PLAGIARISM

DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and to pretend that it is one's own.

2. I have used the *American Psychological Association (APA)* 7th Edition convention for citation and referencing. Each significant contribution to, and quotation in, this essay / report / project / from the work, or works, of other people has been attributed, and has cited and referenced.

3. This project is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing if off as his or her own work.

5. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

SIGNATURE …Signed by candidate……………………………..

A Meta-analytic Review of the Contact, Outgroup Ethnic-Nationality Effect (OENE) and Prejudice Relationships[1]

Jade Mac Donnell
ACSENT Laboratory
Department of Psychology, University of Cape Town

Supervisor: Professor Colin Tredoux

February 2023

[1] Funding received from the University of Cape Town

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. I know the meaning of plagiarism and declare that all the work in the document is my own work and that each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signed by candidate

February 2024

Signed

Date

Word count: 21 889

Abstract

A META-ANALYTIC REVIEW OF THE CONTACT, OUTGROUP ETHNIC-NATIONALITY EFFECT (OENE) AND PREJUDICE RELATIONSHIPS

Jade Mac Donnell
University of Cape Town

Identification performance is poorer for ethnic outgroup members relative to in-group members (Laurence et al., 2016). This outgroup identification deficit is referred to in this thesis as the Outgroup Ethnic Nationality Effect (OENE). Such misidentifications, while often earnest, translate into wrongful convictions with a high fiscal and social cost. Consequently, while the theoretical underpinnings thereof and means with which to attenuate such an identification deficit have been extensively studied, a paradoxical finding has emerged with regard to outgroup contact and the OENE. Outgroup contact is central to theoretical explanations of the OENE and higher levels of outgroup contact have been associated with improvements in outgroup identification performance and yet, the strength of the observed relationship between outgroup contact and the OENE remains small (Meissner & Brigham, 2001; Singh et al., 2021; Young et al., 2012). This research therefore sought to investigate this relationship by including and accounting for the influence of outgroup prejudice. This meta-analysis aggregated 53 years of research on the outgroup prejudice, outgroup contact, and OENE relationships and made use of matched-to-sample outgroup prejudice data to explore the moderating role of outgroup prejudice. Included studies tested memory for outgroup faces via an identification task and the final sample included data from 8418 participants spanning 54 studies. Consistent with prior studies, all outgroup contact-OENE relationships were small in strength and affirmed the associated gains in outgroup identification performance when outgroup contact increased. Higher outgroup prejudice was associated with a reduction in outgroup contact and higher outgroup prejudice was associated with a decrement in outgroup identification performance. The findings support the tri-directional relationship between outgroup contact, outgroup prejudice and the OENE. Future studies should not ignore the interconnected nature thereof and should measure outgroup prejudice alongside outgroup contact and OENE data.

Keywords. Contact, Identification, Prejudice, Outgroup Ethnic Nationality Effect, Meta-analysis, Outgroup Identification, Outgroup Recognition, Merged Prejudice Data, Own race bias, Own race effect, Own group bias

Acknowledgements

Thank you to my supervisor, Professor Colin Tredoux, for his unwavering support, kindness and invaluable knowledge base. Without your support, this would not have been possible.

Thank you to my family for their unending support. Thank you to my mother, Uncle Rob, Aunt Ingrid and the whole UK family. Thank you for your patience, guidance and belief in me.

Thank you to the Eyewitness Lab for their support and companionship throughout this journey

Thank you to the Psychology Department at the University of Cape Town

Thank you to Kyra Scott for your support and all the laughs shared throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Faces, and by extension the perceptual information stored within faces, are special in that they anchor social interactions (McKone et al., 2019). They are the means through which we orientate ourselves socially. Facial familiarity can be conceptualized as existing on a continuum, with the highest level of familiarity reserved for faces of family members or primary caregivers (Kramer et al., 2018). While we are faster and more accurate at identifying familiar faces, the ability to identify unfamiliar, or less familiar faces from limited prior encounters, is equally as important for effective social functioning. The ability to recognize familiarity, attribute contextual information to a face and ultimately attach a name to a face can either build or inhibit the formation of social relationships. Given the number of social interactions across the lifespan, some degree of identification error is to be expected. While base level identification slip ups are to be expected, identifications are more error prone when the to-be-identified face belongs to an ethnic outgroup (Laurence et al., 2016). This phenomenon is referred in this thesis to as the Outgroup Ethnic-Nationality Effect[2] (OENE). This in-group identification advantage, wherein we are better able to discriminate faces belonging to our own relative to other ethnic[3] groups, limits the potential for cross-ethnic friendships; cross-ethnic favourable attitudes; and increases the likelihood of internalized or externalized outgroup prejudice (Brigham & Malpass, 1985; Turner et al., 2007b). When the memory component of identifications is isolated, as is the case in airports or other security settings wherein officials match an in-view face to an identification document, the OENE persists and manifests as an increased number of errors in cross-ethnic face to photo matching (Meissner et al., 2013).

