Empirical Research Paper

Personality and Social Psychology Bulletin
1–23

DOI: 10.1177/01461672211024463
journals.sagepub.com/home/pspb
SAGE

Does Cross-Race Contact Improve Cross-Race Face Perception? A Meta-Analysis of the Cross-Race Deficit and Contact

Balbir Singh[1], Christopher Mellinger[1], Holly A. Earls[1], Janis Tran[1], Brighid Bardsley[1], and Joshua Correll[1]

Abstract

Contact with racial outgroups is thought to reduce the cross-race recognition deficit (CRD), the tendency for people to recognize same-race (i.e., ingroup) faces more accurately than cross-race (i.e., outgroup) faces. In 2001, Meissner and Brigham conducted a meta-analysis in which they examined this question and found a meta-analytic effect of $r = -.13$. We conduct a new meta-analysis based on 20 years of additional data to update the estimate of this relationship and examine theoretical and methodological moderators of the effect. We find a meta-analytic effect of $r = -.15$. In line with theoretical predictions, we find some evidence that the magnitude of this relationship is stronger when contact occurs during childhood rather than adulthood. We find no evidence that the relationship differs for measures of holistic/configural processing compared with normal processing. Finally, we find that the magnitude of the relationship depends on the operationalization of contact and that it is strongest when contact is manipulated. We consider recommendations for further research on this topic.

Keywords

cross-race contact, cross-race recognition deficit, own-race bias, face recognition, meta-analysis, contact, cross-race effect, own-race effect

Received March 8, 2020; revision accepted May 21, 2021

The cross-race recognition deficit (CRD) refers to the fact that recognition and memory are generally worse for cross-race faces (i.e., faces from other races) compared with same-race faces. The CRD is implicated in many social processes, including implicit bias (Lebrecht et al., 2009) and eyewitness misidentification (Brigham et al., 1982; Wilson et al., 2013). Race-based eyewitness misidentification has been recognized as a problem by the legal field for many years (The Harvard Law Review Association, 1988) and remains a problem today (Flevaris & Chapman, 2015; Zalman, 2012). Over the past 50 years, researchers have examined this phenomenon by many names (e.g., own-race bias, own-race effect, cross-race effect). Following Levin (2000), we prefer the term *cross-race recognition deficit* because it is more accurate and more specific than the alternatives. Own-race effect and cross-race effect seem far too vague, as there are many different effects of race. Own-race bias is fundamentally inaccurate because the phenomenon in question reflects a change in sensitivity (not bias). Researchers have also examined the phenomenon from a variety of theoretical perspectives, using several different paradigms, trying to understand how, why, and when the CRD occurs, and how it might be mitigated (Meissner & Brigham, 2001; Young et al., 2012).

One possible strategy for reducing the CRD involves cross-race contact. Several major theories predict that cross-race contact should mitigate the CRD. The purpose of this meta-analysis is to examine the prediction that the CRD is smaller for people with more extensive cross-race contact. We first give a brief description of major theories about the CRD and why they hypothesize a relationship with contact. We then discuss various methodological factors that may affect this relationship. Finally, we present a meta-analysis of the contact–CRD relationship based on studies conducted in the past 38 years.

Contact and the CRD

There are three main hypotheses concerning the mental mechanisms that underlie the CRD: perceptual expertise, category-focused attention, and motivation. The perceptual

[1]University of Colorado Boulder, USA

Corresponding Author:
Balbir Singh, University of Colorado Boulder, Muenzinger D244, 345 UCB, Boulder, CO 80309, USA.
Email: balbir.singh@colorado.edu

expertise hypothesis suggests that recognizing cross-race faces is a skill that can be improved with practice. Perceivers who spend time with cross-race individuals may learn to more effectively extract, integrate, and encode perceptual information that allows them to individuate members of the outgroup. Accordingly, these perceivers may show a reduced CRD. This hypothesis is derived, in part, from the more general observation that experience with any class of stimuli increases expertise in differentiating exemplars from that class (Gauthier et al., 1998, 2000; Richler et al., 2011; Tanaka & Curran, 2001). People who are devoted to birding, for example, may generally be better than nonbirders at differentiating finches from sparrows. Similar patterns of expertise may develop as a function of practice with a wide variety of stimuli (e.g., cars, greebles, dogs, or faces). Expertise seems to affect both neural processing and decision-making.

In electrophysiological and imaging studies, perceivers show different neural signatures to stimuli with which they have extensive experience and expertise compared with stimuli they encounter less frequently (Gauthier et al., 2000; McGugin et al., 2014; Rossion et al., 2002; Tarr & Gauthier, 2000). When dog experts view dogs (rather than birds), and when bird experts view birds (rather than dogs), they demonstrate a more pronounced N170, an event-related potential (ERP) component that is associated with recognition expertise and integration of visual information (Tanaka & Curran, 2001). This suggests that, given practice with a class of stimuli, participants learn to integrate visual information more efficiently. Gauthier and colleagues (1999) measured the activation of brain regions involved in visual integration as they viewed novel face-like stimuli called greebles. After being trained to recognize individual greeble exemplars, participants showed more activation of brain regions associated with integrative processing to novel greebles. Again, practice viewing and individuating greebles evidently enhances participants' ability to process visual information about those stimuli.

In line with these changes to neural processes, practice also improves downstream recognition decisions (Anzures et al., 2012; Goldstein & Chance, 1985; Heron-Delaney et al., 2011; Lebrecht et al., 2009). Lebrecht and colleagues (2009) experimentally manipulated expertise by training White participants to differentiate Black or Asian faces. After five sessions, participants who were trained to individuate Black (or Asian) faces showed a reduced CRD toward novel Black (or Asian) faces. Together, these studies suggest that experience facilitates integrative processing and recognition; the more experience a person has with faces of a particular outgroup, the better their ability to recognize cross-race faces (Anzures et al., 2012; Lebrecht et al., 2009; Tanaka & Pierce, 2009). This set of literature suggests that more contact with members of a racial outgroup should reduce the CRD, particularly if that contact requires individuated processing.

The second major hypothesis concerning the mechanism of the CRD involves the idea of category-focused attention. This explanation posits a different mechanism for the CRD, but it still suggests that cross-race contact should reduce the CRD. According to this account, when participants of one race view faces of another race, their attention is misdirected toward race-specifying information within the face, rather than toward individuating information. Levin (2000) argued that race is encoded by the participant as an additional feature that draws attention and inhibits the participant's ability to individuate the face. So, although attention may be drawn to outgroup members, that attention is effectively wasted on *categorizing* outgroup members instead of *individuating* them (Correll et al., 2017). Two independent literatures in face recognition coalesce to support the attentional hypothesis. First, recognizing individual identity seems to involve different processes than recognizing the race of a face, and there may be a trade-off such that participants who are faster to categorize outgroup faces by race are slower to individuate them (Ge et al., 2009; Levin, 1996). Second, cross-race faces may attract more initial visual attention compared with same-race faces (Dickter & Bartholow, 2007; Donders et al., 2008; Guillermo & Correll, 2016; Ito & Bartholow, 2009; Ito & Urland, 2003, 2005; Lovén et al., 2012). Participants orient to cross-race faces more quickly than same-race faces, and they may fixate on cross-race faces for a longer period of time (Hirose & Hancock, 2007; Trawalter et al., 2008). Critically for the current work, attention seems to be driven by the relative novelty of racial outgroup faces (Dickter et al., 2015). Dickter and colleagues observed that as the number of cross-race friends increased, participants showed less bias in their visual attention toward cross-race faces. Cross-race contact should decrease the novelty of racial outgroups, which again suggests that contact should ameliorate the CRD.

The third major account of the CRD contends that perceivers fail to individuate cross-race faces, in part because they are not motivated to do so (Hugenberg et al., 2007, 2010; Shriver & Hugenberg, 2010). In a 2007 paper, Hugenberg and colleagues explicitly instructed participants to individuate cross-race faces and found that this manipulation decreased and even eliminated the CRD. In a similar line of argument, Bernstein and colleagues (2007) asked participants to individuate a set of *same-race* faces. Critically, some faces were labeled as members of an ingroup (students at the same school) while other faces were labeled as members of an outgroup (students at another school). The researchers observed a CRD-like effect such that ingroup faces were recognized more accurately than outgroup faces. Bernstein and colleagues suggest that cross-race faces may be coded as outgroup members decreasing motivation to individuate them, and thus lack of motivation drives the CRD. Increasing contact with racial groups should increase the personal relevance of the outgroup, and so, increase the motivation to individuate (Allport, 1954; Page-Gould et al., 2008, 2010). Again, this account seems to suggest that increasing contact should decrease the CRD.

The three major psychological explanations of the CRD thus suggest that contact with a racial outgroup should mitigate the CRD. More extensive contact should result in greater perceptual expertise with cross-race faces (e.g., Lebrecht et al., 2009). If cross-race faces are frequently encountered, they should also be less novel and therefore attract less misdirected race-focused attention (e.g., Dickter et al., 2015). Finally, outgroup contact should increase the personal relevance of cross-race faces (e.g., Young & Hugenberg, 2012), increasing perceivers' motivation to individuate them. The goal of this meta-analysis is not to determine which theory is correct in terms of the process by which contact mitigates the CRD. Rather, in light of the clear theoretical consensus, we seek to understand *when and to what extent* contact actually does mitigate the CRD.

In 2001, almost two decades ago, Meissner and Brigham conducted a groundbreaking meta-analysis on the CRD. As part of their work, the authors included 29 studies that measured both contact and the CRD. Examining the relationship, they reported a meta-analytic effect equivalent to $r = -.13$. As contact increased, participants showed better recognition of cross-race faces (a reduction in the CRD). Although this pattern is clearly in line with the theories reviewed above, all of which offer arguments for the idea that contact should moderate the CRD, this meta-analysis showed a very small effect: Meissner and Brigham's results suggest that contact explains about 1.7% of the variation in the CRD.

Part of the motivation for the current work is to clarify why, in spite of seemingly broad theoretical consensus that contact *should* influence recognition, the measured effect seems to be so small. For example, to the extent that cross-race contact was relatively rare in the later decades of the 20th century, it is possible that Meissner and Brigham (2001) underestimated the strength of the relationship. Statistical power to detect any relationship requires that there is meaningful variance on the predictor, so if contact did not vary (much) in the 1970s or 1980s, studies from that time may underestimate its effect. Alternatively or in addition, the relationship between contact and the CRD may vary as a function of the situation. Many researchers focus on the capacity of contact to promote individuated processing and interpersonal connections, but some forms of contact may have other effects. An interracial conflict, a workplace with White managers and Black workers, and a school with segregated classes may all involve nominal contact, but that contact may not reduce the CRD. The nature of the contact–CRD relationship may also depend on the methodology employed to study it (including sampling and measurement). It is possible that certain kinds of contact or certain forms of face processing exhibit stronger relationships.

In the 20 years since Meissner and Brigham's (2001) seminal paper, communities around the world have become more diverse (Organisation for Economic Co-Operation and Development & United Nations Department of Economic and Social Affairs Population Division [OECD-UNDESA], 2013), technology has increased opportunities for contact with other races, and many more studies of the CRD have been conducted. Researchers have developed new scales for measuring contact (e.g., Walker & Hewstone, 2006), new ways to measure the CRD (e.g., Michel et al., 2006), and new theoretical accounts that explain how contact reduces the CRD (e.g., Hancock & Rhodes, 2008). In light of these developments, this article reassesses the relationship between contact and the CRD, updating the average meta-analytic effect and examining a variety of potential moderators of the relationship such as sample characteristics (e.g., does the race of the participant matter?), methodology (e.g., does it matter how researchers measure contact or the CRD?), and variables predicted to matter by various theories (e.g., does contact during childhood have a larger effect?). We begin by outlining a number of variables coded in the papers that explore the contact–CRD relationship.

Moderators

Sample Characteristics

Year of publication. Meissner and Brigham (2001) reported only one moderator of the relationship between contact and the CRD: year of publication. They found that the magnitude of the contact–CRD relationship increased over time. They postulated that this change may have been due to historical events like increasing integration in the United States and/or improvements in methodology. Over the past 20 years, integration has continued. In the United States and across the world, migration and differential birth rates have dramatically increased opportunities for cross-race interaction. Although many communities are still heavily segregated, the trend favors integration with increased access and exposure to members of other races (Frey, 2018; U.S. Census Bureau, 2018). These opportunities may promote variation in cross-race contact, increasing the likelihood of detecting a relationship between contact and the CRD. During this time, methodology has also continued to change (we will consider several methodological factors, over and above year of publication). If the trend Meissner and Brigham (2001) observed continues, we should see that the relationship between contact and the CRD becomes stronger over time.

