = 15.095$, $p < 0.001$): participants were slower to respond to different Black faces compared to different White faces ($t = 2.48$, $p = 0.013$). Unexpectedly, participants in Experiment 2 were also slower to respond to repeated White faces compared to repeated Black faces ($t = 3.29$, $p < 0.001$). Note that unlike Experiment 1, the correct response key in Experiment 2 was imbalanced between the conditions; we assigned the same key to single and different conditions to simplify the task, hence assigning the same correct response key to two thirds of the trials. This design choice probably slowed the responses to repeated trials, as participants had to use an infrequent key to respond correctly to these trials. This, in turn, might have made responses to repeated targets more difficult—a difficulty that manifested particularly strongly for White targets. Error bars indicate standard error of the mean.

Figure 3. Results of the dynamic face localizer task and repetition suppression parameter estimates extracted from the FFA. (a) The spatial extent of the key ROIs of interest, the FFA and the OFA, presented as degree of spatial overlap between participants in normalized space. We

eNeuro Accepted Manuscript

generated these ROIs with the Group-Constrained-Subject-Specific (GcSS) approach (Julian et al., 2012) for the contrast of Faces > Other Categories at the level of each individual participant. See Figure 3-6 for a table listing the full ROI list. (b) Repetition suppression parameter estimates in the FFA for different-face and repeated-face pairs for Black and White male faces. Upper panel: To present repetition suppression effects, we subtracted neural activity in response to single faces (baseline) from the neural response to the different conditions (Differ, Repeat), separately for each race (Black, White). Negative values indicate neural suppression compared to baseline. Across all figures, individual dots represent neural suppression for unique participants. Each figure also visualizes the mean of each condition (as a red dot), the median (solid horizontal line) and 1st and 3rd quartiles (boxplot). Lower panel: mean effect size (the difference in suppression effect) and the bootstrapped 95% confidence intervals for the comparison between Black and White targets in each condition. The results demonstrate more release from suppression (i.e., less negative values) for different White targets compared to different Black targets, but no difference in suppression for repeating Black and White targets. This pattern suggests that the FFA forms unique representations for different individuals, but only for members of one's own racial group. We report the results of the parallel analysis in the OFA and visual cortex in Figure 4. Figures 3-1, 3-2, 3-3, and 3-4 demonstrate the robustness of the effect to different analytic strategies. In Figure 3-5 we report the results of the main analysis with female faces. ** $p < 0.01$

Figure 4. Repetition suppression results for repeated and different Black and White targets for the OFA and primary visual cortex (see Materials and Methods for additional details on analysis approach). To plot suppression effects (upper panel), we subtracted neural activity in response to single faces (baseline) from the neural response to the other experimental conditions (Different, Repeated), separately for each race (Black, White). Negative values indicate neural suppression compared to baseline. The lower panel depicts effect size estimate (the difference in suppression effect) and the bootstrapped 95% confidence intervals for the comparison between Black and White targets in each condition. (a) The OFA demonstrated robust repetition suppression for repeated and different faces ($F_{(1,28)} = 245.88$, $MSE = 12.06$, $p < 0.001$, $\eta^2_p = 0.90$ and $F_{(1,28)} = 247.03$, $MSE = 10.13$, $p < 0.001$, $\eta^2_p = 0.90$ for repeated and different targets, respectively). The OFA showed some release from suppression for different faces ($F_{(1,28)} = 9.93$, $MSE = 1.98$, $p = 0.004$, $\eta^2_p = 0.26$), but we found no indication for an effect of race on suppression for repeated or different targets ($F_{(1,28)} = 0.04$, $MSE = 3.88$, $p = 0.85$, $\eta^2_p = 0.001$ and $F_{(1,28)} = 0.46$, $MSE = 3.62$, $p = 0.50$, $\eta^2_p = 0.02$, respectively). Thus, the OFA was not sensitive to the group features of the different faces. (b) Primary visual cortex (including V1) served as a control condition. No effects were observed in this region (all F's <1.4, all p's > 0.2).

