Real-life outgroup exposure, self-reported outgroup contact, and the other-race effect

Marleen Stelter[1,3], Deja Simon[2], Jimmy Calanchini[2], Oliver Christ[1], & Juliane Degner[3]

[1] Institute of Psychology, FernUniversität Hagen, Hagen, Germany

[2] Department of Psychology, University of California, Riverside, USA

[3] Institute of Psychology, Universität Hamburg, Hamburg, Germany

Author Note

Data Availability Statement: Data is publicly available online (https://osf.io/zr69c/).

Correspondence concerning this article should be addressed to Marleen Stelter, Universitaetsstraße 33, 58097 Hagen, Germany. E-mail: marleen.stelter@fernuni-hagen.de

Abstract

People are better at recognizing faces from their own racial or ethic group compared to faces from other racial or ethnic groups, known as the other-'race' effect (ORE). Several theories of the ORE assume that memory for other-race faces is impaired because people have less contact with members of other racial or ethnic groups, resulting in lower visual expertise. The present research investigates contact theories of the ORE, using self-report contact measures and objective measures of potential outgroup exposure (estimated from participants' residential location and from GPS tracking). Across six studies (total N = 2660), we observed that White American and White German participants displayed better memory for White faces compared to Black or Middle Eastern faces, whereas Black American participants displayed similarly equal or better memory for White compared to Black faces. We did not observe any relations between the ORE and objective measures of potential outgroup exposure. Only in Studies 2a and 2b, we observed very small correlations (rs = -.08 - .06) between 4 out of 30 contact measures and the ORE. We discuss methodological limitations as well as implications for theories of the ORE.

Keywords: other-race effect; own-group bias; cross-race recognition deficit; outgroup contact; GPS tracking; ethnic neighborhood composition

Word count: 8232

Real-life outgroup exposure, self-reported outgroup contact, and the other-race effect

Introduction

People are better at recognizing faces from their own race or ethnicity compared to faces from a different race or ethnicity (Meissner & Brigham, 2001). This other-'race'[1]-effect (ORE) is one of the most replicated and well-investigated effects in the face perception literature. However, in spite of a rich literature of empirical investigations and theorizing on the ORE (see Young, Hugenberg, Bernstein, & Sacco, 2012, for a review), the mechanisms underlying the ORE are still debated.

Theories of the Other-Race Effect

Several theories of the ORE are based on the idea that impaired recognition of other-race faces (at least partly) reflects less visual expertise for other-race compared to own-race faces. People often have less contact with members of other racial groups than with members of their own group, for example, due to geographic segregation or self-selection (e.g., Reardon et al., 2008). Such contact theories of the ORE broadly assume that less contact with members of other racial groups impairs the ability to encode perceptual details that make a face unique and are necessary to recognize a face. Though most theories of the ORE agree that contact is an important factor in causing the ORE, they differ in the assumed mechanisms of how contact differences impair recognition of other-race faces.

Face Space models, for example, assume that our visual systems are trained to recognize the faces we regularly come in contact with. Because we typically come in contact with more members of our racial ingroup than outgroups, our visual systems

[1] We use the term 'race' to refer to social categories and racialized group membership, even if perceivers may rely on phenotypic similarities among faces of different geographical regions of the world to construct these social categories (e.g., *Jena Declaration*, 2019).

become more finely tuned to encode and mentally represent the variability of visual details of own-race faces (Valentine, 1991). Consequently, mental representations of other-race faces are assumed to be less fine-grained than mental representations of own-race faces. Face Space models, thus, assume that suboptimal mental representations of other-race faces cause more errors in recognition of other-race faces.

Alternatively, the holistic processing account of the ORE suggests that higher levels of visual expertise for own-race faces translates into higher degrees of holistic processing (Tanaka, Kiefer, & Bukach, 2004). Holistic processing means that a face is integrated into a unified representation of the overall configuration and individual features. Holistic processing is generally associated with better memory performance (DeGutis, Wilmer, Mercado, & Cohan, 2013). Consequently, differences in holistic processing of own- and other-race faces may explain differences in memory for own- and other-race faces.

Developmental theories add a further perspective on the ORE, focusing on the importance of the point in time when outgroup contact occurs (e.g., Kelly et al., 2007; McKone et al., 2019). In line with theories of perceptual narrowing, developmental theories of the ORE suggest that there is a sensitive period during childhood for learning faces. When infants have little or no contact with people of other racial groups during this period, they are assumed to lose the ability to discriminate between other-race faces.

Finally, the categorization-individuation model (Hugenberg, Wilson, See, & Young, 2013; Hugenberg, Young, Bernstein, & Sacco, 2010) suggests that visual expertise developed through contact over time is a precondition for individuated face processing. Individuated processing focuses on facial features that distinguish one individual from other individuals and is usually applied to own-race faces. However, in the absence of either expertise or motivation, faces are processed in a more shallow category-based manner, with attention focused on facial features shared with other members of the same racial category. In the absence of expertise or motivation to individuate other-race faces, other-race faces are assumed to be processed on a categorical level, which in turn impairs memory for other-race faces.

In summary, many theories of the ORE agree that the extent of outgroup contact moderates the ORE, though they differ in the specific mechanisms by which contact affects memory. However, this theoretical assumption stands in sharp contrast to a large body of empirical evidence regarding the relation between outgroup contact and the ORE. In a meta-analysis from 2001, Meissner and Brigham reported a small correlation of r = -.13 between contact measures and the ORE, averaged across k = 29 studies. This finding was reproduced by a recent meta-analysis from Singh et al. (2021), reporting an equally small correlation of r = -.15 between contact measures and the ORE, averaged across k = 65 studies. In other words, meta-analyses converge on the conclusion that contact measures explain only about 2% of the variance in the magnitude of the ORE. Thus, previous research provides surprisingly little empirical evidence on the relation between contact measures and the ORE. This gap between theorizing and empirical findings emphasizes the need for further research.

Measuring Outgroup Contact

One limitation of previous research on the relation between outgroup contact and the ORE might be its frequent reliance on self-report contact measures (2021). Self-report measures can be biased and may not represent real-life behavior (Baumeister, Vohs, & Funder, 2007), and this is especially true in the context of behaviors that are frequent and seemingly mundane. For example, people may not attend carefully to the number of people they interact with in a given day, and thus may misremember the amount of contact they have with people from their own and other racial or ethnic groups. To address the limitations of self-reported contact, the present research indirectly estimates participants' extent of outgroup contact based on their location. For example, people who live in racially segregated places should be expected to have less daily exposure to and interaction with members of other racial groups than people who live in more diverse places. Thus, we predict that participants' extent of outgroup contact will – at least partly – depend on features of their location, such that people who live in places that afford greater outgroup contact will have smaller OREs. In order to increase the generalizability of the present research, we operationalized participants' location in several different ways. In the first set of studies, we operationalized online participants' locations based on their IP addresses to investigate the effect of the broader geographical context on the ORE. In a second set of studies, participants reported the geographic location (e.g., city districts) in which they live. In a third study, we used GPS tracking to measure participants movements during their daily lives over the course of two weeks. This measurement provides relatively fine-grained information about participants' potential exposure to people from different racial or ethnic groups, offering higher spatial resolution than IP address measurements or ZIP-codes and affording more precise estimates due to repeated measurement (Dixon et al., 2020).