Race of participants. The races of the participants analyzed in any given sample may also moderate the contact–CRD relationship. The CRD has been observed in many different countries using participants of many different races (e.g., Sangrigoli et al., 2005; Wright et al., 2001), suggesting that the CRD may be a universal phenomenon. But there is also some reason to expect differences as a function of perceiver race. Many researchers (Hancock & Rhodes, 2008; Hayward et al., 2013; Michel et al., 2006; Tanaka et al., 2004) argue that the CRD occurs due to a lack of holistic or configural processing toward cross-race faces (further discussion of holistic and configural processing is presented below). For example, White participants would be more likely to process White faces holistically compared with Asian faces (Michel et al., 2006). However, a recent review (Hayward et al., 2013) argues that Asian participants may process faces differently compared with other races, and some studies (Crookes et al., 2013; Michel et al., 2006) have found that Asian participants process both same-race and cross-race faces holistically. This potential difference in processing styles may affect the magnitude of the contact–CRD relationship for Asian participants compared with participants of other races. Most of the samples in our data set use White participants, Asian participants, or Black participants, and we code this as a potential moderator (several studies used participants of other races, but these were singular in their usage so we could not analyze them).

Race of faces. In addition to sampling participants, researchers must sample stimuli. We therefore consider the impact of the race of the faces used to measure the CRD as an aspect of sample characteristics. Almost every sample (188 out of 207) in this data set uses White faces. Some samples compare performance with White faces to performance with Black faces, and some compare performance with White faces to performance with Asian faces. In a few cases, a group other than White, Black, or Asian is used (e.g., Canadian Indigenous or Turkish), but the White–Black and White–Asian comparisons seem to dominate work in this area (142 samples). We test whether the magnitude of the contact–CRD relationship depends on the race of the faces used to assess the CRD.

Methodological Factors

Over the past 20 years, researchers have developed new scales and new measures of face recognition that explore different types of face processing. These changes in methodology may have implications for the magnitude of the contact–CRD relationship. We consider these methodological moderators next.

CRD measurement. The CRD is measured with a number of tasks. The most common paradigm is the encode-recognition task. In an encoding phase, this task asks participants to view and remember a number of faces (typically between 10 and 40 faces). In a subsequent recognition phase, the procedure tests the participant using a set of faces that includes both the learned faces and a number of lures (see Figure 1). A variety of other tasks have also been used to measure the CRD (see Figure 1 for three examples). On each trial of the same–different task, for example, participants view one face, which they are expected to remember. This face is followed quickly by a second face. Participants indicate whether the second face is the same or different from the first face. The forced-choice task is similar to the same–different task in that it presents a single face at encoding, followed immediately by a test trial. But the forced-choice task then presents an array of perhaps five faces from which the participant must select the to-be-remembered face. Experiments that use the same–different or forced-choice tasks typically utilize numerous trials, but each trial requires the participant to hold only a single face in memory, and the retention period is fairly short, lasting only a second or two. From these three tasks, we can illustrate some key differences in CRD measurement that may impact the contact–CRD relationship.

There is recent evidence that the CRD occurs during encoding rather than during retrieval (Hughes et al., 2019; Stelter & Degner, 2018). If we assume that the CRD occurs entirely at encoding (i.e., storage and retrieval play no role), then any source of variation in performance occurring after encoding may obscure the contact–CRD relationship. CRD paradigms vary in memory demands. In a same–different or forced-choice task, participants are asked to hold faces in memory for a few seconds. In a typical encode-recognition task, participants are asked to hold faces for minutes or hours. Similarly, same–different and forced-choice tasks present a single face to encode on each trial. Encode-recognition tasks may ask participants to encode 40 faces or more (e.g., Maclin et al., 2004; Tullis et al., 2014) before they are tested. As set size and retention interval increase, so do the demands on memory (Cowan, 2008), which may reduce the contact–CRD relationship.

Finally, tasks may differ in terms of the number of faces present during a recognition trial. The same–different and encode-recognition tasks ask participants to make recognition decisions about a single face at a time. By contrast, the forced-choice task presents multiple faces during the recognition test and asks participants to decide whether one of the presented faces is the face they recognize. In social-cognitive style laboratory studies, the forced-choice task may present anywhere from two (e.g., De Heering et al., 2010) to six faces during a trial of the recognition test (e.g., Jackiw et al., 2008). In eyewitness studies, participants often choose from a much larger set of faces, perhaps 12 or more (Brigham et al., 1982). The presence of multiple faces at recognition has been shown to reduce recognition ability (R. C. Lindsay & Wells, 1985), potentially adding noise to the measurement and altering the contact–CRD relationship.

Studies that minimize demands unrelated to face processing may provide a cleaner measurement of the CRD. Accordingly, we predict that studies with shorter retention times, smaller encoding sets, and smaller recognition sets will have the largest contact–CRD effects.

Operationalization of cross-race contact. Just as there are a range of tasks that researchers use to measure the CRD, there is also variability in how researchers operationalize cross-race contact. Contact has most often been conceptualized in terms of the amount of exposure to members of other racial groups (e.g., Hancock & Rhodes, 2008), although other researchers have emphasized the importance of the valence

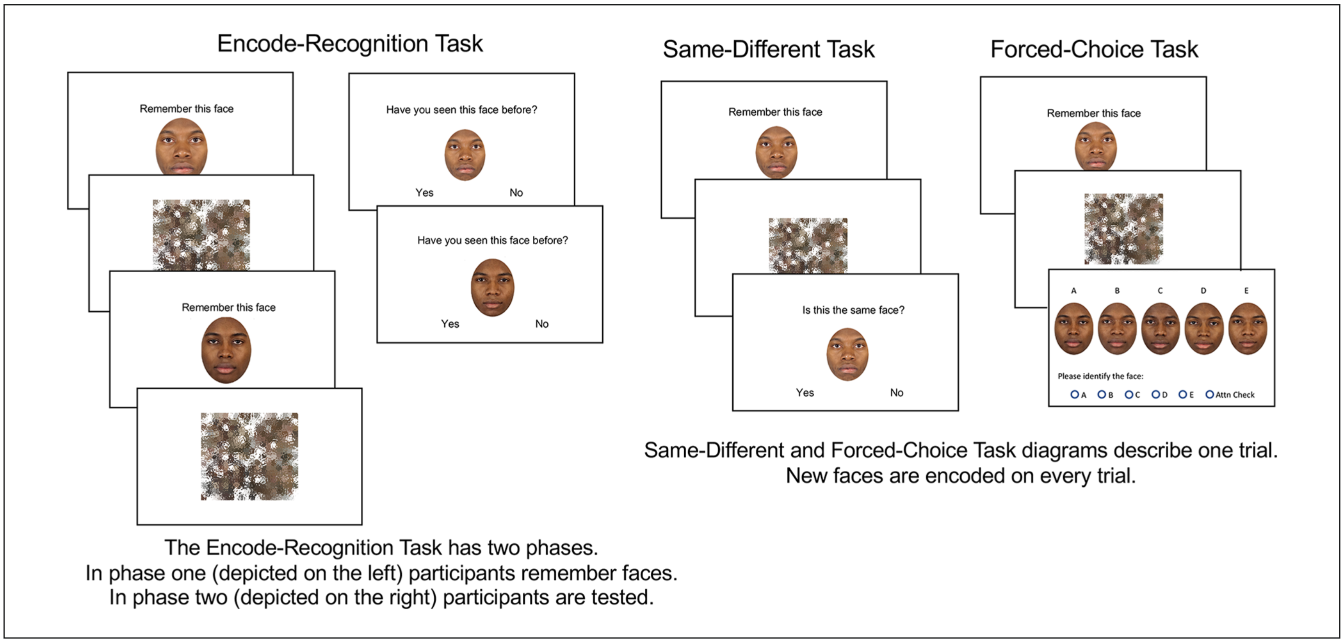

Figure 1. Depiction of the three most common CRD tasks.
Note. CRD = cross-race recognition deficit.

or quality of cross-race contacts (Slone et al., 2000; Walker & Hewstone, 2006). Methodologically, contact is typically operationalized in one of three ways: experimental manipulations, life experience, and self-report measures. Experimental manipulations involve training participants by providing individuated exposure (which we will treat as a form of contact) to cross-race faces (Anzures et al., 2012; Goldstein & Chance, 1985; Heron-Delaney et al., 2011; Lebrecht et al., 2009; Tanaka & Pierce, 2009). These manipulations typically increase the amount of experience a person has with cross-race faces and thus represent an increase in sheer quantity, but they also typically avoid confounding contact with affective dimensions: There is no friendship and no enmity. Other researchers have used a variety of life experience measures as a stand-in for contact (e.g., Sangrigoli et al., 2005). For example, researchers might measure the number of years that participants lived in a country or region predominantly populated by a group of people that was different from the participant's own race (e.g., the number of years a Korean person has spent in France) or compare White participants who live in majority White England with White participants who live in majority Black South Africa (Wright et al., 2001). Although this kind of measure may be a proxy for contact, it is likely a poor one. For example, the number of years that a participant lived in a foreign country is not informative about the number or quality of cross-race relationships the participant may have formed or how much time a participant has spent with outgroup members.

The most common strategy for assessing contact relies on self-report measures. A number of researchers have developed scales for this purpose (e.g., M. M. Davis et al., 2016; Hancock & Rhodes, 2008; Ng & Lindsay, 1994; Slone et al., 2000; Sporer et al., 2007; Walker & Hewstone, 2006). There is a great deal of diversity among these scales in terms of content and psychometrics. Some assess sheer frequency or quantity of contact, and others assess quality and intimacy (Mellinger et al., 2021). In some cases, contact is measured by a single item, in some cases by a multi-item scale designed to assess a single construct, or even by a multi-item scale designed to assess multiple constructs. As the bulk of the samples included in our analysis used scales to measure cross-race contact, we examine whether the most commonly used scales (Hancock & Rhodes, 2008; Slone et al., 2000; Walker & Hewstone, 2006) differ in the magnitude of the contact–CRD relationship.

We predict that manipulating contact will be an effective way to measure the contact–CRD relationship. In addition, commonly used scales (Hancock & Rhodes, 2008; Slone et al., 2000; Walker & Hewstone, 2006) should perform better than idiosyncratic scales. Finally, we predict that quasi-experimental methods (e.g., how long a person has lived in a country) will be a poor measure of contact and thus will weaken the contact–CRD effect.

Theoretically Important Moderators

The CRD is a phenomenon that exists at the interface of neuroscience and cognitive, social, and developmental psychology. It has implications for a number of different theories, and in at least two cases, our meta-analysis can explore questions that have been debated in recent work.

Age of cross-race contact. The time period during which cross-race contact occurs may moderate the contact–CRD relationship. Literature suggests that there is a critical window in processes like language acquisition (Lenneberg, 1967), and some researchers have argued that the effects of contact on face processing may similarly have a critical or sensitive period (McKone et al., 2019; Pascalis et al., 2020; Sangrigoli et al., 2005). The critical window hypothesis suggests that cross-race contact during childhood should result in a larger reduction in the CRD compared with cross-race contact during adulthood. As such, we code the time period during which contact occurred for each sample to examine how this moderator may impact the magnitude of the contact–CRD relationship.

Face processing. Researchers argue that same-race faces are typically processed in a more configural or holistic fashion than are cross-race faces (Crookes et al., 2013; Michel et al., 2006; Richler et al., 2011; Tanaka et al., 2004). As defined by Maurer and colleagues (2002), holistic processing refers to processing the various features of a face in a gestalt fashion, whereas configural processing refers to sensitivity to spatial relationships among features of the face (e.g., relative distance between the eyes). Although configural and holistic processing differ, they both describe a type of integrative processing (Hayward et al., 2013). Several studies have reported that people are less likely to demonstrate holistic or configural processing for cross-race faces, and several researchers have argued that cross-race contact may have effects that are specific to these kinds of integrative processing. For example, Rhodes and colleagues (1989, 2006, 2008) argue that cross-race contact increases the likelihood that people process cross-race faces in a configural manner. In our sample of studies, 14 use tasks designed to specifically measure holistic processing, and an additional two samples use tasks designed to measure configural processing. Although we do not have enough samples to test configural processing separately, we can compare samples that test either configural or holistic processing (combined) with samples that test "normal" processing. Our goal is to test the possibility that cross-race contact has a larger impact on more integrative processing of cross-race faces. Some research (Hancock & Rhodes, 2008; Rhodes et al., 1989, 2006) suggests that relative to normal processing, studies that assess holistic/configural processing will show evidence of a stronger contact–CRD relationship.[1]

Summary

Given new studies, new methods, and new theoretical questions, a new meta-analysis has the potential to shed new light on the nature of the contact–CRD relationship. First, additional data from the new studies can shed light on the overall magnitude of this relationship. Second, given variability in the procedures, samples, and analytical approaches used during the last 20 years, we can examine moderators of the relationship in ways that were not feasible for Meissner and Brigham (2001).