The extent of this identification deficit underscores the importance of focusing on theoretical underpinnings and by extension the means through which this identification deficit can be attenuated. While the exact mechanisms, or rather the extent of their influence on the emergence of the OENE are debated, the role of other or out-group contact is

---

[2] Historically this in-group identification advantage has been referred to as the 'Own Group Bias (OGB)', 'Cross Race Effect', 'Other Race Effect', 'Own Race Bias', "Own Race Advantage", or "Other Ethnicity Effect". The author has updated the terminology to inhibit the usage of race in discussions of this topic and so as to better methodologically define this phenomenon. The umbrella term 'White' should not reflect individuals whose familial ancestry is Middle Eastern and North African (MENA) for example. In this way outgroup is more clearly defined in terms of both ethnicity and ancestral geographic ties. The updated terminology also avoids the pitfalls associated with a broader term such as the OGB which encapsulates not only an in-group ethnic-nationality identification advantage but also an own age and own sex advantage to name a few - as there are a multitude of ways through which an in-group can be constructed.

[3] Ethnic is used as an abbreviated term for ethnic-nationality

uncontested in both its theoretical and applied importance (Young et al., 2012). The amount of other or out-group contact is actionable, it can be changed and therefore has the greatest utility (Meissner & Brigham, 2001). The other or out-group contact – OENE relationship is thus crucially important. Equally as important is a consideration of prejudice. While not always studied concurrently to the outgroup contact-OENE relationship, outgroup prejudice, or attitudes, influence both outgroup contact and the size of the OENE (Pettigrew, 2009). Outgroup prejudice, in addition to providing a greater contextual understanding of the interplay between outgroup contact and the OENE, is changeable. Attitudes by their very nature are dynamic. Thus, research that acknowledges outgroup prejudice equally carries applied utility in reducing the extent of the outgroup identification deficit (OENE). Thus, to explore the influence of the mechanisms of the OENE that simultaneously have an applied benefit, in terms of attenuating the observed strength of the OENE, this research aggregated findings on the outgroup contact-OENE, outgroup contact-outgroup prejudice, and outgroup prejudice-OENE relationships.

Outgroup Ethnic-Nationality Effect

Making use of a face recognition paradigm, the first empirical evidence of the OENE was published in 1969 (Malpass & Kravitz, 1969). Since then, numerous studies have sought to explore both the emergence and consistency of this outgroup identification disadvantage. Empirical findings support the OENE as a relatively robust phenomenon, demonstrable across culturally and geographically distinct samples using a wide array of target ethnicities (Meissner & Brigham, 2001). The robustness of the OENE is further supported by consistent findings across the two dominant experimental methodologies used to test identification performance namely, eyewitness and face recognition methodologies[4]. Additionally, the OENE is demonstrable across the lifespan, with greater in-group identification accuracy noted in both child and adult samples.

Within an identification task, the OENE can be conceptualized as (a) a greater number of correct identifications, or hits, for in-group relative to outgroup members; (b) a greater number of incorrect identification decisions, or false-alarms, for outgroup relative to in-group members; (c) a more liberal decision strategy for outgroup relative to in-group members, or a

[4] Eyewitness methodologies differ in the nature of the identification task, which typically encompasses an identification parade or lineup and has fewer targets or to-be-remembered faces than the number of targets used within a face recognition task. Eyewitness tasks typically include multiple or dynamic views of target faces during study and identifications carry greater weight within the context of a mock criminal investigation. By comparison, face recognition tasks typically make use of static front-facing study faces that are tested sequentially with shorter delays between study and testing.

lower response bias; (d) a greater overall identification accuracy for in-group relative to outgroup members, or discrimination[5]; or it can be demonstrated via (e) an identification difference score[6]. While the OENE is consistently demonstrable across various identification indices, the strength of the OENE varies across indices. The effect size for the OENE, specifically for hits, false-alarms and response bias indices is small-to-medium in strength (Lee & Penrod, 2022; Meissner & Brigham, 2001) Comparatively, the effect size for the OENE using discriminability, or overall identification accuracy, is larger with a medium-to-large strength.