Quality of Outgroup Contact

Another aspect of the ORE that has not been sufficiently or systematically addressed in previous research is the role of contact quality. Situations of outgroup contact may be perceived positively or negatively, such that positive outgroup contact has an ameliorating effect on prejudice and discrimination (Lemmer & Wagner, 2015; Pettigrew & Tropp, 2006), whereas negative contact may exacerbate prejudice and outgroup conflict (e.g., Schäfer et al., 2021). As a complement to our investigation into objective measures of outgroup exposure, the present research also investigates how the perceived quality of self-reported positive and negative contact may relate to the ORE. Based on contact theories of the ORE, positive contact should translate into better face memory, and thus, reduce the ORE. The same may hold true for negative contact, given that negative experiences may increase people's motivation to distinguish between different potentially threatening outgroup members (Hugenberg et al., 2013; e.g., Hugenberg et al., 2010). However, previous research has also shown that negative contact increases category salience (Paolini, Harwood, & Rubin, 2010), which may lead to higher category-based processing and less individuated processing. Therefore, negative contact might instead impair face memory. It thus remains an open question if positive and negative contact may similarly reduce the ORE, or if positive contact may be more effective in reducing the ORE than negative contact.

Overview of Studies

Taken together, the present research aimed at investigating relations among the ORE, different measures of contact (i.e., objective measures of potential exposure versus self-report measures), and quality of contact (i.e., positive versus negative). We investigated these questions in two different cultural contexts. Studies 1a-c relied on the regional racial demographics indicated by online participants' IP addresses as a proxy for potential other-race exposure in the context of Black and White American participants' recognition of Black and White faces. Studies 2a and b relied on the regional racial demographics indicated by participants' self-reported locations, along with different self-report contact measures, to examine ORE effects among White American and White German participants in two high-powered studies. Finally, Study 3 employed GPS tracking as a more fine grained geospatial measurement of potential other-ethnicity contact and compared effects to different self-report contact measures, in the context of White German participants' recognition of White versus Middle Eastern faces. Hypotheses and analyses of Studies 2a, b, and 3 were preregistered at the Open Science Framework. We report how we determined our sample sizes, all data exclusions, all manipulations, and all measures in the study and indicate any deviations from the preregistered methods or analyses plans. The data will be made available at the Open Science Framework after publication.[2]

[2] All analyses were conducted in R (Version 4.2.0; R Core Team, 2019) and the R-packages *corx* (Version 1.0.6.1; Conigrave, 2020), *cowplot* (Version 1.1.1; Wilke, 2019), *data.table* (Version 1.14.2; Dowle & Srinivasan, 2019), *here* (Version 1.0.1; Müller, 2017), *knitr* (Version 1.39; Xie, 2015), *lsr* (Version 0.5.2;

Studies 1a, b and c

The aim of Studies 1a-c was to compare Black and White US American participant's memory for Black and White faces as a function of their regional racial demographics as an indicator of potential outgroup contact[3]. Specifically, we geolocated participants based on their IP addresses to determine the percent of White and Black people in each participant's city. If more contact with members of the racial ingroup versus outgroup causes White participants' recognition deficit for Black faces, we would expect that White participants located in areas with more Black residents would demonstrate greater recognition for Black faces than White participants located in areas with fewer Black residents. Similarly, we would expect that Black participants located in areas with more White residents would demonstrate better recognition for White faces than Black participants located in areas with fewer White residents.

Method

Participants. For Study 1a, we recruited a total sample of $N = 348$ of Black and White American participants. Based on pre-registered exclusion criteria, we excluded data of 158 participants. Furthermore, data of 17 participants were excluded because we were unable to obtain geographic information and 13 participants were excluded due to performance below chance level, leaving final samples of $n = 69$ White participants ($M_{\text{age}} = 41.39$, $SD = 13.90$) and $n = 58$ Black participants ($M_{\text{age}} = 36.84$, $SD = 9.97$).

For Study 1b, we recruited a total sample of $N = 238$ Black American participants. Based on pre-registered exclusion criteria, we excluded data of 56 participants.

Navarro, 2015), *MBESS* (Version 4.9.1; Kelley, 2019), *papaja* (Version 0.1.1; Aust & Barth, 2020), *psych* (Version 2.2.5; Revelle, 2021), *tidyverse* (Version 1.3.1; Wickham, 2017), and *usmap* (Version 0.6.0; Di Lorenzo, 2021).

[3] Studies 1a-c reflect reanalyses of the following studies reported in this Special Issue by Simon, Chen, Sherman, & Calanchini (2022): Study 1a reports data of the control condition from Study 1, Study 1b reports data from Study 3, and Study 1c reports data from Study 4. The analyses reported here were not preregistered because the question of outgroup contact was not central to the focus of Simon et al. (2022).

Furthermore, data of 26 participants were excluded from analyzes because we were unable to obtain geographic information and data of 9 participants were excluded due to performance below chance level, leaving a final sample of $N = 116$ Black participants ($M_{\text{age}} = 37.23$, $SD = 10.92$).

For Study 1c, we recruited a total sample of $N = 126$ Black American participants. Based on pre-registered exclusion criteria, we excluded data of 38 participants. Furthermore, data of 6 participants were excluded because we were unable to obtain geographic information and data of 9 participants were excluded due to performance below chance level, leaving a final sample of $N = 63$ participants ($M_{\text{age}} = 32.51$, $SD = 10.25$).

For further details of participant recruitment and exclusion criteria, see Simon et al. (2022, this Special Issue). Sensitivity analyses indicate that sample sizes in Studies 1a-c allow to test for correlations between rs = .23-.31 with a test power of 1- $\beta = .80$ and $\alpha = .05$ (Faul, Erdfelder, Lang, & Buchner, 2007).

Designs. Study 1a followed a 2x2 mixed design, with a within-participants factor of stimulus race (Black, White) and a between-participant factor of participant race (Black, White [4]). Study 1b followed a 2x2 mixed design with a within-participant factor of stimulus race (Black, White) and a between-participant factor of stimulus gender (female, male). For analyses reported in this manuscript, we collapsed data across stimulus gender. Study 1c followed a repeated measures design with stimulus race (Black, White) and status measure (Black, White) as within-participant variables.

Procedure. After providing informed consent, participants completed an old/new recognition task. In Study 1a and c, we presented a sequence of 36 White faces and 36 Black faces in the study phase and 72 White faces and 72 Black faces in the test phase (of which 36 of each race were from the study phase and 36 of each race were new). In Study 1b, we presented a sequence of 24 White faces 24 Black faces, and 24 Latinx faces in the

[4] The experiment also included Latinx faces, but for comparability across Studies 1a-c, they were excluded from analyses.

study phase, and 48 White faces, 48 Black faces, and 48 Latinx faces in the test phase (of which 24 of each race/ethnicity were from the study phase and 24 of each race/ethnicity were new). Faces were always presented in random order. In the study phase, each face was presented for 2 seconds. In the test phase, participants' task was to indicate whether they had seen each face in the previous task using radio buttons labeled 'yes' and 'no.' Response time in the test phase was not restricted, and target faces remained on screen until a response was made. After the recognition test, participants provided demographic information. Participants' IP addresses were recorded automatically by the survey presentation software (Qualtrics, Provo, UT, USA).

Stimuli. For Studies 1a and c, we selected 72 Black and 72 White female faces from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) that were rated as most racially prototypical. For Study 1b, we selected faces generated by an artificial intelligence algorithm available at Generated Photos (https://generated.photos). A total of 288 Black, White, and Latinx faces were collected based on perceived race prototypicality (144 female and 144 male, 48 of each race/ethnicity).

Results

Other-race effect in recognition memory. Responses in the recognition task were classified into hit and false alarm rates. Based on signal detection theory (Green & Swets, 1966), we calculated sensitivity parameter *d'*, indicating participants' ability to discriminate between old and new faces[5]. Comparisons of *d'* for Black versus White faces are reported in Table 1.

In Study 1a, White participants showed an ORE, with better memory for White faces than Black faces. However, Black participants showed a reverse ORE, with better memory for White faces than Black faces. The absolute value of White participants' memory

[5] *d'* = z(Hits) - z(False Alarms). Hit rates = 1 or false alarm rates = 0 were adjusted following the procedure by Macmillan and Creelman (2005).

Table 1
The other-race effect in recognition measure d'.