Figure 1-1. Reaction time and memory performance for female faces in Experiments 1 and 2. All panels depict estimates computed by generalized linear mixed models, as detailed in the Materials and Methods section. (a) Participants in Experiment 1 were faster to respond to repeated faces than to different faces ($t = 8.43$, $p < 0.001$). However, unlike responses for male faces, there was no evidence for an effect of Race ($t = 0.34$, $p = 0.74$) nor for an interaction of Race with type of Repetition (interaction model comparison: $\chi_{(1)} = 1.93$, $p = 0.16$, suggesting that a model with no interaction term provides a better fit to the data). This pattern hints that Black

eNeuro Accepted Manuscript

female faces were individuated to the same extent as White female faces. When we included all trials to test a full model of the data, with the fixed effects of Race, Repetition, and Gender, we observed a significant three-way interaction (model comparison: $\chi_{(4)} = 10.67$, $p = 0.031$), confirming the difference in individuation between male and female faces. (b) Unlike male faces, participants were *not* more likely to remember White female faces compared to Black female faces (OR = 1.004, *Wald's z* = 0.09, $p = 0.93$). Participants were more likely to remember repeated faces compared to different faces (OR = 1.10, *Wald's z* = 2.20, $p = 0.028$) with no interaction with Race (OR = 1.02, *Wald's z* = 0.456, $p = 0.65$). However, we did not observe a statistically significant three-way interaction effect for gender when it was inserted into the model ($\chi_{(4)} = 6.88$, $p = 0.14$). (c) Similar to responses to male faces, participants in Experiment 2 responded faster to single face trials (model comparison: $\chi_{(2)} = 111.42$, $p < 0.001$). However, unlike response times for male faces, we did not observe a robust effect of Race ($t = 1.97$, $p = 0.049$) nor an interaction of Race and Repetition (model comparison: $\chi_{(2)} = 1.03$, $p = 0.597$; we obtained similar results when omitting the single condition from the analysis). As in Experiment 1, when we included the fixed effects of Race, Repetition, and Gender in the model, we observed a significant triple interaction (model comparison: $\chi_{(4)} = 27.84$, $p < 0.001$), once again confirming the difference in processing between male and female faces. Error bars indicate standard error of the mean.

Figure 3-1. Repetition suppression parameter estimates for the FFA, OFA and primary visual cortex for analyses including participants we excluded in the main text (N=32). Upper panel: repetition effects for all individuals. Lower panel: effect size estimate and the bootstrapped 95% confidence intervals for the comparison between Black and White targets in each condition. Overall, results replicate with this sample (compare to Figure 3b and Figure 4). (a) FFA results. Activation in FFA was suppressed to nearly an equivalent degree for both Black and White repeated targets [repetition effect: $F_{(1,31)} = 167.86$, $MSE = 5.32$, $p < 0.001$, $\eta^2_p = 0.84$; interaction effect: $F_{(1,31} = 1.50$, $MSE = 0.98$, $p = 0.23$, $\eta^2_p = 0.05$; equivalence test for the two repetition effects (mean difference: 0.43): p = 0.056, equivalence interval (-1.02, 0.16)]. FFA also showed more release from suppression for White targets than for Black targets [$F_{(1,31)} = 11.96$, $MSE = 0.56$, $p = 0.002$, $\eta^2_p = 0.31$; interaction between repeated and different faces: $F_{(1,31)} = 3.35$, $MSE = 0.57$, $p = 0.08$, $\eta^2_p = 0.10$]. And again, similar to the main findings, we observed statistically equivalent levels of repetition suppression for different and repeated Black targets [equivalence test (mean difference: 0.46): $p = 0.005$, (-0.80, -0.12)]. (b) The OFA demonstrated robust repetition suppression for repeated and different faces ($F_{(1,31)} = 273.17$, $MSE = 11.83$, $p < 0.001$, $\eta^2_p = 0.90$ and $F_{(1,31)} = 261.09$, $MSE = 10.37$, $p < 0.001$, $\eta^2_p = 0.89$ for repeated and different trials, respectively). The OFA showed some release from suppression for different faces ($F_{(1,31)} = 12.43$, $MSE = 1.87$, $p = 0.001$, $\eta^2_p = 0.29$), but no indication for an effect of race on any of these results ($F_{(1,31)} = 0.05$, $MSE = 3.59$, $p = 0.83$, $\eta^2_p = 0.002$ and $F_{(1,31)} = 0.28$, $MSE = 3.39$, $p = 0.50$, $\eta^2_p = 0.009$ for repeated and different targets, respectively). Thus, the OFA was not sensitive to the group features of the different faces. (c) We did not observe any effects in the primary visual cortex, again replicating the principal analysis (all *F's* <2.02, all *p's* > 0.17).