The outgroup identification disadvantage emerges in early childhood. While a definitive answer regarding the age of emergence is not yet developed, the OENE is demonstrable in participants under one (Kelly et al., 2007a; Walker & Hewstone, 2006a; Wong et al., 2020). In such cases, i.e. working with infant and child samples, identification ability is inferred based on reaction times. Faster reaction times are used as an indicator of familiarity. Studies using a sample between six to nine months old have demonstrated a consistent OENE with White and Asian target-faces (Xiao et al., 2014). However, the youngest sample to demonstrate the outgroup identification disadvantage were only three months old (Kelly et al., 2007b; Sangrigoli & De Schonen, 2004). Developmentally, the emergence of the OENE is theorized to be reflective of a process of perceptual narrowing (Xiao et al., 2014). This perceptual narrowing is influenced by a child's immediate visual environment and therefore their perceptual familiarity with a set of facial objects. If other or out-group faces are not present in equal quantities in the visual environment, perceptual expertise will mimic this asymmetry (Zhou et al., 2019). As a result, they will be primed, based on experience, to better identify in-group faces.

While the OENE is considered relatively robust, it is influenced by both motivation and as noted above the amount of outgroup contact. Thus, the strength of the OENE can vary across samples. With respect to both motivation and level of outgroup contact, the OENE is consistently demonstrable for majority members who are tested on minority outgroup members (Chien et al., 2018; Wong et al., 2020). Conversely, the OENE is noticeably more inconsistent for minority members who are tested on majority outgroup members. There are

---

[5] The terms 'discrimination' and 'discriminability' have been used interchangeably in this work. Both refer to a participant's overall identification accuracy i.e. d-prime.

[6] Typically, this metric is calculated by subtracting the score for outgroup targets/members from that of the score for in-group members. For all metrics beyond false-alarms, a higher identification difference score signifies a greater OENE

however exceptions to the above noted trends. One such exception is a Black South African sample (Sadozai et al., 2019; Seutloali, 2015; Witter & Tredoux, 2020; Wright et al., 2003). While this sample occupies majority status in terms of population metrics, the unique social and historical context of South Africa influences both the degree of motivation to individuate out-group or minority faces as well as the level of outgroup contact. The persisting structural legacy of Apartheid, a legalized forced segregation and amendment to majority members' and all other-than-white members' rights, is one of persisting patterns of residential segregation and great socio-economic disparity (Southall, 2023). Within the context of the South African economic disparity, minority members typically occupy the highest socioeconomic status and hold the highest degree of wealth when compared to majority members. Primarily Black neighborhoods are typically located further away from city centers and important business hubs resulting in increased travel for majority members in search of economic opportunities. As a result, Black South Africans, albeit majority members, have a higher extrinsic level of motivation to motivate minority faces and often have higher levels of outgroup contact than global north majority counterparts. Consequently, Black South Africans typically do not demonstrate an OENE. The exception to the norm, for majority South African members, highlights the importance of acknowledging all pertinent contextual factors in explorations of the OENE.

Theoretical Underpinnings of the Outgroup Ethnic-Nationality Effect

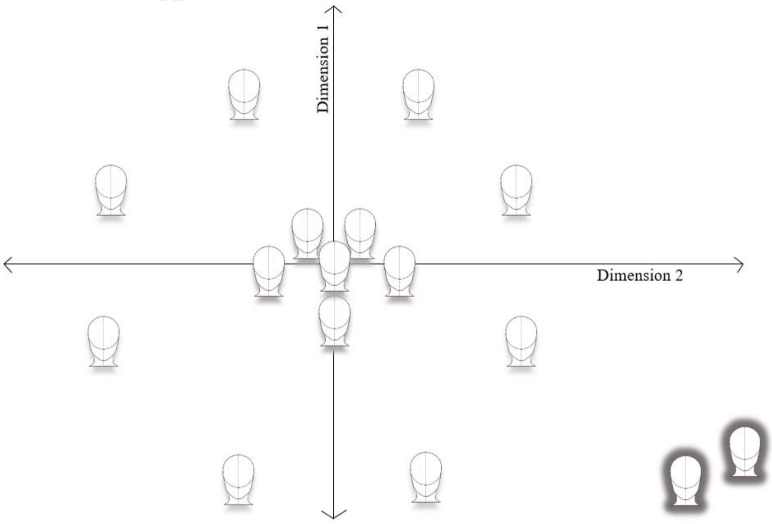
There are competing explanations regarding the emergence of the OENE. The primary theoretical explanations for the outgroup identification deficit include the perceptual expertise model, the socio-cognitive model, and lastly, integrative models that integrate elements of both perceptual expertise and socio-cognitive models (Bernstein et al., 2007; Meissner et al., 2005).