			Ingroup			Outgroup						
Study	Ppts	N	Faces	M	SD	Faces	M	SD	t	p	dz	95% CI
1a	Black	58	Black	0.05	0.35	White	0.87	0.62	-7.81	< .001	-1.03	[-1.34, -0.70]
1a	White	69	White	1.07	0.64	Black	0.03	0.38	10.49	< .001	1.26	[0.94, 1.58]
1b	Black	116	Black	0.44	0.40	White	0.50	0.52	-1.00	.318	-0.09	[-0.28, 0.09]
1c	Black	63	Black	0.09	0.32	White	0.97	0.59	-7.81	< .001	-1.34	[-1.67, -0.99]
2a	White	1079	White	1.13	0.67	Black	0.84	0.59	15.95	< .001	0.49	[0.42, 0.55]
2b	White	1079	White	0.91	0.60	MENAT	0.79	0.56	6.80	< .001	0.21	[0.15, 0.27]
3	White	149	White	1.41	0.65	MENAT	1.28	0.61	2.86	.002	0.23	[0.07, 0.40]

Note. Comparisons of d' values for ingroup vs. outgroup faces in Studies 1a-c were tested in two-sided paired t-tests. Comparisons in Studies 2a, 2b, and 3 were tested in one-sided paired t-tests, as preregistered. Ppts = participants. MENAT = Middle Eastern, North Africa, and Turkey

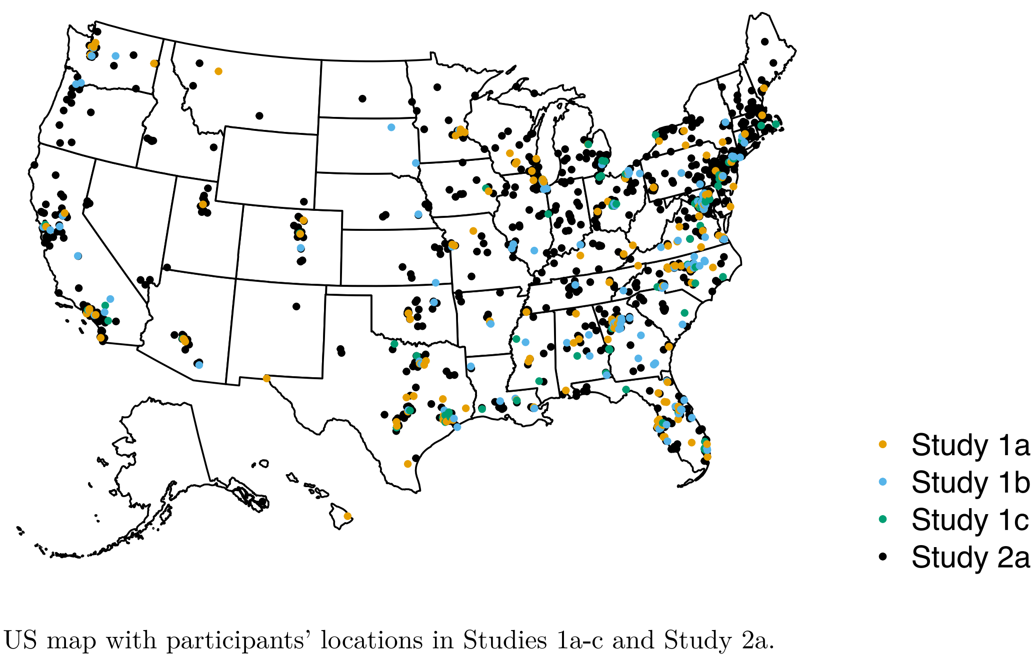

Figure 1. US map with participants' locations in Studies 1a-c and Study 2a.

advantage for White faces (d_z = 1.26) was similar to the absolute value of Black participants' memory advantage for White faces (d_z = -1.03), as revealed by a non-significant comparison of d's between the two groups, $t(125) = 1.59$, $p = .114$, $d =$ 0.28, 95% CI [-0.07; 0.64]. In Study 1b, Black participants showed no ORE on average, with equal recognition of Black and White faces. In Study 1c, Black participants showed a reverse ORE, with better memory for White faces than Black faces. In summary, only White participants displayed an ORE, whereas Black participants showed either no or reverse ORE effects.

Additional analyses of hits, false alarms, and response bias are reported in the supplement.

Potential outgroup exposure and the other-race effect. We examined whether participants' recognition for same- versus other-race faces was related to the percentage of other-race population in each participant's city. Using a location extraction website (whatismyipaddress.com), we identified participants' cities based on their IP addresses. Participants were located all over the United States (see Figure 1). Using the U.S. Census American Community Survey data for 2015-2019, we recorded the percentage of White and Black residents in each participant's city. Across cities, the average percentage of residents was M = 65.96 (SD = 21.02) White and M = 19.41 (SD = 18.97) Black in Study 1a, M = 54.58 (SD = 20.26) White and M = 31.75 (SD = 22.53) Black in Study 1b, and M = 56.75 (SD = 16.82) White and M = 28.37 (SD = 17.68) Black in Study 1c, indicating that overall, White residents were the numerical majority and Black residents a numerical minority.

As measures of the ORE, we computed two different indices (1) by subtracting *d'* for other-race faces from *d'* for own-race faces (i.e., ORE difference score) and (2) by statistically controlling for recognition of own-race faces in regression analyses (i.e. ORE regression score). Both ORE indices have been used in previous ORE research (Stelter & Degner, 2018; see for example Zhao, Hayward, & Bülthoff, 2014) and reflect memory for other-race faces after controlling for overall face memory abilities.

Correlations between ORE indices and percentage of other-race populations are depicted in Figure 2. All ORE indices were unrelated to percentages of other-race local populations.

Discussion

Studies 1a, b, and c investigated memory for own and other-race faces in Black and White American participants as a function of the percentage of other-race residents in participants' cities, which were operationalized as a proxy for potential outgroup exposure. Based on contact theories of the ORE, we hypothesized that people's memory for other-race faces should be related to the percentage of other-race residents in participants' local environments. As expected, we observed an ORE among White participants, who were better at recognizing White faces compared to Black faces. However, and contrary to expectations, we observed a reverse ORE among Black participants, who were also better or equally good at recognizing White faces compared to Black faces.

The absence of ORE effects in Black American participants might be explained by a stimulus selection bias, with selected White faces used in Studies 1a-c being more memorable than the Black faces. However, Studies 1a and 1c relied on pictures of actual people from the Chicago Face Database, whereas Study 1b relied on pictures of people generated by an artificial intelligence algorithm. This consistent pattern of results across stimulus sets speaks against the possibility of stimulus effects. Alternatively, the absence of ORE effects in Black American participants might be explained by the fact that White people represent a numerical majority and occupy a higher social status than Black people (see Simon et al. 2023 for a more detailed discussion).

Of note, participants' performance on the old/new recognition task in Studies 1a and c was generally poor: Black and White participants alike had difficulty distinguishing old from new Black faces, as reflected by d' scores close to zero. Performance close to chance level reduces the chance of finding correlations between memory and any other factors. Our analyses therefore excluded participants performing below chance levels (inclusion criterion: > 50% correct responses) and the pattern of results does not change when applying a stricter inclusion criterion (> 55% correct). The poor performance may be related to the large number of stimuli (i.e., 72 faces in the learning phase), which are more than are often used in recognition tasks like these (e.g., Stelter, Rommel, & Degner, 2021; Wiese & Schweinberger, 2018; but see Sporer & Horry, 2011). However, this pattern of results cannot solely be attributed to task difficulty. If participants were guessing, they should have also demonstrated poor memory for White faces. However, Black and White participants alike reliably distinguished old from new White faces. Consequently, to the extent that the number of trials made this task especially difficult, participants' recognition memory seems to have been impacted systematically rather than randomly. Additionally, and more directly related to the focus of the present research, ORE indices were unrelated to percentages of other-race residents in participants' local environments.