eNeuro Accepted Manuscript

Figure 3-2. Repetition suppression parameter estimates for the FFA, OFA and primary visual cortex for analyses excluding targets that participants incorrectly classified (pre-registered complementary analysis; N=29). Upper panel: repetition effects for all individuals. Lower panel: effect size estimate and the bootstrapped 95% confidence intervals for the comparison between Black and White targets in each condition. As with the previous analysis, results replicate the main findings. (a) FFA results. Activation in FFA was suppressed to nearly an equivalent degree for both Black and White repeated faces [repetition effect: $F_{(1,28)} = 180.11$, $MSE = 4.56$, $p < 0.001$, $\eta^2_p = 0.87$; interaction effect: $F_{(1,28} = 0.28$, $MSE = 1.60$, $p = 0.60$, $\eta^2_p = 0.01$; equivalence test for the two repetition effects (mean difference: 0.25): p = 0.0598, equivalence interval (-1.05, 0.55)]. FFA also showed more release from suppression for White targets than for Black targets [$F_{(1,28)} = 6.94$, $MSE = 0.60$, $p = 0.01$, $\eta^2_p = 0.20$; interaction between repeated and different faces: $F_{(1,28)} = 2.01$, $MSE = 0.94$, $p = 0.17$, $\eta^2_p = 0.07$]. And again, similar to the main findings, we observed statistically equivalent levels of repetition suppression for different and repeated Black targets [equivalence test (mean difference: 0.47): $p = 0.036$, (-0.95, 0.02)]. (b) The OFA demonstrated robust repetition suppression for repeated and different faces ($F_{(1,28)} = 233.52$, $MSE = 12.73$, $p < 0.001$, $\eta^2_p = 0.89$ and $F_{(1,28)} = 267.83$, $MSE = 9.52$, $p < 0.001$, $\eta^2_p = 0.91$ for repeated and different trials, respectively). The OFA showed some release from suppression for different faces ($F_{(1,28)} = 4.70$, $MSE = 3.46$, $p = 0.04$, $\eta^2_p = 0.14$). As before, race did not interact with suppression for repeated or different faces ($F_{(1,28)} = 0.99$, $MSE = 7.52$, $p = 0.33$, $\eta^2_p = 0.03$ and $F_{(1,28)} = 0.08$, $MSE = 4.51$, $p = 0.78$, $\eta^2_p = 0.003$ for repeated and different targets, respectively). Thus, once again, the OFA was not sensitive to the group features of the different faces. (c) We did not observe any effects in the primary visual cortex, again replicating the principal analysis (all F's <1.13, all p's > 0.30).