The cornerstone of the perceptual expertise model is perceptual experience and/or an asymmetrical familiarity with in-group faces as a driver for the emergence of the OENE (Young et al., 2012). Under this theoretical framework, faces, like other stimuli class sets i.e. flowers or dogs, require fine grained discrimination in order to be correctly identified. As discrimination is tethered to the amount of experience, identification ability is theorized to improve with greater stimuli experience. With reference to facial objects, a limited amount of exposure, or contact with outgroup faces, and therefore limited outgroup facial expertise drives the emergence of the OENE. Perceptual learning, and therefore expertise, is best illustrated via Valentine's (1991) multi-dimensional 'face-space' model (Valentine et al., 2016). In this model perceptual learning occurs within the context of a Cartesian multidimensional space. The axes of this Cartesian plane are useful dimensions with which to categorize faces. Each novel face experienced is theoretically represented and positioned within the face-space model according to these useful facial dimensions. As per Valentine's face-space, asymmetric in-group exposure produces a well-developed cluster of in-group faces as well as useful dimensions on which to differentiate in-group faces (see Figure 1A). In the model, similar faces group to the mean, while more distinctive faces are positioned at extremes to the mean. When novel other or outgroup faces are introduced, they are represented on the edge of the perceptual face-space as the pre-developed dimensions, on which to differentiate faces, are not yet optimized to effectively discriminate outgroup faces (see Figure 1B).

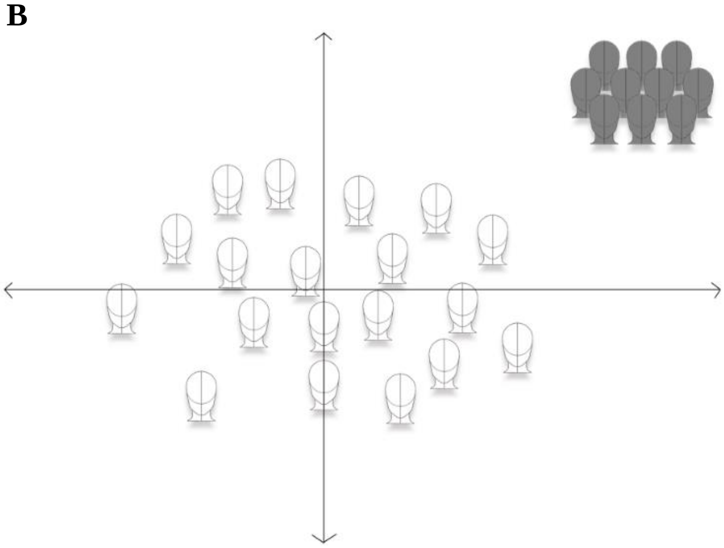
Figure 1

Illustrations of Valentine's Perceptual Face Space Model

A



B



Note. (1A) In this perceptual model typical, or average, faces cluster in the middle of the face space. The dimensions on which the faces are ranked are purposively left blank. These dimensions could be distance between features (large to small) or uniqueness of a feature for example. The exact dimensions utilized are developed in response to the number of perceptual stimuli encountered and the most useful dimensions with which to distinguish them. In this illustration, two distinctive faces are illustrated with a dark grey halo. They are distinct in terms of facial dimensions 2 and 3 and thus are positioned at the edge of the face space. (1B) Spatial distributions of both in and out-group faces are depicted. Out-group faces are more densely populated in a smaller area of space relative to the more open spatial distribution of in-group faces.

As a result, in-group faces tend to be more widely distributed within the face model. By contrast, out-group faces tend to cluster closer together, in a smaller area of space, which hinders efficient out-group discrimination or identification. In order to optimize the global face-space perceptual dimensions, greater out-group contact is needed. Greater perceptual expertise, with both in- and outgroup faces, therefore, facilitates maximal outgroup encoding, storage and retrieval strategies (Meissner et al., 2013). By extension greater outgroup contact is theorized to reduce the OENE.

Socio-cognitive models emphasize the importance of top-down processing on perceptual learning (Hourihan et al., 2013). Attitudes and beliefs determine how bottom-up visual stimuli are processed. More specifically, motivation determines the level of processing for in- and outgroup faces. Greater social importance is typically assigned to in-group members (Meissner et al., 2005). This translates into a greater degree of motivation to individuate in-group faces. Consequently, outgroup faces by extension are categorized as belonging to an 'other' group category and thus the individuating process is theorized to halt. Outgroup identification is therefore limited by the non-processing of the individuating facial details or information needed to accurately identify outgroup faces (Levin, 2000). A key feature of this theoretical model is the fluidity of group membership and identity. As per this theory, group identity or membership is not bound to only ethnicity. Group membership can be formed as a result of any meaningful group affiliation as demonstrated by an out or other university identification deficit or from experimental manipulations of group identity resulting in the traditional OENE for outgroup members (Hourihan et al., 2013; Sadozai et al., 2019).