Method

On February 8, 2018, we entered the following search terms into Google Scholar, "other race effect" or "own race bias" or "face recognition" and contact and expertise. On July 26, 2020, on the recommendation of our reviewers, we repeated the search using Publish or Perish (Harzing, 2007), which allows a user to enter search terms into different databases and save the results. We repeated this second search once more on October 31, 2020, in an attempt to capture every paper that met our criteria (described below). We entered the following search terms in Google Scholar, PsycINFO, Web of Science, and Crossref: "other race effect" or "own race bias" or "face recognition" or "face identification" or "eyewitness identification" or "cross-race effect" and contact and expertise. Between the four databases, we found 3,340 papers. A paper was eligible for inclusion in the analysis if it (a) measured or manipulated cross-race contact, (b) measured recognition of cross-race faces, and (c) tested the relationship between contact and recognition. At least one member of the research team read each of the 3,340 abstracts to determine whether or not the paper might be relevant. Any paper that seemed potentially relevant was reviewed by at least two members of the research team (at least one of whom was a graduate student, postdoctoral fellow, or faculty member). At least two coders (at least one of whom was a graduate student, postdoctoral fellow, or faculty member) independently evaluated each paper and coded it for the final data set, resolving disagreements on a case-by-case basis.

In several cases, it was clear that researchers had collected the requisite data but either did not analyze the association between contact and recognition or did not report the association in a manner that allowed us to use the data. For example, several papers reported only mean levels of contact to suggest that cross-race contact was minimal. These papers clearly measured contact and the CRD, but they did not report the relationship between contact and the CRD. Other papers reported that a relationship between contact and recognition was not significant but did not report the test statistic. These cases were noted, and a member of the research team wrote to the corresponding author of each paper to request supplemental analyses or data. We sent 28 emails and received 10 responses, five of which yielded additional data.

On December 5, 2018, we sent an email to the Society for Personality and Social Psychology listserv asking for any studies, published or unpublished, that met the inclusion criteria. We received three replies, of which we were able to include one.

In the end, we obtained usable data from 207 statistical tests in 96 independent samples from 65 papers. Going forward, we will use the term *sample* to refer to an individual statistical test. We have made our data set, analysis code, and output available on OSF (see https://osf.io/avh3x/). We estimated our power via a sensitivity analysis (see the Supplemental Material for code). Power curves were generated via

Monte Carlo simulation with 1,000 meta-analysis simulations per effect size. Results were extrapolated to test our sensitivity at 80% power in each case. We were powered at 80% to detect a correlation of $r = -.074$ between the CRD and cross-race contact (ignoring moderating variables). As described below, we also tested several categorical moderators. To test two-group moderators with equal numbers of samples in each group, our data set achieved 80% power to detect a difference in correlations between the groups of $r_{\text{diff}} = -.101$. We also tested moderators with unbalanced cell sizes, that is, having many more samples in one category than the other. The worst of these, described below, was the moderator concerning the operationalization of contact where nine samples represented the primary group of interest and the other group was comprised of 198 samples. In this situation, our data set resulted in 80% power to detect a correlational difference of $r_{\text{diff}} = -.337$.

Effect Sizes for the Contact–CRD Relationship

The CRD is defined as more accurate recognition of same-race faces compared with cross-race faces. All of the samples we included used some continuous measure of recognition accuracy. Most papers computed the CRD as a difference in accuracy to same-race versus cross-race faces, although some papers only reported accuracy to cross-race faces, and others used alternatives to a difference score approach (e.g., statistically controlling for performance with same-race faces). Due to this variability, we coded the paper's method for computing the CRD (or recognition of cross-race faces) and included this variable as a moderator in our analysis.

The outcome of interest was the relationship between cross-race contact and the CRD (or recognition cross-race of cross-race faces). These relationships were converted to the effect size r. To analyze the effect sizes, each r coefficient was Fisher's Z transformed, and the direction of the relationships was coded so that negative values indicate that increased contact was associated with a decrease in the CRD. That is, when higher cross-race contact was associated with a lower CRD, as predicted by most theories, the z-statistic had a negative sign. Following Borenstein and colleagues (2009), variances for the z-statistics were calculated as $V_z = 1/(n-3)$. The z-values were converted back to the r metric for interpretation in the "Results" section.

Moderators

As noted above, we explored a number of variables as potential moderators of the relationship between contact and recognition (see Table 1 for the list of moderators and their hypotheses). Here, we provide details about the coding of each of those variables. In our analyses, we used contrast codes for all categorical moderators and mean centered continuous variables.

Sample characteristics

Year of publication. If a paper had been published, we recorded the year of publication. For unpublished work, such as dissertations, we recorded the year listed on the document.

Participant race. Most samples used White participants ($k = 108$). The other two most common participant races were Asian ($k = 40$) and Black ($k = 23$). These three races were coded as separate groups in the data set. Twenty samples used participants that were Latino, Multiracial, or Turkish. Fifteen samples included participants who were White, Black, or Asian but collapsed the results across race. One additional sample collapsed the results across White and Canadian Indigenous participants. The authors of these studies assessed the contact–CRD relationship by comparing same-race faces with cross-race faces. As the researchers did not report the effect for each participant race individually, we could not reasonably separate any effects due to race in these samples. We therefore collapsed these 16 samples along with the 20 non-White, Black, or Asian as an "other" category.

Races of faces. Most samples ($k = 142$) in our data set used Black or Asian faces in addition to White faces. Of those samples, 75 used White and Asian faces, whereas 67 used White and Black faces. We coded any sample that did not cleanly fit into these two categories as "other" ($k = 65$).

Methodological factors

CRD measurement. We coded three variables to capture variation in the nature of the CRD measures: the length of time a face is retained, the number of faces participants studied during encoding, and the number of faces in a single recognition trial. As discussed in the introduction, length of face retention and number of encoding faces alter memory demands for participants. In addition, the number of faces at recognition has been shown to affect recognition ability (R. C. Lindsay & Wells, 1985; Makovski et al., 2010), so we tested whether decreasing recognition ability would alter the contact–CRD relationship.

Length of retention. Our first variable of interest was the length of time participants retained faces. Researchers did not generally report precise timing parameters for retention intervals. We were able to characterize each study as requiring participants to retain face(s) in memory for only a few seconds ("short"; for example, the same–different task, $k = 49$), or for minutes to hours ("long," $k = 158$), but extracting a continuous measure proved impossible. As a result, this moderator was necessarily categorical. Often, researchers had participants wait 5 to 10 min prior to the recognition trials (e.g., Young & Hugenberg, 2012), but some researchers asked participants to hold faces in memory for hours (Brigham et al., 1982) or even days (Berger, 1969).

Table 1. Table of Samples Included With a Description of the Hypothesis, and Our Results.

Moderator	Hypothesis based on prior literature	Confirmed	Explanation
Year	The contact–CRD relationship should strengthen over time.	X	The contact–CRD effect weakens with time.
Race of the participants	The race of the participant should not affect the contact–CRD relationship.	✓	There are no differences as a function of the race of participant.
	The contact–CRD relationship should be weaker for Asian participants compared with White participants.	X	There is no difference in the contact–CRD relationship between Asian and White participants.
Races of faces	The races of the faces used to measure the CRD should not affect the contact–CRD relationship.	✓	There are no differences as a function of the races of faces used.
CRD measurement	Retaining a face(s) for more than a few seconds should weaken the contact–CRD relationship.	X	There are no differences as a function of how long a face is retained.
	Retaining more than one face should weaken the contact–CRD relationship.	X	There are no differences as a function how many faces are retained.
	Tasks that ask participants to retain more than one face for a long period of time will weaken the contact–CRD relationship.	X	There is a nonsignificant hint that suggests tasks that ask participants to retain more than one face for a long period of time weakens the contact–CRD relationship. More research is needed.
	Recognition trials with more than one face should weaken the contact–CRD relationship.	X	There are no differences as a function of how many faces are on a given recognition trial.
Operationalization of contact	Manipulating contact compared with other operationalizations should strengthen the contact–CRD relationship.	✓	Manipulating contact reveals a stronger contact–CRD effect compared with all other operationalizations.
	Scales (as a set) compared with other operationalization should weaken the contact–CRD relationship.	✓	Scales as a set reveal weaker contact–CRD effect compared with other operationalizations.
	Idiosyncratic scales as compared with commonly used scales should weaken the contact–CRD relationship.	X	There is no difference as a function of the scale used.
	Quasi-experimental contact compared with other operationalizations should weaken the contact–CRD relationship.	X	There is no difference between quasi-experimental contact and scales. Quasi-experimental contact compared with manipulated contact marginally weakens the contact–CRD relationship.
Age of contact	Childhood contact should be more effective than adult contact		Contact assessed prior to 18 strengthens the contact–CRD relationship. There is a nonsignificant hint that contact assessed prior to 12 strengthens the contact–CRD relationship. More research is needed.
	Under 18 versus over 18	✓	
	Under 12 versus over 12	X	
Face processing	Integrative processing should strengthen the relationship between contact and the CRD.	X	There is a nonsignificant hint that integrative processing strengthens the relationships between contact and the CRD. More research is needed.

Note. CRD = cross-race recognition deficit.

Number of faces encoded. The second variable of interest was the number of faces a participant encoded at a time. Although this variable could be considered a continuous indicator, its distribution was highly skewed. We present the effect of this moderator as both a continuous indicator and collapsed into two groups: Samples that presented one face at a time were categorized as "one," and samples that presented more than one face at a time were categorized as "more than one." Thirty samples (k = 30) presented only one face at encoding while the rest (k = 177) presented more than one face (M = 27.51, SD = 18.49, minimum = 2, maximum = 80).

Number of recognition faces. Finally, we coded the number of faces that were present on an individual recognition trial during the task. Again, this variable could be considered a continuous indicator, but its distribution is even more skewed than the number of faces encoded. Most researchers present one face during the recognition trial (e.g., in the encode-recognition or same–different tasks, $k = 157$) while others ($k = 50$) present more than one face at a time (e.g., in the forced-choice task, $M = 3.74$, $SD = 2.25$, minimum = 2, maximum = 12).

Operationalization of cross-race contact. The vast majority of studies used a scale to measure contact, and we distinguished between the three most commonly used scales (Hancock & Rhodes, $k = 20$; Walker & Hewstone, $k = 44$; Social Experience Questionnaire, $k = 60$). More unusual scales ($k = 55$) were identified as "idiosyncratic." We also identified samples that manipulated contact through their experimental procedure via training with cross-race faces ($k = 9$, for example, Lebrecht et al., 2009). Finally, we identified 19 samples that used a quasi-experimental, known-groups approach. For example, Wright and colleagues (2001) recruited White participants from England and South Africa, who were assumed to vary in terms of contact with Black people. We coded the samples into six categories: measured through Walker and Hewstone, measured through Hancock and Rhodes, measured through the Social Experience Questionnaire, idiosyncratic scales, manipulated contact, or quasi-experimental.

Theoretically important moderators

Age of cross-race contact. Multiple researchers have argued that the contact–CRD relationship should be more pronounced if the cross-race contact occurs during childhood (e.g., M. M. Davis et al., 2016; McKone et al., 2019; Pascalis et al., 2020; Sangrigoli et al., 2005). McKone and colleagues as well as Pascalis and colleagues suggest that contact is most sensitive before the age of 12. McKone and colleagues further separate childhood into the years prior to schooling (ages 0–5), primary school (5–12), and secondary school (12–18). Davis and colleagues also separate childhood into three categories, ages 0–6, 6–12, and 12–18. Following these researchers, we categorized samples into five groups. The first four groups included samples for which contact occurred early in childhood prior to schooling (under the age of 6, $k = 7$), during primary or elementary school (ages 6–12, $k = 18$), during secondary school (in the United States, this is often the time period of middle and high school; ages 12–18, $k = 16$), or after age 18 ($k = 111$). Fifty-five samples collapsed contact across multiple age groups. These 55 tests were grouped in the "lifetime contact" category.

Face processing. Most samples ($k = 186$) in our data set involved tasks that presented normal, upright faces at encoding and at recognition. We designated these cases as "normal" processing. Sixteen samples utilized tasks that measured the extent of participants' holistic or configural processing for same- and cross-race faces (e.g., using inverted faces, as in Hancock & Rhodes, 2008). There were five samples that did not fall cleanly into either category, which were coded as "other."[2]

Calculation of the CRD. Finally, we coded the article's method for computing the CRD. Most samples in our data set calculated the CRD as cross-race recognition accuracy subtracted from same-race recognition accuracy. This metric represents the extent to which participants were more accurate to faces of their own race than to faces of another race ($k = 109$). Rather than calculating a difference score, some samples ($k = 3$) partialed out the performance on same-race faces by including the same-race recognition score as a covariate in the model. Other samples ($k = 95$) reported the relationship between cross-race contact and recognition accuracy for cross-race faces without adjusting for recognition ability for same-race faces. In our view, using a scoring strategy that does not account for the relative nature of the CRD is not optimal. Rather than excluding these samples from our data set, we included the calculation of the CRD as a moderator. As difference scores and partialed scores are relative outcomes, we coded these samples as relative ($k = 112$) and distinguished them from nonrelative outcomes ($k = 95$). This approach allowed us to test whether the magnitude of the contact–CRD relationship varied as a function of the reported outcome and also to control for any difference when examining other moderators.

Analytic Strategy

Our analysis is based on a random-effects multilevel model. Two considerations informed this choice. First, the random-effects model acknowledges that different studies may have different error terms relative to the true effect. This approach also acknowledges that moderators may not explain all of the variance between studies. This model is generally considered to be more appropriate for common meta-analyses (Pastor & Lazowski, 2018).