Figure 3-3. Parameter estimates for the FFA, OFA and primary visual cortex for the original pre-registered analysis, in which we erroneously modelled single target trials with variable durations based on participants' reaction time to the scrambled image that followed the single face (N=29). Upper panel: repetition effects for all individuals. Lower panel: effect size estimate and the bootstrapped 95% confidence intervals for the comparison between Black and White targets in each condition. As in all previous analyses, we plot neural activity by subtracting the response to single faces (baseline) from the neural response to the different conditions (Different, Repeated), separately for each race (Black, White). Parameter estimates in this analysis are positive and no longer reflect suppression, as the baseline in this analysis reflects the neural activity in response to the scrambled image - rather than an actual face. This response was naturally weaker in face-sensitive regions. Nonetheless, overall pattern of results replicated the principal analysis. (a) FFA results. Activation was different from baseline to an equivalent degree for Black and White repeated faces [repetition effect: $F_{(1,28)} = 51.79$, $MSE = 0.35$, $p < 0.001$, $\eta^2_p = 0.65$; interaction effect: $F_{(1,28} = 0.15$, $MSE = 0.50$, $p = 0.70$, $\eta^2_p = 0.005$; equivalence test for the two repetition effects (mean difference: 0.10): p = 0.001, equivalence interval (-0.55, 0.34)]. FFA also showed more difference from baseline for White targets than for Black targets [$F_{(1,28)} = 6.74$, $MSE = 0.31$, $p = 0.01$, $\eta^2_p = 0.19$; interaction between repeated and different faces: $F_{(1,28)} = 3.65$, $MSE = 0.38$, $p = 0.07$, $\eta^2_p = 0.12$]. And again, similar to the main findings, we observed statistically equivalent levels of difference from baseline for different and repeated Black targets [equivalence test (mean difference: 0.28): $p < 0.001$, (-0.55, -0.003)]. (b) The OFA

eNeuro Accepted Manuscript

demonstrated difference from baseline only for different and not for repeated faces ($F_{(1,28)} = 2.48$, $MSE = 1.48$, $p = 0.13$, $\eta^2_p = 0.08$ and $F_{(1,28)} = 28.66$, $MSE = 1.26$, $p < 0.001$, $\eta^2_p = 0.51$ for repeated and different trials, respectively). The difference between OFA response to different and repeated faces was significant ($F_{(1,28)} = 8.67$, $MSE = 1.94$, $p = 0.006$, $\eta^2_p = 0.24$). As before, race did not interact with difference from baseline for repeated or different faces ($F_{(1,28)} = 0.55$, $MSE = 1.42$, $p = 0.47$, $\eta^2_p = 0.02$ and $F_{(1,28)} = 0.99$, $MSE = 1.26$, $p = 0.33$, $\eta^2_p = 0.03$, respectively). Thus, the OFA was not sensitive to the group features of the different faces. (c) We did not observe any effects of race or interaction with race in the primary visual cortex, again replicating the principal analysis (all *F's* <2.13, all *p's* > 0.16).

Figure 3-4. Parameter estimates in the (a) FFA and (b) OFA for the three experimental conditions for Black and White male faces (without correcting for differences in baseline activity). Upper panels: Neural activity in response to the different conditions (Single, Different, Repeated), separately for each race (Black, White). Individual dots represent neural activity for unique participants. Vertical lines (presented in parallel to the scatter plot) depict the mean and standard deviation for each condition. Lower panel: mean effect size (the repetition suppression effect) and the bootstrapped 95% confidence intervals for the comparison between type of repetition and single-face targets for each race. See Figure 3 and main text for statistical analyses.