Integrative models incorporate the key elements of both perceptual expertise and socio-cognitive models (Hugenberg et al., 2010). Instead of prioritizing contact over the role of motivation and top-down processing, or vice versa, this theoretical model considers both processes to be equally important and co-occurring processes that both influence the emergence of the OENE. This is encapsulated in the model's three processing stages which include a (a) categorization of other or outgroup faces as being 'other' which in turn leads to the deployment of configural processing, or rather the homogenization of outgroup faces. Consequently, (b) greater levels of motivation are required to overcome outgroup homogenization. Greater motivation, and thus increased interaction importance, triggers the individuation of outgroup facial features. This fine-grained outgroup discrimination facilitates outgroup identifications. Lastly, (c) perceptual experience with facial objects is acknowledged as an indicator of overall identification ability. In this way integrative models, a comparatively recent addition to theoretical explanations of the OENE, overcome the intrinsic limitations of the prior models which, while valid, can only explain parts of the OENE phenomenon. Such limitations, in terms of the perceptual-expertise model, include the inconsistent and/or small influence of outgroup contact on the size of the outgroup identification deficit (Young et al., 2012). The small strength of the relationship undermines the core tenet of this model. In addition, perceptual-expertise models cannot provide an explanation for the role of motivation in attenuating the size of the OENE. Similarly, socio-cognitive models cannot account for the role of outgroup contact or why motivation to individuate outgroup member manipulations are ineffective when in- and outgroup members are of a similar socioeconomic status. In this way integrative models, such as Hugenberg and Sacco's (2008) 'Individuation-categorization model' provide greater utility in their broader scope and acknowledgement of the role of both outgroup contact and top-down processing.

Contact-Outgroup Ethnic-Nationality Effect

As illustrated via the perceptual-expertise model, greater outgroup contact or outgroup exposure increases overall perceptual expertise (Singh et al., 2021; Young et al., 2012; Zhou et al., 2014). This in turn is theorized to translate into greater outgroup identification gains (Walker & Hewstone, 2006b). The dimensions in Valentine's face space are akin to distinctive features or facial dimensions (Valentine, 1991; Valentine et al., 2016). They could be first or second order features, such as the distance between the bridge of the nose and chin, or holistic dimensions like face roundness. The finer level of discrimination needed to identify a face is dependent on prior knowledge and/or exposure to said facial dimensions or features. Without sufficient out-group exposure, and therefore limited out-group perceptual expertise, the time taken during facial encoding could be ineffectually spent attending to features which do not maximally assist outgroup identification resulting in an outgroup identification deficit (Tullis et al., 2014). Theoretically level of outgroup contact is therefore crucial in explanations of the OENE.

Advancements in telecommunications and travel have facilitated an unprecedentedly interconnected world (Schieferdecker & Wessler, 2017). Given such advancements, outgroup exposure whether by media broadcasting alone, has increased across time. If greater outgroup contact or exposure in and of itself attenuated the size of the outgroup identification deficit there should be an observable decrement in the OENE across time and yet, the OENE remains a robust phenomenon (Sporer, 2001). While higher levels of outgroup contact, and by extension outgroup expertise, should theoretically decrease the size of the OENE across the lifespan, the relationship between outgroup contact and the OENE is more nuanced. The ability to correctly identify faces mimics the process of language acquisition, namely learning phonemes, in that there is a critical period for learning (McKone et al., 2019). During this critical period, exposure to a new class of facial objects, or following the analogy second language phonemes, must occur in order to be maximally beneficial. As following this critical period, a specialization or pruning process occurs (McKone et al., 2019). Therefore, early other or outgroup contact inoculates against the loss of in-group level discrimination for out-group members later in life. Exposure within this critical period alone does not necessarily equate to a high degree of outgroup discrimination, and thus identification proficiency. It does however ensure that future outgroup contact will train perceptual expertise more efficiently. Similar to learning a second language later in life, post critical period, a greater degree of exposure or outgroup contact will be needed to see incremental increases in base outgroup identification performance. Ultimately, learning in later-life will cap the maximum learning potential of outgroup identification accuracy. In this way, the amount of other or outgroup contact, in conjunction with the period in which the contact first occurred influences the magnitude of the OENE. The critical period for beneficial outgroup exposure is theorized to end at twelve years of age (Pascalis et al., 2020). The amount of other or outgroup exposure needed to inoculate against loss of future out-group identification ability, however, remains unclear. Deep artificial neural networks, that are trained to identify faces, also undergo a period of critical development (Wang et al., 2024). In this period, limited exposure to varied faces – in terms of ethnic-nationality, impairs later facial identification ability.

Despite the theoretical importance of outgroup contact on the OENE, evidence of a small effect size for this relationship contributes to the debate regarding the exact pathways that underscore the emergence of the OENE (Singh et al., 2021; Stelter et al., 2022).