Second, some studies in our data set included multiple effect size estimates based on the same sample. For example, Wiese et al. (2014) reported two correlations from the same sample based on different measures of contact. Our multilevel model accounts for the dependence of these effect sizes while allowing us to retain all of the data to inform the final effect size estimate. We utilized a three-level structure that allows for variance due to sampling error at Level 1, variance between dependent effect sizes at Level 2 (within sample), and variance between independent effect sizes at Level 3 (between sample; Assink & Wibbelink, 2016).

Results

Our primary question involved the magnitude of the relationship between contact and the CRD. We then addressed publication bias or small sample effects using a trim-and-fill method. Following these two analyses, we evaluated each moderator on its own to understand its effect on the relationship between contact and the CRD. Finally, we evaluated a combined model, including all of the moderators as simultaneous predictors. See OSF (https://osf.io/avh3x/) for the complete data and analysis code. Our effect sizes are reported as r statistics throughout. Simple effects are reported simply as r, whereas the difference in effect sizes between two groups, such as the difference between child and adult contact, is reported as r_{diff} for categorical moderators or r_{slope} for continuous moderators. All models were estimated with the metafor package in R (R Core Team, 2019; Viechtbauer, 2010).

Effect Size for the Contact–CRD Relationship

In the simplest model, without controlling for moderators, higher levels of contact were associated with a reduction in the CRD, $r = -.148$, 95% confidence interval (CI) = [−.198, −.096], 95% prediction interval (PI) = [−.468, .208], $F(1, 206) = 31.58, p < .001$. See Figure 2 for a forest plot. Of total variability, a large proportion was attributable to real heterogeneity in effects, $I^2_{\text{overall}} = 66.34,\ \tau = .033$. The vast majority of heterogeneity was due to between-sample variance, $I^2_{\text{between}} = 66.29, \tau_{\text{between}} = 0.032$, with only a tiny proportion attributable to within-sample variation, $I^2_{\text{within}} = 0.045,\ \tau_{\text{within}} < .001$. There was evidence of residual variability in the effect sizes between samples in our data set, $Q_E(206) = 419.08, p < .001$.

While the significance of this first test suggests an observable effect, it is important to consider possible bias in its magnitude. Well-known phenomena, such as the file-drawer issue (Rosenthal, 1979), may have restricted our access to nonsignificant samples, systematically inflating our estimate. Similarly, we were not able to include some samples that reported a nonsignificant relationship between contact and the CRD but failed to report the actual estimate of the relationship (or respond to our email requests). Accordingly, our raw estimate may be inflated. To examine this issue, we first examined the relationship between published and nonpublished papers in our data set. Samples from published studies showed a significant contact–CRD relationship, $r = -.174$, $95\%\ \text{CI} = [-.229, -.116], F(1, 205) = 34.79,\ p < .001$, but samples from unpublished reports did not, $r = -.056$, $95\%\,\text{CI} = [-.162, .052],\ F(1, 205) = 1.05,\ p = .306$. The difference between these estimates was marginally significant, $r_{\text{diff}} = -.119,\ 95\%\,\text{CI} = [-.237, .003],\ F(1, 205) = 3.68$, $p = .056$. It is possible that null effects are not being published due to the file-drawer problem.

To further examine the potential file-drawer problem, we generated a funnel plot, which is presented in Panel A of Figure 3. The figure shows study precision (standard error) as a function of observed effect size. Given a truly random sample of all studies ever conducted, a symmetric pattern should emerge around the average effect, with high precision studies more tightly clustered than low precision studies. Visual examination of the plot suggests some evidence of publication bias. More samples fall to the left of the estimated average effect size than to the right (note that our effect is negative, so the left-hand side of the axis represents effects with larger magnitudes), especially for samples with low precision. This pattern suggests a relationship between study precision and effect magnitude in our data set that would not be present if we had randomly sampled from the population of studies. Some small-effect samples, presumably unpublished, may have been missed by our search. The

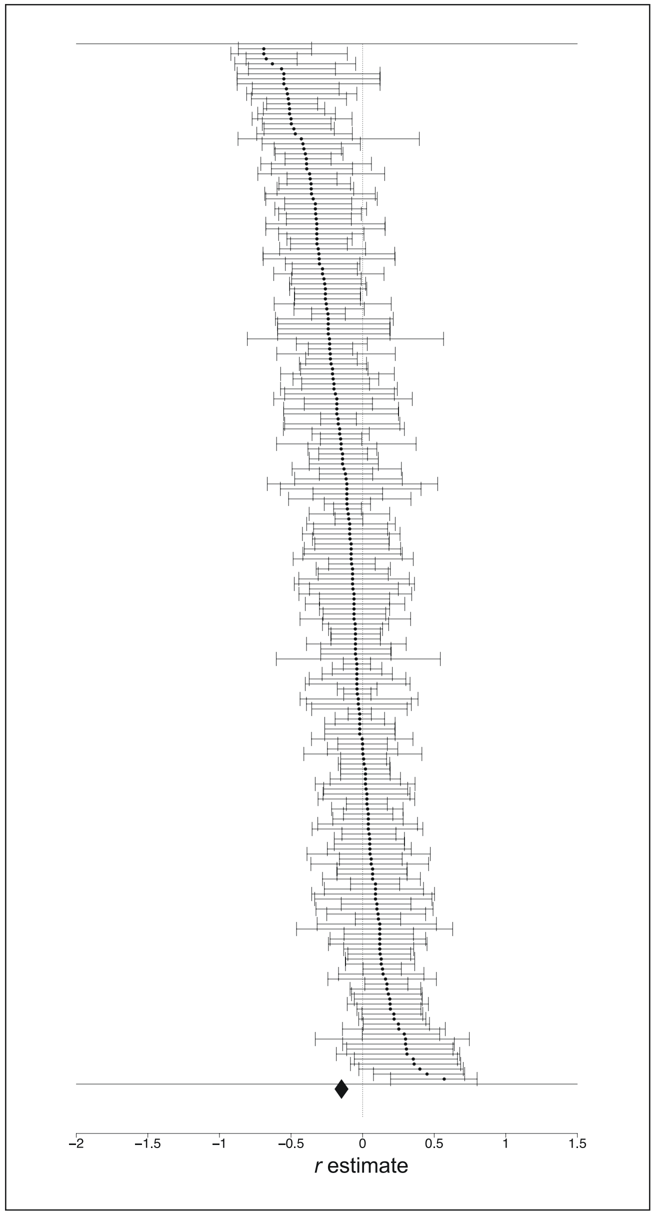

Figure 2. Forest plot derived from the unmoderated model.
Note. Effect sizes are represented as r values. A forest plot with labels is available in the Supplemental Materials.

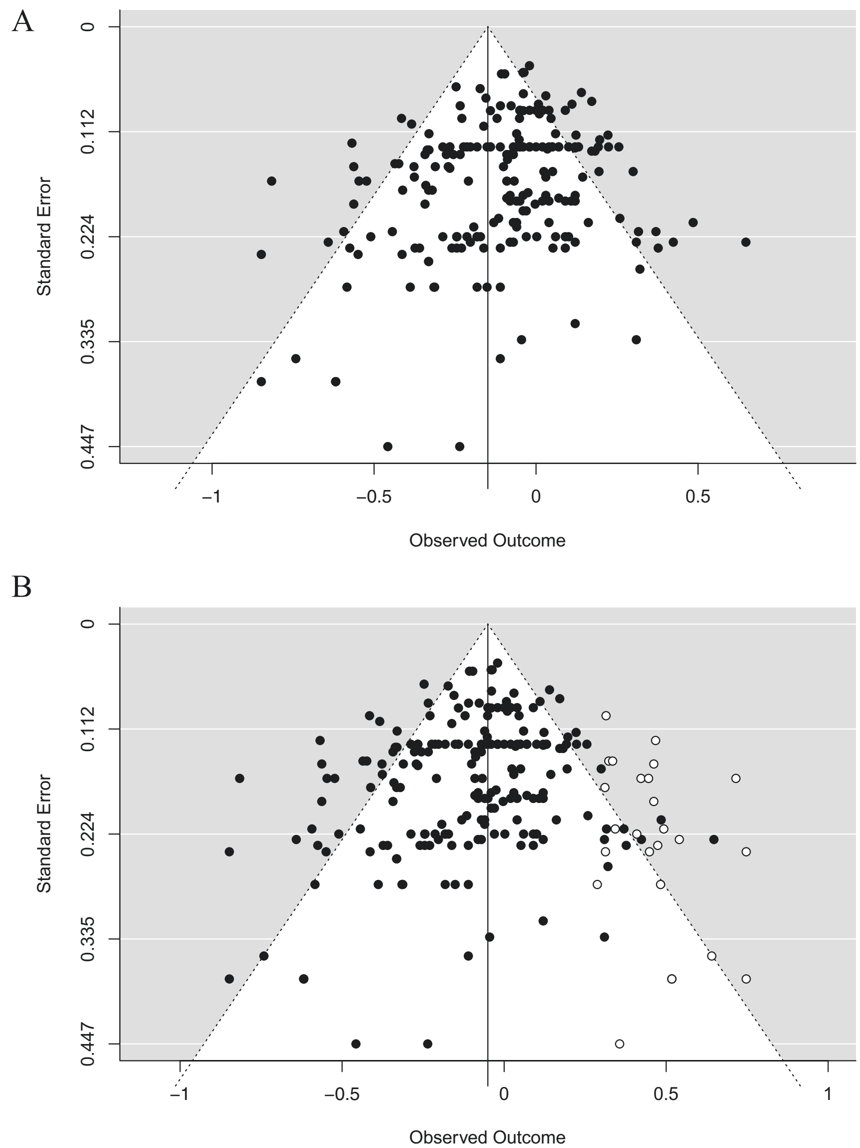

Figure 3. (A) Funnel plot of original data and (B) filled funnel plot from Duval and Tweedie's (2000) trim-and-fill method.

trim-and-fill method suggested by Duval and Tweedie (2000) is one way to estimate the number of missing samples and the resulting bias in effect size. Using this method, samples in the funnel plot causing asymmetry are trimmed from the data, and an adjusted effect size is estimated from the reduced set of samples. The adjusted effect size is considered as the true center, and samples are filled around it to create a symmetric funnel plot to recover the original sample size. Finally, the effect size is re-estimated from the filled data set (Schwarzer et al., 2015). Using the trim-and-fill method, it was estimated that 26 samples may have been missing from our analysis. The filled funnel plot is shown in Panel B of Figure 3. The effect size from the filled data is still significantly different from 0 and still suggests that contact reduces the CRD, but it is markedly smaller than the initial estimate, $r = -.050, 95\%\ \text{CI} = [-.082, -.018], z = 3.03, p = .002$.

Finally, we considered differences between samples that reported zero-order correlations (or the equivalent) of the contact–CRD relationship and those that reported partial relationships. In our data set, we have seven samples that statistically controlled for variables such as age and race categorization when assessing the contact–CRD relationship. These covariates could artificially increase the meta-analytic estimate, although the impact of these covariates should be minimal given that the proportion of partial-effect estimates is rather low (seven of 207 samples). We found no evidence that partial effects are inflating the estimate, $r_{\text{diff}} = .083$, $95\%\text{CI}=[-.125, .284]$, $F(1, 205) = 0.62$, $p = .434$.

Table 2. Simple Effects of the Categorical Moderators.

Moderator	Level of moderator	r	Lower bound CI	Upper bound CI	p
Race of participants $df = 203$	White	−.132	−.188	−.075	<.001
	Black	−.182	−.286	−.073	.001
	Asian	−.092	−.183	−.000	.050
	Other	−.193	−.271	−.112	<.001
Races of faces $df = 204$	White & Black	−.152	−.223	−.079	<.001
	White & Asian	−.173	−.243	−.102	<.001
	Other	−.081	−.178	+.017	.104
Measurement of the CRD $df = 205$	Retain faces for a few seconds	−.152	−.235	−.066	<.001
	Retain faces for minutes/hours	−.146	−.204	−.088	<.001
	Encode a single face	−.162	−.259	−.061	.002
	Encode many faces	−.145	−.199	−.089	<.001
	Recognize one face	−.163	−.222	−.102	<.001
	Recognize multiple faces	−.117	−.199	−.033	.007
Operationalization of contact $df = 201$	Social Experience Questionnaire	−.076	−.196	+.046	.219
	Walker & Hewstone Scale	−.167	−.316	−.010	.038
	Hancock & Rhodes Scale	−.137	−.240	−.030	.013
	Idiosyncratic scales	−.126	−.203	−.047	.002
	Quasi-experimental	−.185	−.278	−.088	<.001
	Manipulated contact	−.382	−.561	−.170	<.001
Age of contact $df = 202$	Early childhood (ages 0–6)	−.335	−.532	−.104	.005
	Primary school (ages 6–12)	−.253	−.348	−.152	<.001
	Secondary school (ages 12–18)	−.257	−.350	−.159	<.001
	Adult contact (ages 18+)	−.128	−.190	−.064	<.001
	Lifetime contact	−.124	−.191	−.055	<.001
Face processing $df = 204$	Normal processing	−.144	−.198	−.089	<.001
	Holistic/configural processing	−.215	−.341	−.081	.002
	Other processing	−.062	−.235	+.115	.490
Calculation of the CRD $df = 205$	Relative scores	−.174	−.235	−.111	<.001
	Nonrelative scores	−.108	−.181	−.034	.005

Note. CI = confidence interval; CRD = cross-race recognition deficit.