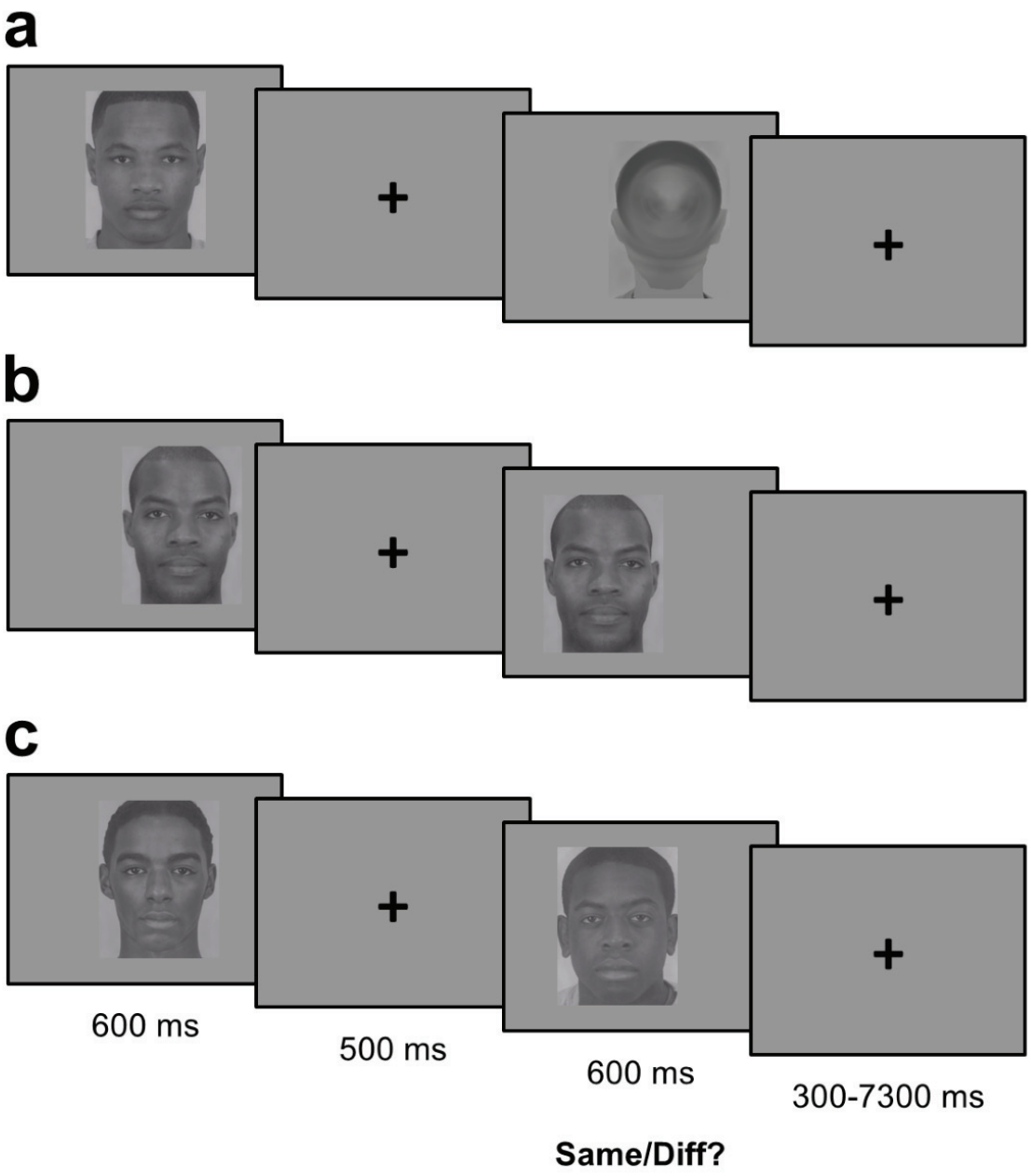
Figure 3-5. Repetition suppression results for female targets for the FFA, OFA and primary visual cortex using the analysis reported in the main manuscript (compare to results for male targets in Figure 3b and Figure 4). We subtracted neural activity in response to single faces (baseline) from the neural response to the different conditions (Different, Repeated), separately for each race (Black, White) (upper panel). The lower panel depicts effect size estimate (the difference in suppression effect) and the bootstrapped 95% confidence intervals for the comparison between Black and White targets in each condition. (a) The FFA demonstrated robust repetition suppression for repeated and different faces ($F_{(1,28)} = 392.66$, $MSE = 2.48$, $p < 0.001$, $\eta^2_p = 0.93$ and $F_{(1,28)} = 290.91$, $MSE = 2.66$, $p < 0.001$, $\eta^2_p = 0.91$ for repeated and different trials, respectively). The FFA also demonstrated release from suppression for different faces ($F_{(1,28)} = 14.66$, $MSE = 0.78$, $p < 0.001$, $\eta^2_p = 0.34$), an effect that was qualified by an interaction with race ($F_{(1,28)} = 6.76$, $MSE = 0.38$, $p = 0.01$, $\eta^2_p = 0.19$). However, the FFA did not show significant simple