Contact-Prejudice

Prejudice is defined as an unsubstantiated pejorative attitude or stereotypically reductionist belief set that stems from an innate characteristic, affiliated identity or group membership (Durrheim et al., 2016; Reicher, 2007). Outgroup prejudice is not solely the result of group membership (Wright et al., 2001). Level of outgroup contact and outgroup prejudice can be conceptualized as having a bi-directional relationship. According to the 'familiar face over-generalization hypothesis' outgroup prejudice can arise following limited outgroup contact (Zebrowitz & Montepare, 2008). Following this theory, outgroup faces would be perceptually atypical, and this atypicality in perceptual expertise automatically elicits unfavourable outgroup attitudes and by extension outgroup prejudice. This reaction is theorized to be attenuated with greater outgroup exposure.

Individuals exhibiting a high level of outgroup prejudice are less likely to engage in outgroup contact for fear of negative outgroup interactions (Brigham & Malpass, 1985). Yet, outgroup contact can demonstrably reduce outgroup prejudice. This is supported by evidence of a small-to-medium effect size for the outgroup contact and outgroup prejudice relationship (Hsieh et al., 2022; Meissner & Brigham, 2001; Zhou et al., 2019). The attenuated level of prejudice in such scenarios is the result of positive outgroup contact increasing favourable outgroup attitudes, leading to a reduction in stereotypical beliefs, which in turn reduce outgroup prejudice (Holtman et al., 2005). The importance of positive contact is encapsulated in Allport's contact theory (Pettigrew & Tropp, 2008). This theory emphasizes the importance of not only the quantity of outgroup contact but also the quality of outgroup interactions both of which are important in necessitating a shift in outgroup prejudice. High quality contact, i.e. both sufficient quantity and quality of outgroup contact, encourages perspective taking, greater outgroup empathy and reduces intergroup anxiety (Pettigrew & Tropp, 2008). Empathy and intergroup anxiety, both affective components of social interactions therefore mediate the relationship between outgroup contact and prejudice.

Contact Theory therefore requires optimal situation specific factors in order for the outgroup contact to be maximally effective. These individual, interaction specific factors include a need to co-operate – the greater the vested interest in the interaction the better; equal power dynamics; and support from authorities (Holtman et al., 2005; Turner et al., 2007a; Wagner et al., 2003). One should recognize though that whereas high quality positive contact can improve attitudes, conversely and as noted above, negative or poor outgroup contact can increase outgroup prejudice. Negative contact can in particular increase inter-group anxiety (Voci & Hewstone, 2003). This anxiety manifests as group level avoidant behaviour. The crux of avoidant behaviour is the belief that the social interactions in question will likely produce an unfavourable outcome. Anxiety leads to an attentional narrowing and a reliance on over simplified social schemata to navigate the social interaction. This in turn can result in self-fulfilling confirmatory practices, which may exacerbate group anxiety, re-affirm negative group level attitudes, stereotypical beliefs and therefore prejudice (Voci & Hewstone, 2003; Walker & Hewstone, 2008). Negative contact can also strengthen in-group identity. All of which increases the likelihood of a greater other or outgroup identification deficit.

Prejudice – Outgroup Ethnic-Nationality Effect

Outgroup prejudice is a broad term which encompasses both implicit and explicit outgroup prejudice. Different processes underpin implicit and explicit outgroup prejudice (Dovidio et al., 2002; Ferguson et al., 2001). Conceptually, implicit and explicit outgroup prejudice can be differentiated in terms of speed and intention. Implicit outgroup prejudice is presumably well-practiced and would therefore be an immediate unconscious process. Explicit outgroup prejudice would be slower by comparison and would therefore require conscious processing. Therefore, in the context of a timed identification task, both the amount of time allocated to the categorization of other or outgroup faces as being ‘other’, as well as the amount of time available to search for featural cues to aid outgroup identification, if motivated, will differ according to the engaged processing level. Theoretically both types of outgroup prejudice would in turn impact the size of the observed OENE. Implicit outgroup prejudice can only be measured indirectly via an evaluation of overall task performance. Implicit outgroup prejudice measures include the Implicit Association Test (IAT; Greenwald et al., 1998). By comparison, measurement of explicit outgroup prejudice involves deliberate outgroup attitudinal assessments that are often completed via self-report. Outgroup prejudice plays an important role in the formation of group associations or identity (Holtman et al., 2005; Turner et al., 2007a). Highly prejudiced individuals are more likely to categorize prejudiced groups as being other or out-group (Agadullina & Lovakov, 2018; Dotsch et al., 2008). This categorization phenomenon will result in the homogenization of the other group, which will reduce the ability to accurately identify outgroup individuals, resulting in a theorized increase in the size of the OENE. Highly prejudiced individuals are also more likely to engage in ‘other’ group contact avoidance or group alienation. This would likely result in the underdevelopment of perceptual expertise for other or outgroup faces and would in turn compound the outgroup identification deficit. While this seems intuitive, studies have reported inconsistent findings or no relationship between outgroup prejudice and the OENE (Ferguson et al., 2001; Meissner & Brigham, 2001).