In the following section, we consider moderators of the effect using our entire data set.

Moderators

As planned, moderator analyses were conducted. We first analyzed each moderator individually and explored relevant simple effects, which are reported in Table 2. Finally, we ran a model including all of the moderators as simultaneous predictors. Each step is presented below.

Sample characteristics

Year of publication. Meissner and Brigham (2001) found that the contact–CRD effect got stronger for more recent studies. We tested for the linear and quadratic effects of year. As a set, these two variables were marginally related to the contact–CRD effect, $F(2,204) = 2.34, p = .099$. This marginal effect is driven by a significant *positive* relationship between year and the contact–CRD relationship, $r_{slope} = .007$, 95% CI = [.000,.014], $F(1, 204) = 4.309, p = .039$. Contrary to the hypothesized effect, the contact–CRD effect was weaker in later years.

Race of participants. We entered three contrasts coding four participant-race groups: Black, Asian, White, and other. There was no evidence that the contact–CRD relationship depended on participant race, $F(3, 203) = 1.373, p = .252$. We examined the difference between White participants and those that are Black or Asian (combined) and found no difference, $r_{diff} = 0.006$, 95% CI = [−.069,.080], $F(1, 203) = 0.02, p = .884$. Samples of Black participants did not differ from samples of Asian participants, $r_{diff} = -.091$, 95% CI = [−.220,.041], $F(1, 203) = 1.86, p = .174$. Hayward and colleagues (2013) argue that Asians process faces differently than people of other races, specifically Whites. We found no evidence for a difference in the contact–CRD relationship for Asian versus White participants, $r_{diff} = .040$, 95% CI = [−.049,.129], $F(1, 203) = .790$, $p = .375$, nor for Asian versus all other participants in our data set, $r_{diff} = .009$, 95% CI = [−.051, .223], $F(1, 203) = 1.57, p = .212$. For participants in each racial group, the contact–CRD relationship was significant (see Table 2 for simple effects). Although there are no significant differences, in line with Hayward and colleagues' (2013) predictions, the effect is directionally smallest among Asian participants.

Races of faces. In any investigation of the relationship between contact and the CRD, researchers must decide what races of faces to use to measure the CRD. By necessity, the same-race faces need to coincide with the race of the participant, which for the bulk of the samples in our data set was White. The choice of the cross-race face, however, is more flexible. We found no overall difference as a function of the pairs of faces that researchers tested, $F(2, 204) = 1.50, p = .226$. Samples that tested the two most common pairs, White–Black versus White–Asian, did not differ in the contact–CRD relationship, $r_{\text{diff}} = .022, 95\%\ \text{CI} = [-.074, .118], F(1, 204) = 0.21, p = .65$.

Methodological factors

CRD measurement. We used three variables to assess how the CRD was measured. We first present each variable's effect on the contact–CRD effect individually. Then, we examine whether the length of retention and number of faces encoded interacted. Both of these variables may impact memory demands for the participant, altering the contact–CRD relationship.

Length of retention. We examined whether the magnitude of the contact–CRD relationship depended on whether a face was retained for a few of seconds or for minutes/hours. We found no difference between these two categories, $r_{\text{diff}} = -.006, 95\%\ \text{CI} = [-.102, .091], F(1, 205) = .013, p = .910$. Both categories revealed a significant contact–CRD effect.

Number of faces encoded. Next, we considered whether the number of faces a participant was asked to encode affected the contact–CRD relationship. Conceptualized as a categorical variable (one vs. more than one), there was no difference, $r_{\text{diff}} = -.017, 95\%\ \text{CI} = [-.124, .090], F(1, 205) = .102, p = .750$. We next considered the number of faces as a continuous predictor. Due to the relatively high number of studies that asked participants to encode only one face ($k = 30$), we used a square root transformation to normalize the variable. Again, we found no evidence that the magnitude of the contact–CRD relationship depended on the number of faces a person was asked to encode, $r_{\text{slope}} = .0002, 95\%\ \text{CI} = [-.0002, .0005], F(1, 205) = .092, p = .762$.

Number of recognition faces. The last variable that we considered was the number of faces present on a recognition trial. There was no significant difference between whether researchers presented one face or multiple faces on recognition trials, $r_{\text{diff}} = -.047, 95\%\ \text{CI} = [-.145, .052], F(1, 205) = .867, p = .353$. The vast majority of samples only presented one face on each recognition trial (157 out of 207), so we were unable to find a transformation that normalized the continuous form of this variable.

Memory demands. As discussed in the introduction, tasks may vary in the extent to which they rely on multiple memory processes. We allowed the length of retention and the number of faces encoded to interact in a 2 × 2 analysis of variance. Contrary to our hypothesis, these two moderators as a set did not impact the contact–CRD relationship, $F(3, 203) = 0.44, p = .725$. The average contact–CRD effect was marginally different than zero, $r = -.105, 95\%\ \text{CI} = [-.213, .006], F(1, 203) = 3.48, p = .064$. There was no main effect of length of retention, $r_{\text{diff}} = -.071, 95\%\ \text{CI} = [-.283, .147], F(1, 203) = 0.41, p = .523$, or a main effect of the number of faces encoded, $r_{\text{diff}} = .047, 95\%\ \text{CI} = [-.171, .261], F(1, 203) = 0.178, p = .674$. In addition, the interaction between retention time and number of faces that were encoded was not significant, $r_{\text{diff}} = -.240, 95\%\ \text{CI} = [-.598, .198], F(1, 203) = 1.18, p = .279$. In Panel D of Figure 4, the simple effects of this interaction are presented. When researchers asked participants to encode one stimulus for a few seconds (i.e., same–different task), the contact–CRD relationship was significant. The same holds true when researchers asked participants to encode many stimuli for minutes or hours (i.e., encode-recognition task).

In addition to the coding structure, tested above, we analyzed the data for several tasks that are commonly used to measure the CRD, including encode-recognition, same–different, forced-choice, eyewitness, and the Cambridge Face Memory Task. This seems like a useful addition to the analysis above because researchers typically choose one of these tasks (rather than, say, considering whether they prefer a short versus a long retention interval). There were no significant differences between the tasks. Figure 4 presents the mean effect size for each task.

Operationalization of contact. Overall, there was no evidence that operationalization of contact moderated the effect of contact on CRD, $F(5, 201) = 1.64, p = .203$. We examined whether the contact–CRD relationship was impacted by the use of scales (on one hand) compared with manipulated and quasi-experimental operationalizations of contact (on the other). We found that scales were associated with a weaker relationship, $r_{\text{diff}} = .323, 95\%\ \text{CI} = [.063, .542], F(1, 201) = 5.91, p = .016$. Within the scales, we found no difference between idiosyncratic scales and scales that were more commonly used (Social Experience Questionnaire, Walker & Hewstone, and Hancock & Rhodes), $r_{\text{diff}} = .001, 95\%\ \text{CI} = [-.105, .107], F(1, 201) = 0.001, p = .981$. We also tested whether manipulating contact differed from the scales as a set. Manipulating contact yielded a stronger relationship than measuring contact via scales, $r_{\text{diff}} = -.269, 95\%\ \text{CI} = [-.473, -.037], F(1, 201) = 5.17, p = .024$, and a stronger relationship than all of the other operationalizations considered together as

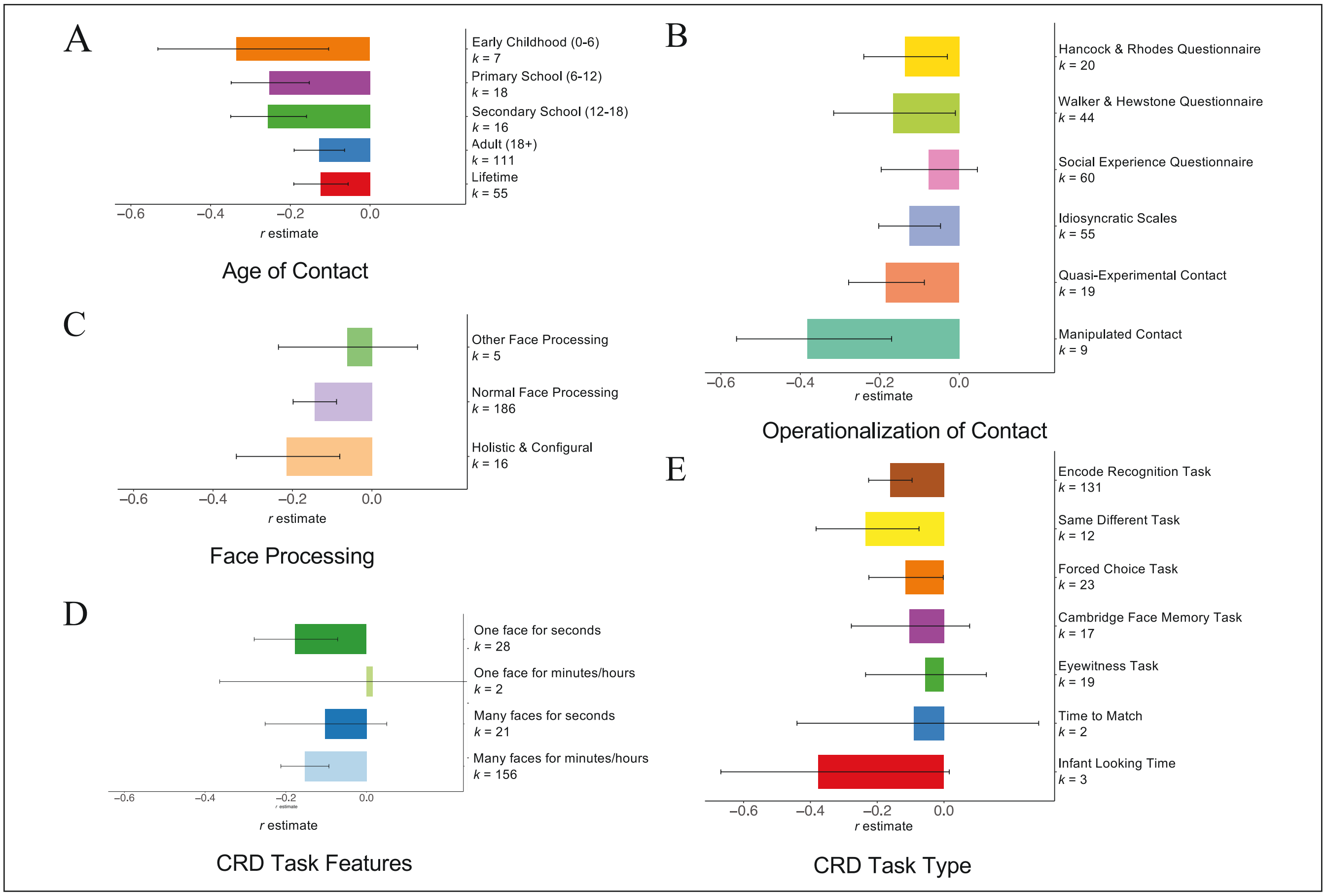

Figure 4. Bar graphs of selected moderators: Age when contact occured (A), how contact was operationalized (B), the type of face processing assessed in the measurement of the CRD (C), features of the CRD measure (D), the type of task used to measure the CRD (E).
Note. Error bars are 95% confidence intervals. CRD = cross-race recognition deficit.

a set, $r_{\text{diff}} = -.258$, 95% CI = [–.463, –.026], $F(1, 201) = 4.79$, $p = .030$. We next examined how manipulated contact differed from each of the other operationalizations individually. Manipulated contact was marginally better than samples that measured contact quasi-experimentally, $r_{\text{diff}} = .212$, 95% CI = [–.035, .436], $F(1,201) = 2.87$, $p = .092$ and significantly better than samples that used the Hancock and Rhodes scale, $r_{\text{diff}} = .259$, 95% CI = [.011, .478], $F(1, 201) = 4.22$, $p = .041$; Social Experimental Questionnaire, $r_{\text{diff}} = .315$, 95% CI = [.065, .528], $F(1, 201) = 6.06$, $p = .015$; and idiosyncratic scales, $r_{\text{diff}} = .270$, 95% CI = [.032, .478], $F(1, 201) = 4.22$, $p = .041$ Manipulated contact did not differ from samples that used the Walker and Hewstone scale, $r_{\text{diff}} = .230$, 95% CI = [–.046, .473], $F(1, 201) = 2.72$, $p = .101$. Examining the simple effects, all of the operationalizations except one (the Social Experience Questionnaire) were associated with significant effects.