effects (i.e., differences in suppression) between Black and White female targets ($F_{(1,28)} = 2.41$, $MSE = 1.58$, $p = 0.13$, $\eta^2_p = 0.08$ and $F_{(1,28)} = 0.08$, $MSE = 1.51$, $p = 0.77$, $\eta^2_p = 0.003$ for repeated and different faces, respectively). As we did not have an a priori hypothesis for female targets. we cannot offer a reliable interpretation of these results. (b) Much like for male targets, the OFA demonstrated robust repetition suppression for repeated and different faces ($F_{(1,28)} = 517.23$, $MSE = 6.92$, $p < 0.001$, $\eta^2_p = 0.95$ and $F_{(1,28)} = 371.93$, $MSE = 8.28$, $p < 0.001$, $\eta^2_p = 0.93$ for repeated and different trials, respectively). The OFA showed some release from suppression for different faces ($F_{(1,28)} = 6.70$, $MSE = 2.79$, $p = 0.02$, $\eta^2_p = 0.19$), but no indication of an effect of race on any of these results ($F_{(1,28)} = 3.25$, $MSE = 4.84$, $p = 0.08$, $\eta^2_p = 0.10$ and $F_{(1,28)} = 2.08$, $MSE = 4.71$, $p = 0.16$, $\eta^2_p = 0.07$ for repeated and different targets). Thus, as for male targets, the OFA was not sensitive to the group features of the different female faces. (c) Primary visual cortex (including V1) results. No effects were observed in this region