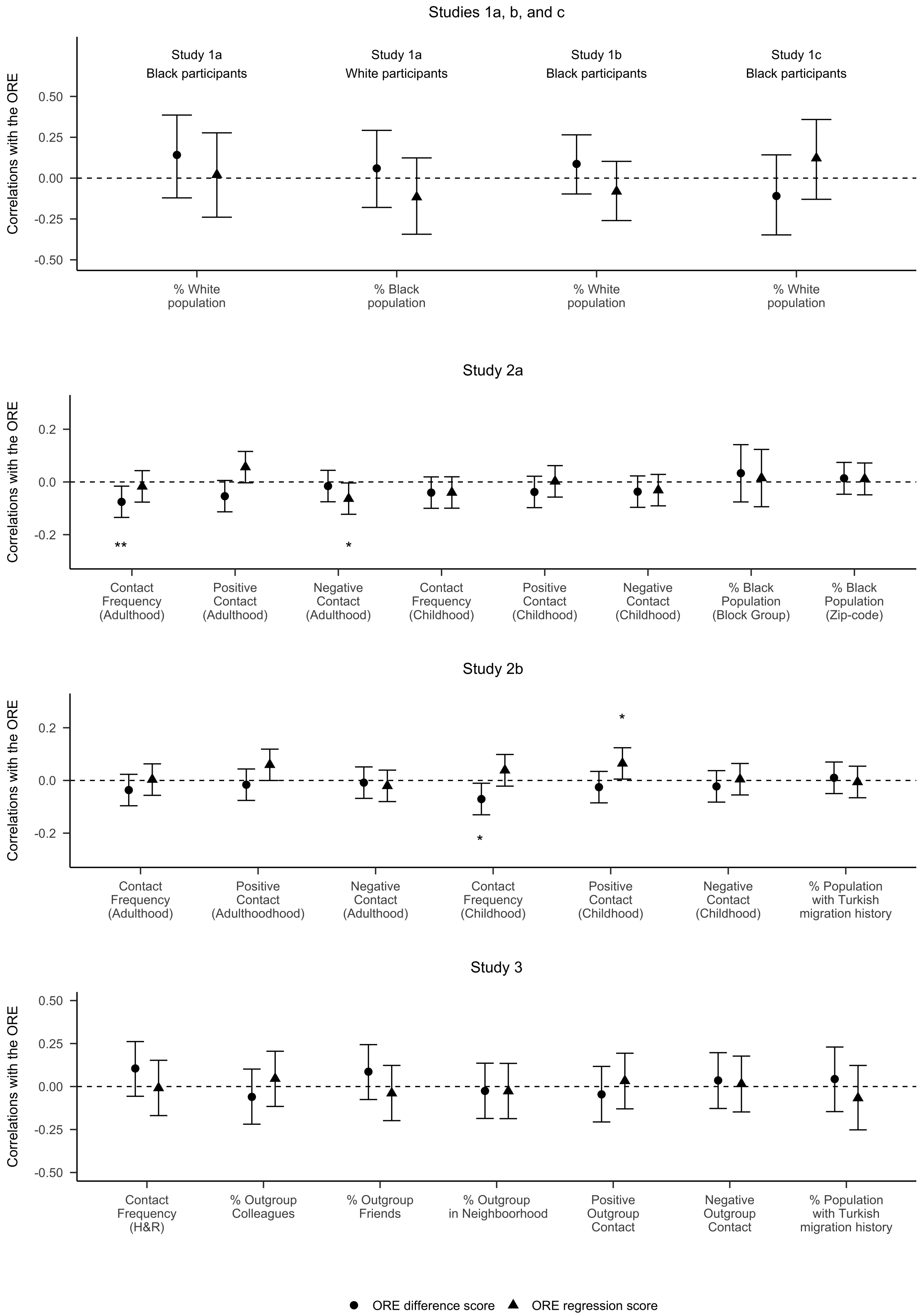

Figure 2. Correlations of ORE indices with self-report contact measures and percentage of outgroup populations as objective measures of potential outgroup exposure. * $p < .05$; ** $p < .01$. Error bars represent 95% confidence intervals.

Taken together, Studies 1a-c do not provide evidence in support of the relation between contact and the ORE. However, it should be noted that the sample sizes recruited in Studies 1a-c did not provide sufficient test power to detect very small correlations between contact and the ORE. Even descriptively small effects can have societally-large consequences when they affect many people, or the same people repeatedly (Götz, Gosling, & Rentfrow, 2022; Greenwald, Banaji, & Nosek, 2015; Rosenthal, 1990). Given the frequency with which most people interact with members of their racial ingroup and outgroups, Studies 2a and b addressed this limitation by recruiting large participant samples.

Studies 2a and b

Studies 2a and b aimed at further investigating relations between different contact measures and the ORE in two high powered studies conducted in two different cultural contexts. Study 2a measured recognition for White versus Black faces in White American participants. Study 2b measured recognition for White versus MENAT[6] faces in White German participants. People with ancestors from the MENAT region represent a sizable ethnic minority in the German society (Statistisches-Bundesamt, 2020) and previous research has observed reduced memory among White Germans for Turkish or Middle Eastern faces compared to White German faces (Horry & Wright, 2009; Singmann, Kellen, & Klauer, 2013; Sporer, 2001; Stelter & Degner, 2018; Stelter et al., 2021; but see Sporer & Horry, 2011).

As an objective measure of potential outgroup exposure, we assessed zip-code areas and census block groups (Study 2a) and city districts (Study 2b). Furthermore, participants reported quantity of outgroup contact (i.e., percentage of people perceived as MENAT/Black) in different areas of participants' daily lives during adulthood and

[6] We use the term "MENAT" to refer to people who are perceived to have ancestors from Middle Eastern countries, North African countries, or from Turkey (see Awad, Kia-Keating, & Amer, 2019).

childhood (i.e., at professional places such work and school, among family and close friends, among acquaintances, and among people in their neighborhood). Furthermore, participants rated the quality of outgroup contact (i.e., how often they have had positive and negative experiences with MENAT/Black people in the respective area of their daily lives).

Methods

Participants. We aimed to recruit N = 1073 White American participants in Study 2a and N = 1073 White German participants in Study 2b. Sample sizes were estimated to test for correlations of r = .10 with a test power of .95 and $\alpha = .05$. We preregistered several exclusion criteria and thus planned to oversample and recruit 1180 participants for each study.[7]

Participants in Study 2a were recruited via Prolific and reimbursed with the equivalent of 5 £/h. Participants were required to live in the United States and to self-identify as White. We recruited a total sample of N = 1339 participants. Following the preregistration, we excluded data of 118 participants who did not self-identify as White, 58 participants who did not live in the US for a minimum of ten years, and 84 participants who performed below chance in the recognition task. The final sample included N = 1079 participants ($M_{\text{age}} = 32.29$, $SD = 60.65$).

Participants in Study 2b were recruited via the German panel provider Respondi. Participants had to be over 18 years of age, self-identify as White, and were required to live in the German cities Berlin, Hamburg, Cologne, Frankfurt or Duesseldorf. These cities were selected because they provided publicly accessible information regarding ethnic composition of the population within city districts. We recruited a total sample of N = 1172 participants. Following the preregistration, we excluded data of 93 participants who performed below chance in the recognition task. The final sample included N = 1079

[7] See preregistrations here: https://osf.io/p2ejq and https://osf.io/gf3jk.

participants ($M_{\text{age}} = 52.74$, $SD = 14.30$).

Procedure. Participants completed the study online. First, participants provided written consent. In Study 2b, participants then provided demographic information regarding their age, gender, ethnicity and the city in which they live, in order to meet the screening criteria. In Study 2a this information was provided after completion of the old/new recognition task.