Age of cross-race contact. The omnibus test showed evidence that the magnitude of the relationship depended on the age when contact occurred, $F(4, 202) = 3.35$, $p = .011$. We first tested samples that separated effect sizes by the age that contact occurred (e.g., contact during elementary school) versus those that collapsed the effect of contact across multiple age groups (e.g., contact during elementary school and contact during college). We found that samples that were more granular revealed a stronger contact–CRD relationship, $r_{\text{diff}} = -.125$, 95% CI = [–.223, –.023], $F(1, 202) = 5.85$, $p = .016$.

Based on past research (McKone et al., 2019; Pascalis et al., 2020), we expected that contact that occurred before the age of 12 would be most effective in reducing the CRD. Although trending in the predicted direction, this sensitive period hypothesis was not clearly supported, $r_{\text{diff}} = -.213$, 95% CI = [–.452, .055], $F(1, 202) = 2.46$, $p = .118$. To further investigate whether contact during childhood strengthened the contact–CRD relationship, we considered a more general operationalization of childhood. We compared samples that assessed contact prior to 18 and contact after 18. According to this general definition of childhood, childhood contact revealed a stronger contact–CRD effect, $r_{\text{diff}} = -.160$, 95% CI = [–.259, –.058], $F(1, 202) = 9.42$, $p = .002$. The simple effects presented in Table 2 seem to show a pattern of weakening effects, dropping gradually from $r = -.335$ for early childhood contact to $r = -.124$ for lifetime contact.

Face processing. Overall, there was no significant evidence that the nature of face processing moderated the effect of contact on CRD, $F(2, 204) = 1.85, p = .161$. Hancock and Rhodes (2008) argue that contact decreases the CRD by promoting configural processing of cross-race faces. We tested this prediction by comparing samples that use either holistic or configural measures of face memory with samples that assess normal face processing. Although trending in the direction proposed by Hancock and Rhodes, processing type did not significantly moderate the contact–CRD relationship, $r_{\text{diff}} = -.073$, 95% CI $= [-.215, .072], F(1, 204) = 0.99, p = .322$.

Calculation of the CRD. Some researchers did not calculate the CRD by computing a difference scores or statistically controlling for performance with same-race faces. They examined the relationship between contact and a raw measure of cross-race recognition (e.g., the relationship between contact and d' scores for cross-race faces). We tested to see whether the difference in calculation made a difference for the meta-analytic effect and found that it did not, $r_{\text{diff}} = -.067$, 95% CI $= [-.160, .028], F(1, 205) = 1.93, p = .166$. Samples that used relative scores, and those that used non-relative scores were both significant, although, as expected, the effect is directionally stronger when a relative measure is used.

Combined Model

We ran a combined model estimating the overall relationship between contact and the CRD, simultaneously testing all of the moderators discussed above. We tested this model using the contrasts reported above, with two exceptions. For the operationalization of contact, we found a difference between manipulating contact on one hand and the other operationalizations on the other. In addition, we found a difference between samples that assessed contact that occurred before the age of 18 versus contact that occurred after the age of 18. In the combined model, we wanted to make sure these differences were not an artifact of other characteristics of our data set (e.g., maybe children's contact appears more influential because childhood contact is more likely to be manipulated), so we included a set of contrasts that would allow us to directly test the contrast between manipulated rather than other operationalizations of contact and the contrast that tested contact that occurred prior to 18 versus contact that occurred above 18.

The average effect size controlling for all moderators was significantly different from zero, $r = -.239$, 95% CI = $[-.344, -.129], F(1, 184) = 17.68, p < .001$. Note that this effect is larger than the overall meta-analytic effect because it is estimating an unweighted average effect across the set of moderators (e.g., as if there were equal numbers of samples in each age-of-contact category, when in fact studies that measure childhood contact are less frequent). As a set, the moderators accounted for variance in the effect sizes, $F(22, 184) = 1.89, p = .012$. Even after accounting for the moderators, there was significant residual heterogeneity, $Q_E(184) = 303.82, p < .001$.

Five moderators showed significant or marginal effects. First, controlling for the effect of the other moderators, effect sizes derived from manipulating contact were larger than effect sizes derived from other operationalizations, $r_{\text{diff}} = -.305$, 95% CI $= [-.522, -.051], F(1, 184) = 5.55, p = .019$. Second, controlling for the effect of the other moderators, effect sizes derived from contact that occurred prior to 18 were larger than effect sizes derived from contact that occurred after 18, $r_{\text{diff}} = -.132$, 95% CI $= [-.238, -.023], F(1, 184) = 5.68, p = .018$. Controlling for the effect of the other moderators, the linear effect of year remained significant, such that more recent effect sizes were weaker, $r_{\text{slope}} = .007$, 95% CI $= [.0002, .014]$, $F(1, 184) = 4.13, p = .044$. In addition, the quadratic effect of year was marginal, such that the linear trend of weakening of effect sizes seems to be accelerating, $r_{\text{slope}} = .0002$, 95% CI $= [-.00004, .0005], F(1, 184) = 2.88, p = .09$. Finally, the race of the participants emerged as a significant predictor. Samples that collapsed across races or did not use those races had larger effect sizes compared with samples with only White, Black, or Asian participants, $r_{\text{diff}} = -.113$, 95% CI $= [-.216, -.006]$, $F(1, 184) = 4.38, p = .038$. Two reasons give us hesitation in drawing strong conclusions from this last effect. First, this effect only emerges in the combined covariate model. Second the "other" participant code represents an amalgam of samples that defies simple categorization, including 20 samples of a single group (but a group that is not White, Asian, or Black) and 16 samples that collapsed across multiple races. Given the heterogeneity of this group, the effect is hard to interpret.

General Discussion

Our meta-analysis provides strong evidence that cross-race contact reduces the CRD. The overall relationship between contact and the CRD, without considering any moderators, is modest ($r = -.148$). The magnitude of this effect suggests that contact explains 2.18% of variation in the CRD. Our results are largely consistent with the effect reported by Meissner and Brigham (2001). In the following sections, we will discuss the moderators of this relationship and recommendations for future research.

When and How Contact Reduces the CRD

Age of cross-race contact. The results of the meta-analysis strongly suggests that cross-race contact is effective at reducing the CRD throughout the life span. Regardless of when contact occurs, the relationship between contact and the CRD is significant. This suggests that contact is relatively plastic both during childhood and adulthood. However, contact seems to be more effective at reducing the CRD if it occurs prior to adulthood (i.e., before a person is 18 years old). This is both in agreement and at odds with prior research. On one hand, the general assertion that cross-race contact should be more sensitive for younger people (M. M. Davis et al., 2016; Sangrigoli et al., 2005) is supported by the results of our meta-analysis. On the other hand, we found no evidence for the sensitivity period hypothesis as defined by McKone and colleagues (2019) and Pascalis and colleagues (2020). Both argue that contact should be most sensitive if it occurs for children prior to 12 years old. When comparing contact that occurs prior to 12 years old versus after 12 years old, we found no difference in the contact–CRD effect. Rather contact that occurs prior to schooling (0–6), during primary/elementary school (6–12), and during secondary/ middle & high school (12–18) all reveal a strong contact–CRD effect. The effect during each of these time periods (tested individually) is marginally or significantly stronger than adult contact.

It is important to note that our data set had a relatively low proportion of samples that measured contact during childhood (41 out of 207). Some of these samples recruited children and asked about their current contact while other samples asked adults to report on their childhood contact. Of these 41 samples, only seven involved childhood contact under the age of six. This relatively low number of samples does not allow us to get a precise estimate of early childhood contact. Although the point estimate of early childhood contact is the largest ($r = -.335$), the standard error of the simple effect of early childhood contact ($SE = 0.124$) is more than double the standard error of every other time period. We recommend that researchers investigate early childhood contact, as it is possible that this time period of contact may be especially influential in the contact–CRD relationship, but more data would be valuable.

Our data set also had a large proportion of adult contact ($k = 111$) and contact that occurred during multiple time periods ($k = 55$). The latter is especially problematic in attempting to examine the question of what time period of contact is most effective at reducing the CRD. For these 55 samples, we cannot be sure which time period of contact helped reduce the CRD. We would highly recommend that future researchers report contact as a function of the time period it occurs.

Face processing. In our meta-analysis, we did not find support for the hypothesis that contact mitigates the CRD by increasing configural/holistic processing, specifically. When we examined the simple effects, we observed significant relationships for both normal and for holistic/configural face processing.

This null result could have occurred for many reasons. Hayward and colleagues (2013) reviewed a variety of different holistic/configural tasks (e.g., the part-whole task, tasks with blurred or inverted faces). They argue that these tasks may be probing different processes. In addition, some paradigms may yield noisy effects. For example, some samples using inverted faces show inversion decrements for both same-race and cross-race faces, whereas others only show an inversion decrement for same-race faces (the expected outcome). We may have been unable to see any difference as a function of face processing due to heterogeneity within the configural/holistic tasks.

Another possibility is that cognitive styles and processing styles differ across races (Hayward et al., 2013). In some cases, researchers report that Asians process faces differently than Whites, in that Asians do not show a holistic processing decrement for cross-race faces (Michel et al., 2006, 2013, but see Crookes et al., 2013). Although the CRD is robust across cultures, it is possible that the recognition deficits occur through different mechanisms in each culture. We were unable to fully test this hypothesis in our data set as only three samples assessed holistic and/or configural processing in Asian participants. We instead examined whether there was any difference in the contact–CRD relationship for Asians versus Whites or Asians versus other races and observed no difference in either contrast.

Although we did not observe a stronger relationship when the CRD was assessed using a holistic measure, it is important to note that the direction of the difference is consistent with Hancock and Rhodes's (2008) predictions. It is possible that a difference does exist between holistic, configural, and normal face processing, but at this point, more research is needed in order for a definitive conclusion.

Operationalizations of Contact and the CRD

Cross-race contact. The strongest moderator in our analysis was the operationalization of contact. Samples that manipulated contact showed a significantly larger contact–CRD relationship regardless of whether the variable was considered in isolation, or in the context of every other moderator we considered. To provide context, the simple effect among samples that manipulated contact was $r = -.382$, which equates to contact explaining 14.61% of the variation in the CRD.

Manipulating a variable versus measuring one is generally seen as the gold standard in psychology and other sciences (Shadish et al., 2001), so it should come as no surprise that manipulating contact would decrease the CRD more than other operationalizations. Still, the strength of the effect suggests that manipulating contact may be a more reliable way to measure the contact–CRD effect and also hints that the true effect between contact and the CRD may be larger than our measured overall contact–CRD estimate. As discussed in the introduction, some forms of contact probably do not promote the kind of processing that *should* reduce the CRD. Quasi-experimental or self-reported contact may actually represent conflict or disinterested coexistence rather than meaningful person-to-person contact in which the perceiver learns to conceptualize members of another race as individuals.

Examining the other operationalizations, we see two interesting patterns. First, all but one of the other operationalizations of contact showed a significant contact–CRD effect. The lone exception was the Social Experience Questionnaire. As we understand it, the first version of this questionnaire was constructed by Brigham in 1993. It was further developed over the years by the same lab, growing from a 16-item scale to a 56-item scale (Slone et al., 2000). The heterogeneity in the scale may have contributed to the weak relationship between contact and the CRD. We chose to group the different versions of this scale as one because they were authored by the same lab and because it was often difficult to know which version of the scale researchers were using.

In addition, self-reported scales as a set had weaker effect sizes compared with quasi-experimental and manipulated contact. The relative weakness of the scales may reflect a lack of variation in cross-race contact in the population: If a predictor does not vary, it cannot predict an outcome. It is also perhaps worth noting that most of the scales examined in this meta-analysis include items that are face valid, but that have not been subjected to rigorous psychometric analysis. Many of these scales were developed ad hoc, in the context of studies that sought to test the effects of contact rather than in studies that were focused exclusively on scale development. A more rigorous approach may yield measures with better reliability and better construct validity.

Measurement of the CRD. Although our meta-analysis found no differences in the way the CRD was measured, the moderator still deserves some discussion. We reported tests from three variables in the meta-analysis: the length of time a face was retained for, the number of faces a person had to encode before being tested, and the number of faces a person saw in a given recognition trial. We found no differences in any of these variables when considered in isolation, in additive models, or interactive models. In addition to analyzing all three variables as categorical predictors, we also considered the number of encoding faces and the number of recognition faces as continuous predictors and again found no differences. Finally, in the supplemental analysis, we considered differences between tasks, relying simply on the paradigm names to classify them (e.g., encode-recognition vs. same–different vs. Cambridge Face Memory Task). Although we did not find any differences in the different task frameworks, samples that used the Cambridge Face Memory Task (CFMT; Duchaine & Nakayama, 2006) or idiosyncratic tasks did not show a significant contact–CRD effect. The CFMT is a unique face perception task that asks participants to hold six faces in memory for a few seconds, repeatedly exposing and testing participants on the same faces in different positions (full frontal, three fourth profile, etc.). The CFMT was originally designed to test for prosopagnosia (Duchaine & Nakayama, 2006) and has also been used to test for super-recognizers (Bobak et al., 2016). However, general face recognition ability may be orthogonal to the CRD and its relationship with contact (Correll et al., 2021). According to the results of our meta-analysis, the CFMT does not seem well suited to detecting a relationship between contact and the CRD.