(all *F's* <1.2, all *p's* > 0.2). Overall, we did not observe any neural evidence for the outgroup homogeneity effect for female targets, paralleling the behavioral results.

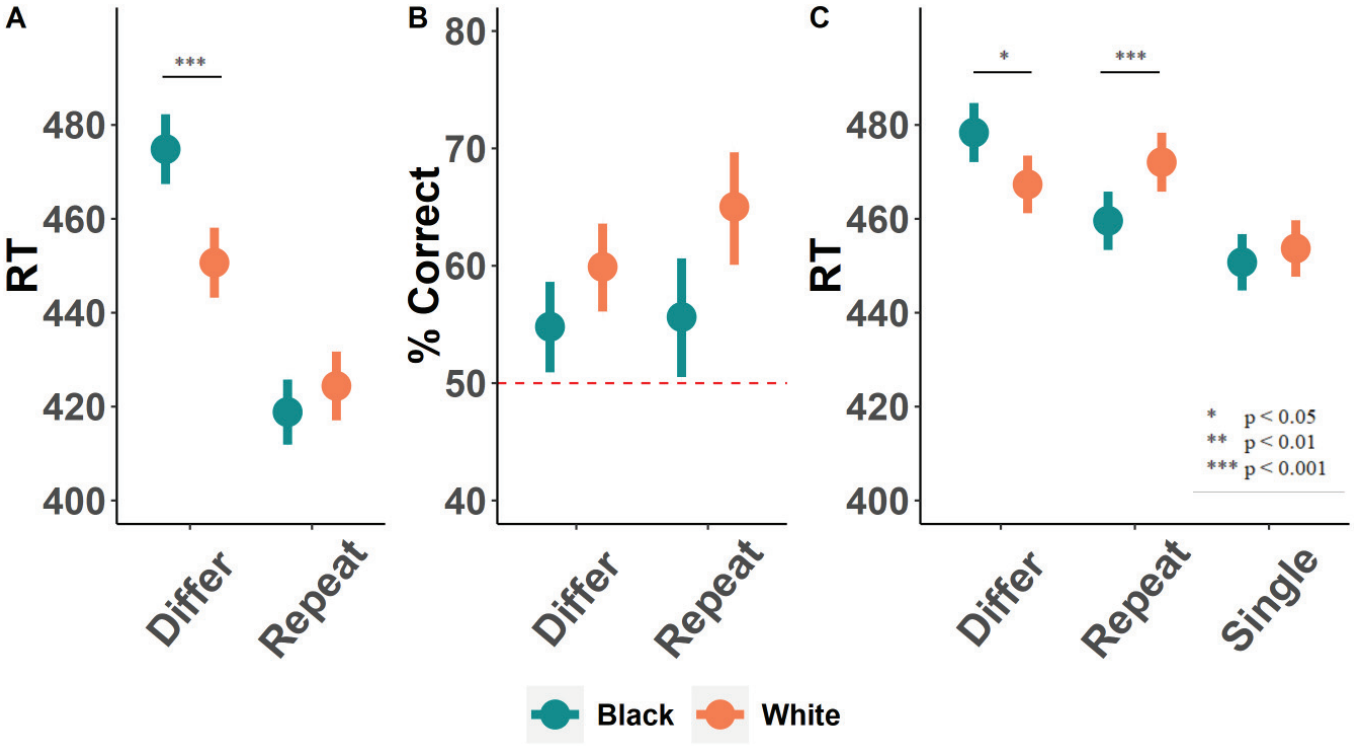
Figure 3-6. Results from the main functional ROI analysis (Group-Constrained Subject-Specific; GcSS; see Materials and Methods). For each ROI, the GcSS algorithm defines a parcel to intersect with activation for each individual participant, thus reliably identifying the same functional ROIs in all participants. We conducted the GcSS analyses with the contrast of Faces > Other Categories in independent localizer scans. A priori ROIs are presented in italics.