Forensic and Applied Relevance of the OENE

In 1978, an Ohio resident named William Bernard Jackson was put on trial for the sexual assault of two women (Johnson, 1984; Welnhofer, n.d.). The alleged crime had taken place three years prior to the commencement of criminal proceedings. Both women had been assaulted in their place of residence after a male assailant had illegally gained entry to their respective properties. In the absence of any physical evidence linking Mr. Jackson to the crime scene(s), the prosecution's case solely relied on the women's eyewitness testimony to secure a conviction. Both women had identified Mr. Jackson as the assailant who had forcibly entered their residence. Mr. Jackson maintained his innocence throughout the criminal proceedings and was able to supply numerous alibi witnesses, accounting for his whereabouts on the night(s) in question. Despite such evidence, so compelling was the eyewitness testimony proffered that Mr. Jackson was convicted of both counts of sexual assault and sentenced to a mandatory minimum of fourteen years in prison and a maximum sentence of fifty years. Five years into Mr. Jackson's sentence, new information came to light which corroborated his claim of innocence. At this time an investigation was underway into Dr. Edward Jackson. Dr. Jackson had been caught illegally entering the residences of women. During the course of the investigation, a book was seized which logged the details of every crime Dr. Jackson had committed. The meticulous self-written logs of criminal activity included accounts of crimes for which Mr. Jackson had been convicted and this exculpatory evidence secured an exoneration.

To further frame the contextual importance of this case study, both women who gave eyewitness testimony in the trial of Mr. Jackson were White and both women made a cross-ethnic identification (Johnson, 1984; Welnhofer, n.d.). A side by side visual comparison of both men showed some base level similarities. Namely, both men belonged to the same ethnic group, had an approximate physical build, both had facial hair, and both had relatively short hair. This however, was the extent of their visual similarities. Their resemblance to one another in terms of facial features was rough at best and there was a sizeable age gap of 14 years between both men. Despite their featural dissimilarity, both women were earnestly mistaken in their incorrect other or out-group identifications.

Eyewitness testimony includes both verbal and written testimony regarding the details of the crime as well as identifications. Eyewitness testimony is singular in its ability to not only link the suspect to the criminal act, but also to link the suspect to both the time and place the criminal act occurred. As a result, eyewitness testimony is afforded great weight within criminal proceedings (Garrett et al., 2020). So great is the evidentiary weight of eyewitness testimony that it eclipses the relative weight of forms of direct evidence, such as fingerprints, which can only link the suspect to the place where the crime occurred. This inherent weight underscores the importance of incorrect identifications, which can translate into wrongful convictions, as was the case for Mr. Jackson. Memory in general is fallible and some degree of identification error is to be expected. The trade-off for true convictions, if true convictions could be known, is that some degree of wrongful, or incorrect, convictions will occur. The acceptable base rate for incorrect versus true convictions is for the purview of the criminal justice system and beyond the scope of this research. Framing this problem, however, is a confluence of factors that impact a witness' ability to accurately identify their assailant, which most often is unfamiliar to the witness, making identification that much more difficult. These factors include those related to the crime itself – referred to as estimator variables, the collection of identification evidence - system variables, and witness specific factors (Wells et al., 2020). All of these factors impact the strength of encoding and the witness' retrieval ability. Witness specific factors include ethnic in-group identity. While incorrect identifications occur, a disproportionate number of cross-ethnic misidentifications occur as a direct result of the OENE.

While the true extent of the cross-ethnic identification problem is not known, the work of advocate groups such as the Innocence Project suggests a severe systemic problem. Of the cases handled to date, by the advocacy group the Innocence Project based in the United States of America, 69% of all worked cases involved convictions based on incorrect identification evidence (Innocence Project, n.d.). Of those cases, 42% involved witnesses making an incorrect other or out-group identification. Ground truth cannot always be known. Overturned convictions require new evidence, most often DNA evidence that was not tested previously, to corroborate innocence. Such evidence is not always available, as this is dependent on the type of crime committed or has degraded to an untestable level since the evidence was first collected. This subsample of exonerated cases is therefore only a fraction of the ground truth or actual number of wrongful cross-ethnic identification. The true extent of the cross-ethnic identification problem logically must be larger than reported and is therefore a considerable issue. An examination of the Innocence Project's cross-ethnic identification cases yields a disproportionate minority bias. The ethnic majority in the United States of America is White whilst Black is an ethnic minority along with Hispanic and Asian (Cortiz et al., 2023; Schmid, 2003) Yet, of the overturned wrongful convictions, 58% involved ethnic minority members while only 33% were ethnic majority members (Innocence

Project, n.d.). A Black individual within the United States of America is seven times more likely to be wrongfully convicted when compared to ethnic majority members.