So which task would best reveal the effect between contact and the CRD? The results from our meta-analysis are at best inconclusive. We will review the literature on measurement of the CRD to establish some guiding principles. However, we want to make it clear to the reader that in our meta-analysis there was no differences between any of the features of the tasks.

For a researcher trying to decide how to measure the CRD, we would suggest a task that reduces memory demands. The point estimates were directionally stronger for tasks that reduced memory demands (again, this difference was not significant). Furthermore, it seems reasonable to suggest that contact improves a participant's ability to visually process *individual* cross-race faces. It seems much less likely that contact improves either long-term memory or the participant's ability to process multiple faces at the same time. To the extent that measurement tasks involve processes that are not related to contact (which, for the current investigation, constitute "constructs of disinterest," Judd & McClelland, 1998), they should reduce the contact–CRD relationship by adding systematic error variance.

Other Moderators

In addition to the theoretically important and study design moderators, we also examined moderators focused on how and when the sample was collected. Based on the results of our meta-analysis, neither the participants' race nor the races of the stimuli seem to matter. Although no null effect is ever conclusive, our results suggest that contact can reliably reduce the CRD in people of multiple races and multiple race pairings.

We did not expect the year the study was published to be correlated with weaker contact–CRD effect sizes. The relationship may reflect more complete reporting in recent years, including the reporting of null results and the availability of unpublished dissertations.

Is the Relationship Between Contact and the CRD Actually Small?

It is quite possible that the true relationship between contact and the CRD is indeed small. The point estimates between our meta-analysis and Meissner and Brigham's are quite similar. This is especially notable as our meta-analysis had more than twice as many studies and many more tests of the effect. Over the past 20 years, researchers have looked at the relationship between contact and the CRD in numerous ways with many different methodologies, and yet the relationship in both meta-analyses is remarkably similar.

However, the heterogeneity in our results gives us pause. Operationalizations of contact and measurement of the CRD varied widely in our data set. For example, operationalizations of contact ranged from asking where a person lives to detailed 56-item scales to manipulations that involved five days of intensive training on cross-race faces. It seems to us that the field would benefit from greater clarity about the differences in operationalization and about differences in the very definition of contact. Theoretically, there is no reason to expect that contact defined by sheer frequency of cross-race encounters; contact defined by positive, intimate friendship with one or two cross-race peers; and contact defined by individuated processing of large numbers of cross-race co-workers should have similar effects on the CRD.

Beyond ambiguity in the construct, even in 2021, there may not be enough variance in cross-sectional measures of contact to consistently detect a relationship with the CRD. The large effect size associated with manipulations of contact hints at this possibility. Notably, many papers in this data set report very low levels of cross-race contact. Although the world is trending toward integration (Frey, 2018; U.S. Census Bureau, 2018), and opportunities for cross-race contact may be increasing, much of the world is still quite segregated. As we have already discussed, low variance in contact will reduce its correlations with the CRD. Clearer conceptual definitions of the relevant constructs, better operationalizations, and studies that maximize variation on contact (either by manipulating it or by sampling more strategically) may yield a much stronger relationship.

Conclusion

There are three primary conclusions from this analysis. First, this work largely replicates Meissner and Brigham (2001). Although we examine a much larger sample of studies using diverse methods and populations, the results are consistent with their original estimate of the average relationship between cross-race contact and the CRD. This consistency reinforces the idea that the contact–CRD relationship is real, but small. However, the second conclusion is that this "weak" relationship may be systematically dampened by the methodological choices researchers make. The most commonly used methods for operationalizing contact (self-report measures) seem to yield small effects, whereas manipulating contact yields much larger effects ($r = -.382$). This difference may suggest that the true relationship between contact and the CRD is larger than the overall effect we observed. At the same time, there are relatively few samples that manipulate contact, so more research is needed. The third conclusion is that although contact throughout the life span can reduce the CRD, early contact seems to be especially effective.

Acknowledgments

We would like to thank the entire Correll Lab at CU Boulder for their assistance on this project, especially Brittany Birdsey and Jackson Baker, who contributed to the initial stages of this meta-analysis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Support for this work was provided by National Science Foundation Standard Grant 1946788.

ORCID iDs

Balbir Singh https://orcid.org/0000-0002-3120-1791
Christopher Mellinger https://orcid.org/0000-0001-6182-6170

Notes

1. Note that the type of processing can be assessed with a variety of tasks. Encode-recognition, forced-choice, and same–different task structures can all be used while assessing normal or holistic/configural processing, for example, by utilizing inverted faces.
2. Two of these samples tested participants on features of the to-be-remembered face. These featural tasks do not fall cleanly as holistic/configural tasks. Because holistic processing refers to gestalt processing of a face, the presentation of a subset of the face precludes this type of processing to occur. The other three samples combined holistic and normal processed faces. These two samples fall in the middle of our holistic/configural normal distinction and thus were also classified as "other."

Supplemental Material

Supplemental material is available online with this article.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

*Anzures, G., Wheeler, A., Quinn, P. C., Pascalis, O., Slater, A. M., Heron-Delaney, M., Tanaka, J. W., & Lee, K. (2012). Brief daily exposures to Asian females reverses perceptual narrowing for Asian faces in Caucasian infants. *Journal of Experimental Child Psychology*, *112*(4), 484–495. https://doi.org/10.1016/j.jecp.2012.04.005

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

*Bennett, L. B. (2004). *The development of the "Other-Race Effect" in children's face recognition memory* (Publication No. 3282569) [Doctoral dissertation, The Florida State University]. ProQuest Information & Learning.

*Berger, D. G. (1969). *They all look alike* [Unpublished doctoral dissertation]. Vanderbilt University.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological*

Science, *18*(8), 706–712. https://doi.org/10.1111/j.1467-9280.2007.01964.x

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, *30*(1), 81–91. https://doi.org/10.1002/acp.3170

*Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, *42*(4), 673–681. https://doi.org/10.1037/0022-3514.42.4.673

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

*Bukach, C. M., Cottle, J., Ubiwa, J., & Miller, J. (2012). Individuation experience predicts other-race effects in holistic processing for both Caucasian and Black participants. *Cognition*, *123*(2), 319–324. https://doi.org/10.1016/j.cognition.2012.02.007

*Burgess, M. C. R. (1997). *The cross-race effect in facial recognition: A function of expertise?* (Publication No. 9817306) [Doctoral dissertation, The Florida State University]. ProQuest Information & Learning.

*Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research*, *38*, 2455–2468.

*Carroo, A. W. (1986). Other race recognition: A comparison of Black American and African subjects. *Perceptual and Motor Skills*, *62*(1), 135–138. https://doi.org/10.2466/pms.1986.62.1.135

*Carroo, A. W. (1987). Recognition of faces as a function of race, attitudes, and reported cross-racial friendships. *Perceptual and Motor Skills*, *64*(1), 319–325. https://doi.org/10.2466/pms.1987.64.1.319

*Chance, J., Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory for faces. *The Journal of Social Psychology*, *97*(2), 243–253. https://doi.org/10.1080/00224545.1975.9923344

*Childs, M. J., Jones, A., Thwaites, P., Zdravkovic, S., Thorley, C., Suzuki, A., Shen, R., Ding, Q., . . . Tree, J. J. (in press). Do individual differences in face recognition ability moderate the other ethnicity effect? *Journal of Experimental Psychology: Human Perception and Performance*.

*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 Section A*, *48*(4), 879–894. https://doi.org/10.1080/14640749508401421

*Corenblum, B., & Meissner, C. A. (2006). Recognition of faces of ingroup and outgroup children and adults. *Journal of Experimental Child Psychology*, *93*(3), 187–206. https://doi.org/10.1016/j.jecp.2005.09.001

Correll, J., Hudson, S. M., Guillermo, S., & Earls, H. A. (2017). Of kith and kin: Perceptual enrichment, expectancy, and reciprocity in face perception. *Personality and Social Psychology Review*, *21*(4), 336–360. https://doi.org/10.1177/1088868316657250

*Correll, J., Ma, D. S., & Davis, J. P. (2021). Perceptual tuning through contact? Contact interacts with perceptual (not memory-based) face-processing ability to predict cross-race recognition. *Journal of Experimental Social Psychology*, *92*, Article 104058. https://doi.org/10.1016/j.jesp.2020.104058

*Correll, J., Singh, B., & Mellinger, C. (2021). *High power test of contact on CRD using Qualtrics Panels 1*. Unpublished manuscript.

Cowan, N. (2008). Chapter 20: What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Essence of memory* (Vol. 169, pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9

Crookes, K., Favelle, S., & Hayward, W. G. (2013). Holistic processing for other-race faces in Chinese participants occurs for upright but not inverted faces. *Frontiers in Psychology*, *4*, Article 29. https://doi.org/10.3389/fpsyg.2013.00029

*Davis, E. E., Matthews, C. M., & Mondloch, C. J. (2020). Ensemble coding of facial identity is not refined by experience: Evidence from other-race and inverted faces. *British Journal of Psychology*, *112*, 265–281. https://doi.org/10.1111/bjop.12457

*Davis, M. M., Hudson, S. M., Ma, D. S., & Correll, J. (2016). Childhood contact predicts hemispheric asymmetry in cross-race face processing. *Psychonomic Bulletin & Review*, *23*(3), 824–830. https://doi.org/10.3758/s13423-015-0972-7

*De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect: Experience and the other-race effect. *Developmental Science*, *13*(1), 181–187. https://doi.org/10.1111/j.1467-7687.2009.00876.x

Dickter, C. L., & Bartholow, B. D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, *2*(3), 189–198. https://doi.org/10.1093/scan/nsm012

Dickter, C. L., Gagnon, K. T., Gyurovski, I. I., & Brewington, B. S. (2015). Close contact with racial outgroup members moderates attentional allocation towards outgroup versus ingroup faces. *Group Processes & Intergroup Relations*, *18*(1), 76–88. https://doi.org/10.1177/1368430214527854

*Digiovanni, L. (2010). *Accuracy of eyewitness identification of Black Americans as a function of social dominance orientation and cross race contact* (Publication No. 3405267) [Doctoral dissertation, Hofstra University]. ProQuest Information & Learning.

Donders, N. C., Correll, J., & Wittenbrink, B. (2008). Danger stereotypes predict racially biased attentional allocation. *Journal of Experimental Social Psychology*, *44*(5), 1328–1333. https://doi.org/10.1016/j.jesp.2008.04.002

*Dore, H. S. (2002). *The own-race bias in children's eyewitness memory* (Publication No. 3047111) [Doctoral dissertation, The Florida State University]. ProQuest Information & Learning.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

*Elliott, E. S., Wills, E. J., & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society*, *2*(2), 71–73. https://doi.org/10.3758/BF03327717

*Feinman, S., & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development*, *47*(2), 506. https://doi.org/10.2307/1128809

*Finklea, K. M. (2008). *Investigating inappropriate cue utilization in the own-race bias* (Publication No. 3307561) [Doctoral dissertation, University of California San Diego]. ProQuest Information & Learning.

Flevaris, T. V., & Chapman, E. F. (2015). Cross-racial misidentification: A call to action in Washington State and beyond. *Seattle University Law Review*, *38*(3), Article 861.

Frey, W. (2018). *Black-white segregation edges downward since 2000, census shows*. The Brookings Institution. https://www.brookings.edu/blog/the-avenue/2018/12/17/black-white-segregation-edges-downward-since-2000-census-shows/

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573. https://doi.org/10.1038/9224

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, *38*(15–16), 2401–2428. https://doi.org/10.1016/S0042-6989(97)00442-2

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., Slater, A., Tian, J., & Lee, K. (2009). Two faces of the other-race effect: Recognition and categorisation of Caucasian and Chinese faces. *Perception*, *38*(8), 1199–1210. https://doi.org/10.1068/p6136

*Goldstein, A. G., & Chance, J. E. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society*, *23*(3), 211–214. https://doi.org/10.3758/BF03329829

Guillermo, S., & Correll, J. (2016). Attentional biases toward Latinos. *Hispanic Journal of Behavioral Sciences*, *38*(2), 264–278. https://doi.org/10.1177/0739986316639525

*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

*Harvey, A. J. (2014). Some effects of alcohol and eye movements on cross-race face learning. *Memory*, *22*(8), 1126–1138. https://doi.org/10.1080/09658211.2013.872278

Harzing, A. W. (2007). *Publish or perish*. https://harzing.com/resources/publish-or-perish

Hayward, W. G., Crookes, K., & Rhodes, G. (2013). The other-race effect: Holistic coding differences and beyond. *Visual Cognition*, *21*(9–10), 1224–1247. https://doi.org/10.1080/13506285.2013.824530

*Heron-Delaney, M., Anzures, G., Herbert, J. S., Quinn, P. C., Slater, A. M., Tanaka, J. W., Lee, K., & Pascalis, O. (2011). Perceptual training prevents the emergence of the other race effect during infancy. *PLOS ONE*, *6*(5), Article e19858. https://doi.org/10.1371/journal.pone.0019858

*Hilliar, K. F., Kemp, R. I., & Denson, T. F. (2010). Now everyone looks the same: Alcohol intoxication reduces the own-race bias in face recognition. *Law and Human Behavior*, *34*(5), 367–378. https://doi.org/10.1007/s10979-009-9204-x

Hirose, Y., & Hancock, P. J. B. (2007). Equally attending but still not seeing: An eye-tracking study of change detection in own- and other-race faces. *Visual Cognition*, *15*(6), 647–660. https://doi.org/10.1080/13506280601069578

*Hjartarson, K. H. (2017). *High-level vision in dyslexia: The possibility of inefficient visual expertise as a contributing factor to reading problems* [Unpublished doctoral dissertation]. Heilbrigðisvísindasvið, Háskóla Íslands.