In both studies, participants then completed an old/new recognition task programmed in PsychoPy and hosted on Pavlovia. In Study 2a, the recognition task included Black and White faces; in Study 2b, the recognition task included MENAT and White faces. After the recognition task, participants were forwarded to a Qualtrics survey in which they completed self-report contact measures and a one-item group preference measure. Finally, participants were debriefed about the study aims and provided consent for use of their data. The studies took approximately 15 minutes to complete.

Measures.

Old/new recognition task. The recognition task consisted of a study phase and a test phase. During the study phase, participants viewed a randomized series of 48 faces (Study 2a: 24 Black, 24 White; Study 2b: 24 MENAT, 24 White), each presented for 4 seconds. The faces in Study 2a were selected from the Chicago Face Database (Ma et al., 2015). The faces in Study 2b were selected from the Amsterdam Face Database (Schalk, Hawk, Fischer, & Doosje, 2011), the Radboud Face Database (Langner et al., 2010), and a Google Image search. All faces were cropped to oval shapes to exclude external features of the faces. During the subsequent test phase, participants viewed a series of 96 faces, 48 of which had been presented during the study phase (i.e., “old” faces) and 48 had not been presented during the study phase (i.e., “new” faces). Participants responded via key press if a face was old or new. Response times were not restricted and the target faces remained on screen until a response was made.

Self-reported outgroup contact. As a measure of contact quantity, participants were asked to think about colleagues at work, other people at work, family and close friends, acquaintances, and people in their neighborhood, and estimate the percentage of these people they perceive to be White, Black/MENAT, and neither White nor Black/MENAT, with which they had contact in each context (Cronbach's α = 0.83 - 0.81). Each question about contact quantity was followed by a rating of contact quality, in which participants rated how often they have had positive and negative experiences with Black/MENAT people in the respective area of their everyday lives (1 = "never," 7 = "very often"; Cronbach's α = 0.80 - 0.91). Next, participants indicated contact quantity and quality during childhood, rating the percentage of contact at school, with childhood friends, and with people in their neighborhood (Cronbach's α = 0.84 - 0.90). Again, each question about contact quantity during childhood was followed by a rating of contact quality, in which participants rated how often they have had positive and negative experiences with Black/MENAT people in the respective area of their childhood lives (1 = "never," 7 = "often"; Cronbach's α = 0.77 - 0.92).

Self-reported prejudice. We used a single-item measure (Axt, 2018) to assess relative preference for Black versus White people in Study 2a (1 = "I strongly prefer Black people to White people."; 7 = "I strongly prefer White people to Black people.") and MENAT versus White people in Study 2b (1 = "I strongly prefer MENAT people to White people."; 7 = "I strongly prefer White people to MENAT people.").

Objective measure of potential outgroup exposure. In Study 2a, participants reported their zip-code. Additionally, as a more geographically-precise index of location, participants were asked to find their census block group number using an online tool (https://www.arcgis.com/apps/webappviewer). Only n = 325 participants (30.12%) reported their census block group information. Based on the U.S. Census American Community Survey data for 2015-2019, we retrieved the percentage of Black population for each zip-code area and census block group.

In Study 2b, participants named the city district in which they currently lived and reported the number of years they have lived in the district (i.e., "< 1 year," "> 1 year," "> 2 years," "> 3 years," "> 4 years," and "> 5 years"). If they have lived in a district for less than a year, they additionally named the district in which they have lived before. Based on local census data, we derived the percentage of residents with Turkish migration history [8] for each city district.

Results

Other-race effect in recognition memory. Comparisons of *d'* for Black versus White faces in Study 2a and MENAT versus White faces in Study 2b are reported in Table 1. In both studies, participants displayed significant OREs, reflected in lower *d'* values for Black faces than White faces (Study 2a) and lower *d'* values for MENAT faces than White faces (Study 2b). Exploratory comparisons of hits, false alarms and response bias are reported in the supplement.

Contact and the other-race effect. Following the preregistration, we computed an ORE difference-score. For additional exploratory analyses, we also computed a regression-based ORE score. Correlations between ORE indices and contact measures are depicted in Figure 2, and complete correlation matrices are reported in the supplement. Following the preregistration, correlation of the subtraction-based ORE index with frequency of contact with Black people and positive contact with Black people was tested one-sided. All other correlations were tested two-sided.

Out of 30 comparisons with contact measures, we observed 4 significant correlations (see Figure 2). In Study 2a, the ORE difference score correlated negatively with the frequency of contact with Black people during adulthood ($r[1077] = -0.08$, $p_{\text{one-sided}} = 0.01$, 95% CI [-1.00; -0.03]), indicating that higher levels of contact during adulthood were

[8] Residents with Turkish migration history are consistently reported across all local census reports and were therefore used as an estimate for potential exposure with people perceived as MENAT.

associated with smaller ORE effects. The ORE regression score correlated negatively with negative contact with Black people during adulthood ($r[1077] = -0.06$, $p_{\text{two-sided}} = 0.04$, 95% CI [-0.12; 0.00]), indicating that more negative contact during adulthood was associated with *better* memory for other-race faces. In Study 2b, the ORE difference score correlated negatively with the frequency of contact during childhood with people perceived as MENAT ($r[1060] = -0.07$, $p_{\text{one-sided}} = 0.01$, 95% CI [-1.00; -0.02]) and the ORE regression score correlated positively with the amount of positive contact during childhood with people perceived as MENAT ($r[1076] = 0.06$, $p_{\text{two-sided}} = 0.03$, 95% CI [0.01; 0.12]); thus, more negative and more positive contact during childhood was associated with smaller ORE effects. In addition to these relationships between contact and the ORE, group preference also correlated with the ORE difference score ($r[1075] = 0.07$; $p_{\text{two-sided}} = 0.03$, 95% CI [0.01; 0.13]), indicating that participants with higher preference for White people displayed a larger ORE. No other correlations were reliably different from zero.

Discussion

Studies 2a and b investigated relationships between self-reported contact, objective geographical measures of potential outgroup exposure, and the ORE in two high-powered studies. We observed ORE effects in both studies: White American participants displayed better memory for White versus Black faces, and White German participants displayed better memory for White compared to MENAT faces. In addition, we observed correlations between contact in the ORE in 4 out of 30 tests, of which three indicated that more contact was related to smaller ORE effects. All correlations between contact measures and the ORE were very small, with contact explaining less than 1% of the variance in ORE effects – with observed effect sizes that are even smaller than the average effect sizes reported in previous meta-analyses (Meissner & Brigham, 2001; Singh et al., 2021). Studies 2a and b operationalized potential outgroup exposure in terms of local demographics, which were inferred from participants' residential locations. However, local demographics in and of themselves do not provide information about the actual places visited by participants – and, thus, the people participants come in contact with. Study 3 addresses this limitation by employing GPS tracking as a more precise measure of participants' location.

Study 3

Study 3 aimed to further investigate the relation between the ORE and potential outgroup exposure by employing a more fine-grained measure of participants' location. To do so, we tracked participants' GPS locations for the course of two weeks. Due to residential ethnic segregation, people's potential outgroup exposure may vary depending on the city districts in which participants spend their time. This GPS-based measurement approach has the advantage of being able to locate participants in their everyday lives and not just to their place of residence, thus painting a more complete picture of potential outgroup exposure. In addition to the GPS-based measure, we included several different self-report contact measures that distinguish between contact frequency and contact quality.

Methods

Participants. We aimed to recruit a sample of N = 156 White German participants, which allows detecting a small effect of $d_z = 0.20$ for the mean difference between *d'* for White faces and *d'* for MENAT faces with a test power of 1-β =.80 [assuming a correlation of r = .5 between measures and α = 0.05; Faul et al. (2007)]. This sample size would also provide sufficient test power to detect a correlation of r = .22 with 1-β = 0.80 and α = 0.05. We preregistered to oversample and recruit 200 participants [9]. Due to the onset of Covid-19 restrictions, we were unable to recruit the full sample. Participants were compensated with 10 to 30 Euros, depending on if they completed only

[9] See preregistration here https://osf.io/jakfq.

the first part or both parts of the study.