eNeuro Accepted Manuscript

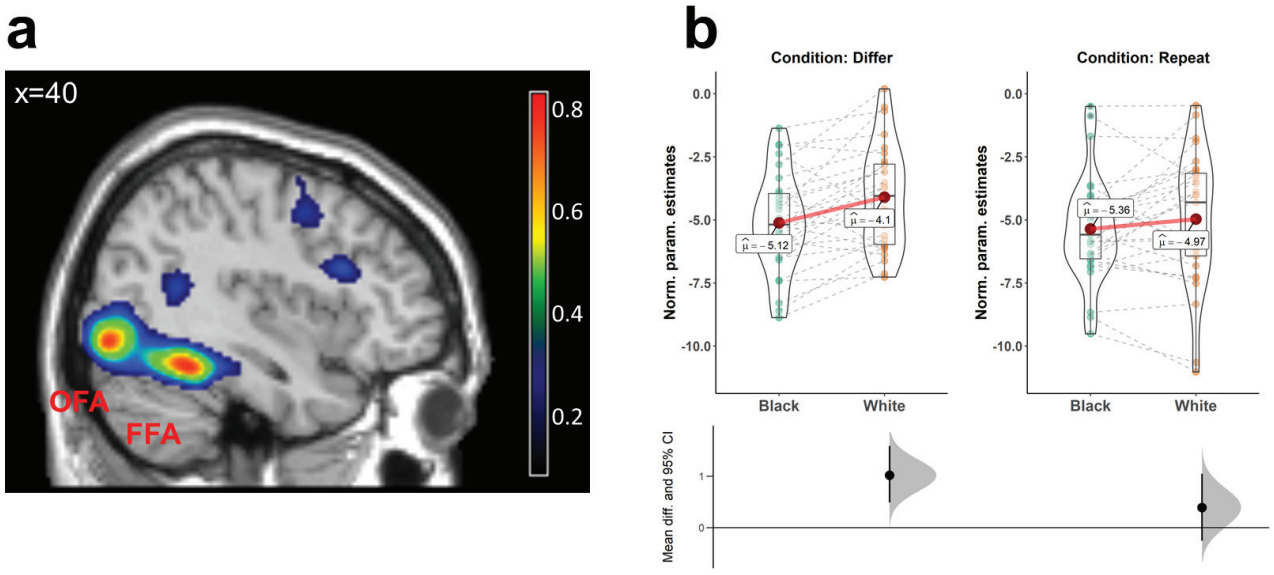
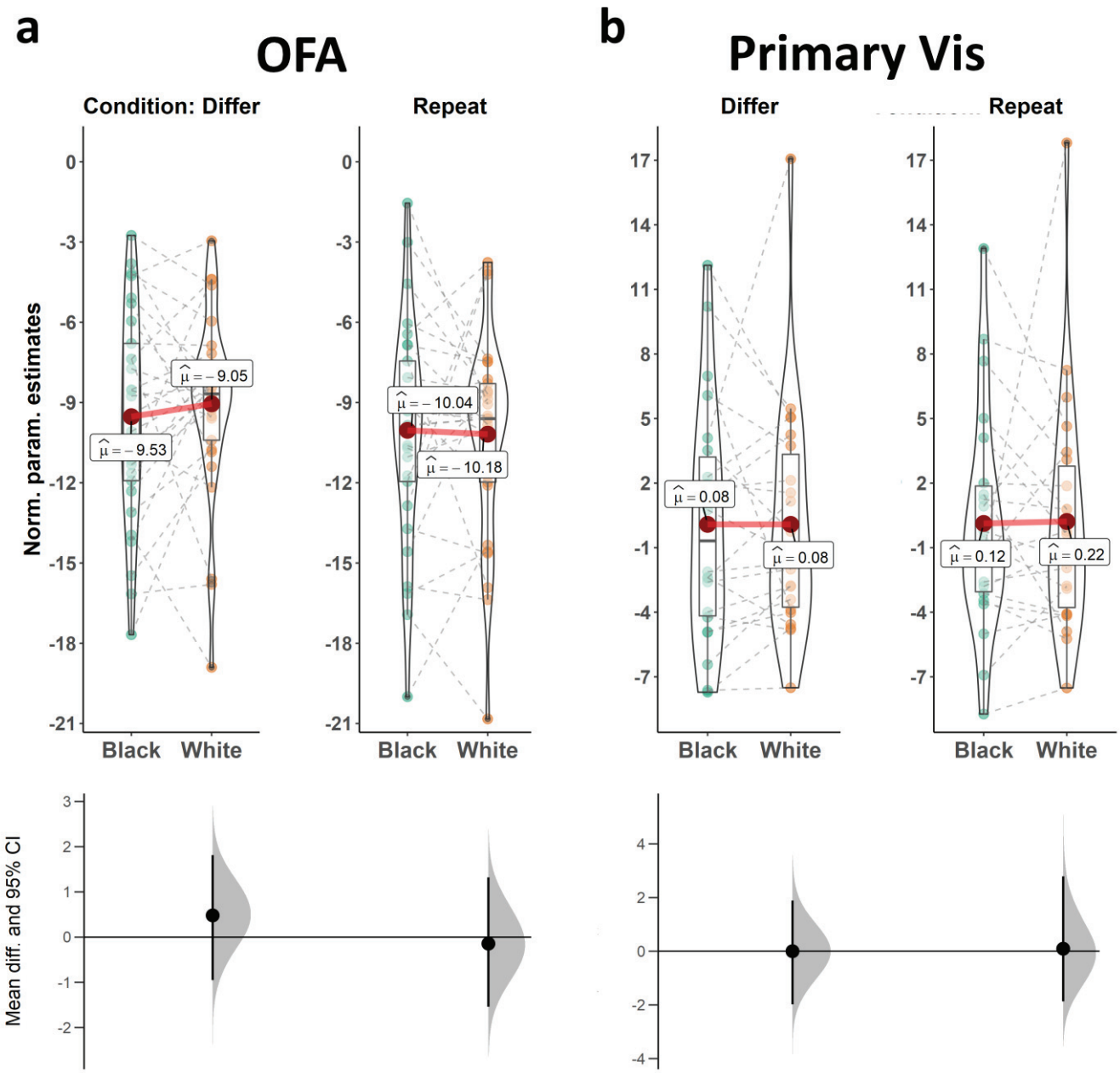
eNeuro Accepted Manuscript



eNeuro Accepted Manuscript



eNeuro Accepted Manuscript



eNeuro Accepted Manuscript