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), 334–340. https://doi.org/10.1016/j.jesp.2006.02.010

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

Hughes, B. L., Camp, N. P., Gomez, J., Natu, V. S., Grill-Spector, K., & Eberhardt, J. L. (2019). Neural adaptation to faces reveals racial outgroup homogeneity effects in early perception. *Proceedings of the National Academy of Sciences*, *116*(29), 14532–14537. https://doi.org/10.1073/pnas.1822084116

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences*, *13*(12), 524–531. https://doi.org/10.1016/j.tics.2009.10.002

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*(4), 616–626. https://doi.org/10.1037/0022-3514.85.4.616

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 21–36. https://doi.org/10.3758/CABN.5.1.21

*Jackiw, L. B., Arbuthnott, K. D., Pfeifer, J. E., Marcon, J. L., & Meissner, C. A. (2008). Examining the cross-race effect in lineup identification using Caucasian and First Nations samples. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *40*(1), 52–57. https://doi.org/10.1037/0008-400x.40.1.52

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 637–679). McGraw-Hill; Distributed exclusively by Oxford University Press.

*Lampinen, J. M., Roush, A., Erickson, W. B., Moore, K. N., & Race, B. (2015). The effects of simulated distance on recognition of same race and other race faces. *Visual Cognition*, *23*(6), 678–698. https://doi.org/10.1080/13506285.2015.1084071

*Lavrakas, P. J., Buri, J. R., & Mayzner, M. S. (1976). A perspective on the recognition of other-race faces. *Perception & Psychophysics*, *20*(6), 475–481. https://doi.org/10.3758/BF03208285

*Lebrecht, S., Pierce, L. J., Tarr, M. J., & Tanaka, J. W. (2009). Perceptual other-race training reduces implicit racial bias. *PLOS ONE*, *4*(1), Article e4215. https://doi.org/10.1371/journal.pone.0004215

Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1364–1382. https://doi.org/10.1037/0278-7393.22.6.1364

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of*

Experimental Psychology: General, *129*(4), 559–574. https://doi.org/10.1037/0096-3445.129.4.559

*Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology*, *76*(4), 587–589. https://doi.org/10.1037/0021-9010.76.4.587

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556–564. https://doi.org/10.1037/0021-9010.70.3.556

Lovén, J., Rehnman, J., Wiens, S., Lindholm, T., Peira, N., & Herlitz, A. (2012). Who are you looking at? The influence of face gender on visual attention and memory for own- and other-race faces. *Memory*, *20*(4), 321–331. https://doi.org/10.1080/09658211.2012.658064

*Maclin, O. H., Van Sickler, B. R., Maclin, M. K., & Li, A. (2004). A re-examination of the cross-race effect: The role of race, inversion, and basketball trivia. *North American Journal of Psychology*, *6*(2), 189–203.

Makovski, T., Watson, L. M., Koutstaal, W., & Jiang, Y. V. (2010). Method matters: Systematic effects of testing procedure on visual working memory sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1466–1479. https://doi.org/10.1037/a0020851

*Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334. https://doi.org/10.1037/h0028434

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4

McGugin, R. W., Newton, A. T., Gore, J. C., & Gauthier, I. (2014). Robust expertise effects in right FFA. *Neuropsychologia*, *63*, 135–144. https://doi.org/10.1016/j.neuropsychologia.2014.08.029

*McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., Kidd, E., & Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Scientific Reports*, *9*(1), Article 12820. https://doi.org/10.1038/s41598-019-49202-0

*Meissner, C. A. (2001). *The influence of perceptual learning in memory for faces: A process-tracing approach to investigating the own-race bias* (Publication No. 3021544) [Doctoral dissertation, The Florida State University]. ProQuest Information & Learning.

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

*Mellinger, C., Singh, B., & Correll, J. (2021). *The integrated cross-race contact scale*. Manuscript submitted for publication.

*Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*(1), 55–73. https://doi.org/10.1080/13506280500158761

Michel, C., Rossion, B., Bülthoff, I., Hayward, W. G., & Vuong, Q. C. (2013). The contribution of shape and surface information in the other-race face effect. *Visual Cognition, 21*(9–10), 1202–1223. https://doi.org/10.1080/13506285.2013.823141

*Ng, W., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-cultural Psychology*, *25*(2), 217–232. https://doi.org/10.1177/0022022194252004

Organisation for Economic Co-Operation and Development & United Nations Department of Economic and Social Affairs Population Division. (2013). *World migration in figures*. http://www.oecd.org/els/mig/World-Migration-in-Figures.pdf

Page-Gould, E., Mendoza-Denton, R., Alegre, J. M., & Siy, J. O. (2010). Understanding the impact of cross-group friendship on interactions with novel outgroup members. *Journal of Personality and Social Psychology*, *98*(5), 775–793. https://doi.org/10.1037/a0017880

Page-Gould, E., Mendoza-Denton, R., & Tropp, L. R. (2008). With a little help from my cross-group friend: Reducing anxiety in intergroup contexts through cross-group friendship. *Journal of Personality and Social Psychology*, *95*(5), 1080–1094. https://doi.org/10.1037/0022-3514.95.5.1080

Pascalis, O., Fort, M., & Quinn, P. C. (2020). Development of face processing: Are there critical or sensitive periods? *Current Opinion in Behavioral Sciences*, *36*, 7–12. https://doi.org/10.1016/j.cobeha.2020.05.005

Pastor, D. A., & Lazowski, R. A. (2018). On the multilevel nature of meta-analysis: A tutorial, comparison of software programs, and discussion of analytic choices. *Multivariate Behavioral Research*, *53*(1), 74–89. https://doi.org/10.1080/00273171.2017.1365684

*Platz, S. J., & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology*, *18*(11), 972–984. https://doi.org/10.1111/j.1559-1816.1988.tb01187.x

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

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80*(3), 313–331. https://doi.org/10.1111/j.2044-8295.1989.tb02323.x

*Rhodes, G., Ewing, L., Hayward, W. G., Maurer, D., Mondloch, C. J., & Tanaka, J. W. (2009). Contact and other-race effects in configural and component processing of faces. *British Journal of Psychology*, *100*(4), 717–728. https://doi.org/10.1348/000712608X396503

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review*, *13*(3), 499–505. https://doi.org/10.3758/BF03193876

Richler, J. J., Wong, Y. K., & Gauthier, I. (2011). Perceptual expertise as a shift from strategic interference to automatic holistic processing. *Current Directions in Psychological Science*, *20*(2), 129–134. https://doi.org/10.1177/0963721411402472

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rossion, B., Curran, T., & Gauthier, I. (2002). A defense of the subordinate-level expertise account for the N170 component. *Cognition*, *85*(2), 189–196. https://doi.org/10.1016/S0010-0277(02)00101-4

*Sadozai, A. K., Kempen, K., Tredoux, C., & Robbins, R. A. (2019). Can we look past people's race? The effect of combining race and a non-racial group affiliation on holistic processing. *Quarterly Journal of Experimental Psychology*, *72*(3), 557–569. https://doi.org/10.1177/1747021818760482

*Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V., A., G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*(6), 440–444. https://doi.org/10.1111/j.0956-7976.2005.01554.x

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer.

*Seutloali, G. M. (2015). *Own-race bias in facial recognition amongst black, coloured and white participants* [Unpublished doctoral dissertation]. University of Cape Town.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

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

*Sigurdardottir, H. M., Hjartarson, K. H., Gudmundsson, G. L., & Kristjánsson, Á. (2019). Own-race and other-race face recognition problems without visual expertise problems in dyslexic readers. *Vision Research*, *158*, 146–156. https://doi.org/10.1016/j.visres.2019.02.010

*Slone, A. E., Brigham, J. C., & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in whites. *Basic and Applied Social Psychology*, *22*(2), 71–84. https://doi.org/10.1207/S15324834BASP2202_1

*Sporer, S. L. (2010). *The own-ethnicity effect in facial recognition: Do Turks also look alike to Germans?* Unpublished manuscript.

*Sporer, S. L., Gehrke, J., & McClure, K. A. (2020). Patrolling the borders: Experience with multiple out-group members is associated with the own-ethnicity effect. *Journal of Police and Criminal Psychology*. https://doi.org/10.1007/s11896-020-09387-x

*Sporer, S. L., Trinkl, B., & Guberova, E. (2007). Matching faces: Differences in processing speed of out-group faces by different ethnic groups. *Journal of Cross-Cultural Psychology*, *38*(4), 398–412. https://doi.org/10.1177/0022022107302310

*Stahl, J., Wiese, H., & Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: An event-related potential study. *NeuroReport*, *19*(5), 583–587. https://doi.org/10.1097/WNR.0b013e3282f97b4d

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

*Susa, K. J., Meissner, C. A., & de Heer, H. (2010). Modeling the role of social-cognitive processes in the recognition of own- and other-race faces. *Social Cognition*, *28*(4), 523–537. https://doi.org/10.1521/soco.2010.28.4.523

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*(1), 43–47. https://doi.org/10.1111/1467-9280.00308

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

*Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective & Behavioral Neuroscience*, *9*(1), 122–131. https://doi.org/10.3758/CABN.9.1.122

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*(8), 764–769. https://doi.org/10.1038/77666

*The Harvard Law Review Association. (1988). Developments in the law: Race and the criminal process. *Harvard Law Review*, *101*(7), 1472–1641. https://doi.org/10.2307/1341402

Trawalter, S., Todd, A. R., Baird, A. A., & Richeson, J. A. (2008). Attending to threat: Race-based patterns of selective attention. *Journal of Experimental Social Psychology*, *44*(5), 1322–1327. https://doi.org/10.1016/j.jesp.2008.03.006

*Tullis, J. G., Benjamin, A. S., & Liu, X. (2014). Self-pacing study of faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect. *Memory & Cognition*, *42*(6), 863–875. https://doi.org/10.3758/s13421-014-0409-y

U.S. Census Bureau. (2018). *2013-2017 American community survey five-year estimates*. American Community Survey.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Walker, P. M., & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology*, *20*(4), 461–475. https://doi.org/10.1002/acp.1191

*Walker, P. M., & Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception*, *32*(9), 1117–1125. https://doi.org/10.1068/p5098

*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

*Wheat, E. (2010). *The other race effect: The role of experience and social attitudes on face recognition* (Publication No. 695) [Master's Thesis]. University of Richmond.

Whitehurst, G. J., Reeves, R. V., & Rodrigue, E. (2016). *Segregation, race, and charter schools: What do we know?* Center on Children and Families at Brookings. https://www.brookings.edu/wp-content/uploads/2016/10/ccf_20161021segregation_version-10_211.pdf

*Wiese, H., Kaufmann, J. M., & Schweinberger, S. R. (2014). The neural signature of the own-race bias: Evidence from event-related potentials. *Cerebral Cortex*, *24*(3), 826–835. https://doi.org/10.1093/cercor/bhs369

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The cross-race effect and eyewitness identification: How to improve recognition and reduce decision errors in eyewitness situations: CRE and eyewitness identification. *Social Issues and Policy Review*, *7*(1), 83–113. https://doi.org/10.1111/j.1751-2409.2012.01044.x

*Wong, H. K., Stephen, I. D., & Keeble, D. R. T. (2020). The own-race bias for face recognition in a multiracial society. *Frontiers in Psychology*, *11*, Article 208. https://doi.org/10.3389/fpsyg.2020.00208

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law*, *7*(1), 119–133. https://doi.org/10.1037/1076-8971.7.1.119

*Wylie, L. E., Bergt, S., Haby, J., Brank, E. M., & Bornstein, B. H. (2015). Age and lineup type differences in the own-race bias. *Psychology, Crime & Law*, *21*(5), 490–506. https://doi.org/10.1080/1068316X.2014.989173

*Young, S. G., & Hugenberg, K. (2012). Individuation motivation and face experience can operate jointly to produce the own-race bias. *Social Psychological and Personality Science*, *3*(1), 80–87. https://doi.org/10.1177/1948550611409759

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

Zalman, M. (2012). Qualitatively estimating the incidence of wrongful convictions. *Criminal Law Bulletin*, *48*(2), 221–279.

*Zhao, M., Hayward, W. G., & Bulthoff, I. (2014). Face format at encoding affects the other-race effect in face memory. *Journal of Vision*, *14*(9), 6–6. https://doi.org/10.1167/14.9.6