Participants were recruited via the Hamburg University job board or via targeted Facebook advertising and had to be over 18 years of age and to own a smartphone to be invited to the study. We recruited a total sample of N = 186 participants. Following the preregistration, we excluded data of 35 participants who did not self-identify as White and data of 2 participants who performed below chance in the recognition task. A total of N = 149 participants (M_{age} = 27.89, SD = 10.13) completed the first part of the study, and n = 109 of those participants also completed the second part of the study.

Procedure. Participants were tested in the Social Psychology Department computer lab at Hamburg University in groups up to eight. Participants were informed that the study consisted of two parts: (1) a memory task and questionnaires in the laboratory and (2) measurement of GPS-data via a smartphone app during the following two weeks. Participants then signed a written consent form. Next, participants completed an old/new recognition task with MENAT and White faces. After the recognition task, participants completed self-reported contact measures, a one-item group preference measure, and provided demographic information. At the end of the first part of the study, participants were debriefed about the study aims and provided consent for use of their data. Participants who did not want to complete the second part of the study (n = 40) were compensated with 10 Euros.

Next, participants signed a second written consent form and installed the GPS-tracking app on their android smartphones. The app measured GPS-signal and motion data for the next two weeks. Participants could interrupt data collection at any time within the app. Furthermore, they could delete the app at any time without any further consequences. After installing the app, participants were compensated with 30 Euros.

Measures.

Old/New recognition task. The procedure and stimuli used in Study 3 were identical to Study 2a.

Self-reported outgroup contact. Self-reported outgroup contact was measured using a translated and adapted version of the contact questionnaire from Hancock and Rhodes (2008) consisting of 15 statements asking about contact with people perceived as MENAT and people perceived as White (e.g., "I interact with people perceived as MENAT/White on a daily basis."; 1 = Very strongly disagree; 6 = Very strongly agree). Responses for items assessing contact with people perceived as MENAT were averaged to a mean contact score (Cronbach's $\alpha = 0.80$). To measure outgroup contact in different settings, participants additionally estimated the percentages of people perceived as MENAT, people perceived as White, and people perceived as neither MENAT nor White (a) among their colleagues or fellow students, (b) among their friends, and (c) in their neighborhood (see McKone et al., 2019). Finally, to measure quality of outgroup contact, we used a questionnaire translated and adapted from Dixon and colleagues (2020) consisting of 10 statements on positive and negative outgroup contact, assessing the extent to which participants experienced contact with people perceived to be MENAT as *friendly*, *helpful*, *welcoming*, *positive*, and *cooperative* versus *negative*, *disrespectful*, *unfriendly*, *verbally abusive*, and involving *ridicule*. Responses to items measuring positive and negative contact were averaged to two separate mean scores (Cronbach's $\alpha = 0.82$ - 0.83).

Self-reported prejudice. We used a single-item measure (Axt, 2018) asking for relative preference for MENAT versus White people (1 = "I strongly prefer MENAT people to White people."; 7 = "I strongly prefer White people to MENAT people.").

Objective measure of potential outgroup exposure. The tracking app collected GPS coordinates and timestamps whenever it registered a change in location. GPS coordinates were translated into corresponding city districts and data points that were outside of [city name] were removed from analyses. Based on the timestamps, we computed "time visited" scores, indicating the amount of time spent in a city district. For each city district, we retrieved the proportion of residents with Turkish migration history reported by the local 2018 microcensus[10]. The proportion of residents with Turkish migration history in a given district was weighted by the time participants spend in the district and averaged to a score indicating participants' individual level of potential outgroup exposure. As a fictional example, a person spending half of the time in a district with 10% residents with Turkish migration history and the other half of the time in a district with 0% residents with Turkish migration history would have a potential outgroup exposure score of 5%.

Results

Other-race-effect in recognition memory. Comparison of recognition measure *d'* for MENAT versus White faces is reported in Table 1. Participants displayed an ORE, reflected in lower *d'* values for MENAT faces than for White faces. Preregistered comparisons of hits, false alarms and exploratory comparison of response bias are reported in the supplement.

Contact and the Other-Race Effect. Correlations between contact measures and ORE indices are depicted in Figure 2 and reported in Table 2. Following the preregistration, we computed the ORE difference-score. For additional exploratory analyses, we also computed the ORE regression-score. Following the preregistration, correlations between the ORE difference-score with contact measured via the Hancock and Rhodes (2008) scale, frequency of contact, and positive contact were tested one-sided.

[10] The 2018 microcensus reports only on the proportion of people with Turkish migration history; information on residents with migration history from Middle Eastern or North-African countries is not reported.

Table 2
Correlations between ORE indices, self-reported contact, and potential real-life exposure.

	1	2	3	4	5	6	7	8	9	10	*M*	*SD*
1. ORE difference score	-	149	149	147	146	149	149	149	109	149	0.13	0.56
2. ORE regression score	-.87***	-	149	147	146	149	149	149	109	149	0.00	0.49
3. HR contact scale	.11	-.01	-	147	146	149	149	149	109	149	2.75	0.92
4. Positive contact	-.05	.03	.10	-	144	147	147	147	108	147	1.77	0.68
5. Negative contact	.04	.02	.52***	-.02	-	146	146	146	107	146	3.00	0.76
6. Contact among colleagues	-.06	.05	.41***	.21*	.33***	-	149	149	109	149	17.04	15.50
7. Contact among friends	.09	-.04	.45***	.11	.47***	.65***	-	149	109	149	9.60	13.67
8. Contact in neighborhood	-.03	-.03	.24**	.37***	.15	.57***	.43***	-	109	149	20.19	18.63
9. % Residents with Turkish migration history	.04	-.07	.07	.31**	.06	.08	.00	.45***	-	109	11.02	4.80
10. Prejudice	-.13	.20*	-.22**	.20*	-.49***	-.18*	-.32***	-.04	.06	-	4.60	0.81

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; values in the lower triangle represent correlation coefficients; values in the upper triangle represent number of pair-wise cases. HR = Hancock and Rhodes (2008)

Neither of the ORE scores correlated with any of the contact measures. However, exploratory analyses showed that self-report contact measures were inter-correlated (see Table 2). Additionally, reported positive contact and reported percentage of people in the neighborhood perceived as MENAT correlated with the GPS measure of potential outgroup exposure. This pattern of correlations supports the validity of our contact measures as assessing their intended constructs. Finally, reported group preference was correlated with the ORE regression score ($r[147] = 0.20$; $p = 0.02$, 95% CI [0.04; 0.35]), indicating that more positive evaluations of people perceived as MENAT was associated with better memory for MENAT faces.

Discussion

Study 3 investigated contact theories of the ORE by examining relations among memory performance for own versus other-ethnicity faces and two types of contact measures: self-reported contact and an objective measure of potential outgroup exposure. We observed an ORE in recognition memory: White German participants better recognized White faces than MENAT faces. However, and contrary to predictions, the ORE was not related to any of the self-report contact measures, nor to the GPS-based measure of potential outgroup exposure. Instead, the only significant correlation that emerged in Study 3 was between the ORE regression-score and the group preference measure, indicating that people with stronger preference for White versus MENAT people exhibited larger ORE effects.

General Discussion

A wide variety of theoretical perspectives converge to posit a central role of contact in the ORE, predicting that this intergroup recognition deficit is attenuated among people who have relatively more contact with outgroup members. However, empirical evidence to date provides mixed support for the moderating role of outgroup contact in the context of the ORE. The present research investigated relationships between memory for own- versus other-race faces and a range of contact measures. In addition to relying on widely-used self-report measures of contact frequency and contact quality, the present research also included objective measures of potential outgroup exposure based on the demographic composition of participants' locations.

We observed ORE effects that replicated previous research: Members of majority groups (i.e., White people in the US and Germany) better remembered the faces of racial ingroup versus outgroup members (i.e., Black and MENAT people, respectively). The exception to this pattern of results was Black people in America, who either remembered the faces of racial outgroup members (i.e., White people) better than of their ingroup, or showed no difference in recognition performance between both groups. Moreover, and of central importance to the present research, across a wide variety of operationalizations of contact, we observed no consistent relations between contact and the ORE. Only in Studies 2a and 2b, using very large samples, we observed small but significant correlations (rs = -.08 - .06) between ORE-indices and 4 out of 30 contact measures. In addition to being small, these correlations were inconsistent across operationalizations of the ORE (e.g., some forms of contact correlated with difference-based ORE indices but not regression-based ORE indices). These findings add to the small and inconsistent correlations between ORE and contact measures reported in previous research (see Singh et al., 2021).

Contact theories of the ORE assume that impaired memory for other-race faces is caused by reduced visual expertise due to limited contact with members of other racial groups (e.g., Tanaka et al., 2004; Valentine, 1991). However, small or absent relationships between contact measures and the ORE in previous research seem to speak against contact theories of the ORE. That said, previous research may be limited by its reliance on self-reported measures of contact, which may be susceptible to biases such as demand characteristics, self-presentation concerns, or the limitations of accurate memory. Consequently, in the present research we relied on geographical information to infer participants' extent of outgroup contact. Given the wide variety of operationalizations of contact and participant samples reflected here, the present research would seem to be well-positioned to find relations between contact and the ORE. However, and in contrast to the predictions of contact theories of the ORE, we found that the ORE was unrelated to most operationalizations of contact examined here – and the few significant correlations we found were very small. Our findings converge with the small and inconsistent relations between contact and the other-race effect reported in the literature, which suggests that contact theories of the ORE currently lack strong empirical support. However, abandoning contact theories of the ORE altogether might be too hasty, as several alternative explanations may account for the small or absent correlations between contact measures and the ORE.

First, it is possible that our novel operationalizations of outgroup contact in terms of geographic location did not tap into the kind of contact conceptualized by contact theories of the ORE. The same criticism can reasonably be applied to conventional self-report measures, which we also used in the present research, and have been widely used in previous research. We cannot rule out the possibility that none of these measures capture the essential features of contact necessary for reliable outgroup memory. Notably, most theories of the ORE do not explicitly specify what kind of contact is necessary to form more accurate representations of other-race faces. Some perspectives posit that long-term, or even lifetime, experience is necessary to reduce or eliminate the ORE (e.g., Chiroro & Valentine, 1995; Sangrigoli, Pallier, Am Argenti, Ventureyra, & Schonen, 2005; Wan, Crookes, Reynolds, Irons, & McKone, 2015). In contrast, our objective measures of explicit outgroup exposure tap only into present or most recent outgroup exposure, and may not fully reflect people's exposure to and contact with outgroups over the course of their lives – especially in contexts with high individual mobility (e.g., when participants move between areas with different population composition). Alternatively, it is possible that the kind of contact most effective in reducing the ORE does not necessarily need to be in-person contact. For example, indirect outgroup contact (i.e., via media) may suffice to improve outgroup face memory. From this perspective, self-report and objective contact measures may underestimate the totality (i.e., direct and indirect) of people's exposure to and expertise with outgroup faces. Finally, contact that is effective in reducing the ORE might depend on motivation to closely pay attention to another person's face. For example, people are motivated to attend to others of higher social status (e.g., Shriver & Hugenberg, 2010) or in situations of outcome dependency (e.g., Baldwin, Keefer, Gravelin, & Biernat, 2013). Consequently, measures that assess mere outgroup exposure without also accounting for motivations may be insufficient to capture the kind of contact experiences necessary to reduce the ORE. Future theorizing should be more explicit in defining the essential characteristics of contact experiences that matter for face recognition, such that future research can investigate relations between the ORE and contact measures that more closely tap into specific contact situations in which individuated face processing is applied.

Another explanation for the lack of correlations observed between contact measures and the ORE may lie in a lack of meaningful variance in either of the measures and/or low measurement reliability. Measurement reliability places an upper limit on how strongly two measures can correlate (Spearman, 1904), so unreliable measures of contact or the ORE may have concealed true relations between the two constructs. The self-report contact measures used in the present research demonstrate satisfactory internal consistencies (ranging from α = .77 to .92). Conversely, the old/new recognition task routinely used to measure the ORE has been documented to have quite low internal consistencies (e.g., DeGutis et al., 2013; Stelter & Degner, 2018). In the present studies, split-half reliabilities of old/new recognition tasks ranged from 0.78 in Study 1a to -0.29 in Study 3. Thus, low reliability of the ORE measurement may have posed a limitation for some of the studies, as reliable measurement is a precondition for finding correlations with other measures. Previous research has reported that versions of the Cambridge Face Memory Test might be a better alternative to measure ORE effects with high internal reliability (McKone et al., 2012). So far, versions of this test are available only for White and Asian faces and are therefore not applicable in the cultural contexts of the present studies. Future research may benefit from creating new versions of the Cambridge Face Memory Test including Black or MENAT faces.

Lastly, outgroup contact may not always be the most important predictor of the ORE, and we may eventually need to abandon experience-based explanations of the ORE for some contexts. Specifically, alternative theories of the ORE highlight the role of social-motivational factors in recognition memory for outgroup faces (e.g., Hugenberg et al., 2010; Rodin, 1987; Sporer, 2001). Broadly speaking, these theories assume that people are generally *able* to encode and recognize other-race, but are *less motivated* to employ sufficient effort to attend to individuating information in other-race faces, and instead attend to category-related information. In line with these theories, previous research has shown that the ORE can be reduced when participants are motivated to focus on individuating features of other-race faces (e.g., Hugenberg, Miller, & Claypool, 2007). These effects might be particularly pronounced in contexts in which participants have sufficient visual expertise but little motivation to recognize other-race faces (Wan et al., 2015). In the context of the present studies, most participants have reported at least a minimum amount of contact with members of other racial or ethnic groups. It is thus possible that most participants in the present studies had enough visual expertise to encode other-race faces. If this was the case, the ORE effects that emerged in the present studies may primarily reflect motivational factors. Future research should attempt to test contact and social-motivational theories in direct comparison, in order to examine the kinds of context in which either one or the other theories are more powerful predictors of the other.

In summary, the present research provides little empirical support for a relationship between the ORE and a wide variety of outgroup contact. Furthermore, the few reliable correlations between ORE and contact that emerged in the present research were very small, accounting for less than 1% of variance in the ORE. Given the potentially harmful consequences of the ORE (McKone et al., 2021; Wells & Olson, 2001), it remains important to better understand how the ORE is caused. Future research should ideally take a joint focus on contact theories and social-motivational accounts to derive more precise predictions of how the ORE can be reduced.

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Awad, G. H., Kia-Keating, M., & Amer, M. M. (2019). A model of cumulative racial–ethnic trauma among Americans of Middle Eastern and North African (MENA) descent. *American Psychologist*, *74*(1), 76–87. https://doi.org/10.1037/amp0000344

Axt, J. R. (2018). The Best Way to Measure Explicit Racial Attitudes Is to Ask About Them. *Social Psychological and Personality Science*, *9*(8), 896–906. https://doi.org/10.1177/1948550617728995

Baldwin, M., Keefer, L. A., Gravelin, C. R., & Biernat, M. (2013). Perceived importance of cross-race targets facilitates recall: Support for a motivated account of face memory. *Group Processes & Intergroup Relations*, *16*(4), 505–515. https://doi.org/10.1177/1368430212460893

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science*, *2*, 396–403.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *48*(4), 879–894. https://doi.org/10.1080/14640749508401421

Conigrave, J. (2020). *Corx: Create and format correlation matrices*. Retrieved from https://CRAN.R-project.org/package=corx

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Di Lorenzo, P. (2021). *Usmap: US maps including alaska and hawaii*. Retrieved from https://CRAN.R-project.org/package=usmap

Dixon, J., Tredoux, C., Davies, G., Huck, J., Hocking, B., Sturgeon, B., … Bryan, D. (2020). Parallel lives: Intergroup contact, threat, and the segregation of everyday activity spaces. *Journal of Personality and Social Psychology*, *118*(3), 457–480. https://doi.org/10.1037/pspi0000191

Dowle, M., & Srinivasan, A. (2019). *Data.table: Extension of ‘data.frame‘*. Retrieved from https://CRAN.R-project.org/package=data.table

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small Effects: The Indispensable Foundation for a Cumulative Psychological Science. *Perspectives on Psychological Science*, *17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Oxford, England: John Wiley.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561. https://doi.org/10.1037/pspa0000016

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45–56. https://doi.org/10.1348/000712607X199981

Horry, R., & Wright, D. B. (2009). Anxiety and Terrorism: Automatic Stereotypes Affect Visual Attention and Recognition Memory for White and Middle Eastern Faces. *Applied Cognitive Psychology*, *23*(3), 345–357.

https://doi.org/10.1002/acp.1465

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334. https://doi.org/10.1016/j.jesp.2006.02.010

Hugenberg, K., Wilson, J. P., See, P. E., & Young, S. G. (2013). Towards a synthetic model of own group biases in face memory. *Visual Cognition*, *21*(9-10), 1392–1417. https://doi.org/10.1080/13506285.2013.821429

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The Categorization-Individuation Model: An Integrative Account of the Other-Race Recognition Deficit. *Psychological Review*, *117*(4), 1168–1187. https://doi.org/10.1037/a0020463

Jena Declaration. (2019). Retrieved from https://www.uni-jena.de/en/190910_JenaerErklaerung_EN

Kelley, K. (2019). *MBESS: The MBESS r package*. Retrieved from https://CRAN.R-project.org/package=MBESS

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & Knippenberg, A. van. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*(2), 152–168. https://doi.org/10.1002/ejsp.2079

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide, 2nd ed.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

McKone, E., Dawel, A., Robbins, R. A., Shou, Y., Chen, N., & Crookes, K. (2021). Why the other-race effect matters: Poor recognition of other-race faces impacts everyday social interactions. *British Journal of Psychology*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/bjop.12508

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., … Pelleg, M. (2012). A Robust Method of Measuring Other-Race and Other-Ethnicity Effects: The Cambridge Face Memory Test Format. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0047956

McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., … Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Scientific Reports*, *9*, 12820. https://doi.org/10.1038/s41598-019-49202-0

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces - A meta-analytic review. *Psychology Public Policy and Law*, *7*(1), 3–35. https://doi.org/10.1037//1076-8971.7.1.3

Müller, K. (2017). *Here: A simpler way to find your files.* Retrieved from https://CRAN.R-project.org/package=here

Navarro, D. (2015). *Learning statistics with r: A tutorial for psychology students and other beginners. (Version 0.5).* Adelaide, Australia: University of Adelaide. Retrieved from http://ua.edu.au/ccs/teaching/lsr

Paolini, S., Harwood, J., & Rubin, M. (2010). Negative Intergroup Contact Makes Group Memberships Salient: Explaining Why Intergroup Conflict Endures.

Personality and Social Psychology Bulletin, *36*(12), 1723–1738. https://doi.org/10.1177/0146167210388667

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

R Core Team. (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography*, *45*(3), 489–514. https://doi.org/10.1353/dem.0.0019

Revelle, W. (2021). *Psych: Procedures for psychological, psychometric, and personality research.* Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Rodin, M. (1987). Who is memorable to whom: A study of cognitive disregard. *Social Cognition*, *5*(2), 144–165. https://doi.org/10.1521/soco.1987.5.2.144

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist*, *45*(6), 775–777. https://doi.org/10.1037/0003-066X.45.6.775

Sangrigoli, S., Pallier, C., Am Argenti, Ventureyra, V. A. G., & Schonen, S. de. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*(6), 440–444. Retrieved from https://search.proquest.com/docview/195593669?accountid=11262

Schäfer, S. J., Kauff, M., Prati, F., Kros, M., Lang, T., & Christ, O. (2021). Does negative contact undermine attempts to improve intergroup relations? Deepening the understanding of negative contact and its consequences for intergroup contact research and interventions. *Journal of Social Issues*, *77*(1), 197–216. https://doi.org/10.1111/josi.12422

Schalk, J. van der, Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907–920. https://doi.org/10.1037/a0023853

Shriver, E. R., & Hugenberg, K. (2010). Power, individuation, and the cross-race recognition deficit. *Journal of Experimental Social Psychology*, *46*(5), 767–774. https://doi.org/10.1016/j.jesp.2010.03.014

Simon, D., Chen, J. M., Sherman, J. W., & Calanchini, J. (2022). A recognition advantage for members of higher-status racial groups. *British Journal of Psychology*, *00*, 1–24. https://doi.org/10.1111/bjop.12587

Singh, B., Mellinger, C., Earls, H. A., Tran, J., Bardsley, B., & Correll, J. (2021). Does Cross-Race Contact Improve Cross-Race Face Perception? A Meta-Analysis of the Cross-Race Deficit and Contact. *Personality and Social Psychology Bulletin*, 01461672211024463. https://doi.org/10.1177/01461672211024463

Singmann, H., Kellen, D., & Klauer, K. C. (2013). Investigating the other-race effect of Germans towards Turks and Arabs using multinomial processing tree models. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*. Retrieved from https://escholarship.org/uc/item/3nr6504j

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*, 72–101. https://doi.org/10.1037/11491-005

Sporer, S. L. (2001). Recognizing faces of other ethnic groups - An integration of theories. *Psychology Public Policy and Law*, *7*(1), 36–97. https://doi.org/10.1038//1076-8971.7.1.36

Sporer, S. L., & Horry, R. (2011). Recognizing Faces from Ethnic In-groups and Out-groups: Importance of Outer Face Features and Effects of Retention Interval. *Applied Cognitive Psychology*, *25*(3), 424–431.

https://doi.org/10.1002/acp.1709

Statistisches-Bundesamt. (2020). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2020 [Results from the German Microcensus 2020].* Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/Downloads-Migration/migrationshintergrund-2010220207004.pdf

Stelter, M., & Degner, J. (2018). Investigating the other-race effect in working memory. *British Journal of Psychology*, *109*(4), 777–798. https://doi.org/10.1111/bjop.12304

Stelter, M., Rommel, M., & Degner, J. (2021). (Eye-) Tracking the Other-Race Effect: Comparison of Eye Movements During Encoding and Recognition of Ingroup Faces With Proximal and Distant Outgroup Faces. *Social Cognition*, *39*(3), 366–395. https://doi.org/10.1521/soco.2021.39.3.366

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, *93*(1), B1–B9. https://doi.org/10.1016/j.cognition.2003.09.011

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, *43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition*, *144*, 91–115. https://doi.org/10.1016/j.cognition.2015.07.011

Wells, G. L., & Olson, E. A. (2001). The other-race effect in eyewitness identification - What do we do about it? *Psychology Public Policy and Law*, *7*(1), 230–246. https://doi.org/10.1037//1076-8971.7.1.230

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wiese, H., & Schweinberger, S. R. (2018). Inequality between biases in face memory: Event-related potentials reveal dissociable neural correlates of own-race and own-gender biases. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *101*, 119–135. https://doi.org/10.1016/j.cortex.2018.01.016

Wilke, C. O. (2019). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=cowplot

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.org/knitr/

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*(2), 116–142. https://doi.org/10.1177/1088868311418987

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research*, *105*, 61–69. https://doi.org/10.1016/j.visres.2014.09.006