While ground truth, meaning the absolute objective truth, cannot always be knowable within real-world investigative settings, it can however, always be known within experimental settings (Wells et al., 2006, 2020). In experimental settings, the variables of interest are directly manipulated and controlled by the researcher. In real-world settings, wherein detectives are testing their hypothesis that the apprehended suspect is the true perpetrator, all that can be known is (a) whether or not the suspect was identified, the suspect could be innocent or guilty; (b) whether or not a lineup member or filler was identified; or (c) whether a lineup decision was made or not. Aggregate estimates from experimental studies find that eyewitnesses are 1.4 times more likely to accurately identify or get a 'hit' for in-group targets when their performance is compared to 'hits' for other or outgroup targets (Meissner & Brigham, 2001). In the forensic context, 'hits' speak to the likelihood of apprehending the true perpetrator. Aggregate estimates for false- alarms, when an innocent suspect is selected within an identification task, report that other or outgroup targets, false-alarms are 1.56 times more likely than false-alarms for in-group targets (Meissner & Brigham, 2001). Experimental research therefore corroborates the extent of the problem brought to light by the work of the Innocence Project.

Incorrect identifications and more precisely wrongful convictions carry a large fiscal, human and emotional cost (Haney, 2002; Norris et al., 2020; Stevenson, 2006). Wrongful convictions cause psychological distress stemming from the injustice of loss of freedom. Human potential and growth is capped. The wrongfully accused lose years of their lives, years in which they would have had the potential to contribute to society in a meaningful way. Wrongful convictions separate families, creating emotional instability and financial loss if the breadwinner is wrongfully accused. A prison sentence carries an objective fiscal cost. Prison systems are often already overburdened, or more specifically overpopulated, with finite resources (Haney, 2002). Thus, compounding the importance of research into ways to reduce the OENE and consequently, how to reduce cross-ethnic misidentifications.

Wrongful convictions also have a real cost in terms of trust, specifically society's collective trust in the criminal system. When minority ethnic groups are disproportionately wrongfully convicted, social division is fostered that can be tremendously difficult to reverse (Smith & Hattery, 2011). This division can stem from fostered feelings of being unsafe and unsupported. Pragmatically, wrongful convictions do not deter crime. The perpetrator remains free and able to commit further crimes with a bolstered sense of impunity. Further framing the importance of the cross-ethnic identification problem is the difficulty of overturning a prior conviction (Wells et al., 2020). Besides new evidence, typically in the form of DNA, the appeal process is time and labour intensive. It requires capital and access to skilled professionals with the available time to start the appeal process (Stevenson, 2006). This is further impacted by the number of incarcerated who claim innocence (Struve, 2018). The resources are limited. In instances where exoneration is achieved, the wrongfully accused will suffer lifelong reputational damage and will likely remain associated with the accusations made. Exonerations cannot undo that damage.

Rationale, Aims and Hypotheses

In multi-ethnic societies the likelihood that the perpetrator is a member of an ethnic outgroup is significant. When there are high levels of residential segregation, opportunity for outgroup contact is reduced leading to poorer memory and identifications of other groups. Given the high level of wrongful convictions as a result of the OENE, and the associated fiscal and social costs attached to those wrongful convictions, research into the OENE more broadly and changeable mechanisms through which the OENE can be attenuated are crucially important. This research therefore sought to aggregate 53 years of empirical findings on the outgroup contact-OENE, outgroup prejudice-OENE and outgroup contact-outgroup prejudice relationships. Meta-analytic efforts in this area of research rarely include a risk of bias assessment. A critical appraisal of the quality of included studies is necessary to not only ensure methodological rigour but also to validate the strength of meta-analytic findings (Cooper & Dent, 2011). Thus, to address this gap in research, this research developed and employed a tailored risk of bias assessment

Given the interdependent nature of outgroup contact, outgroup prejudice and OENE relationships, it is crucially important to assess all relationships simultaneously. Explorations of the influence of outgroup prejudice have previously been limited due to a limited sample that both measures and reports of outgroup prejudice metrics. In order to address this limitation, this research used matched-to-sample implicit and explicit outgroup prejudice data from a global dataset to explore the moderating role of outgroup prejudice. The inclusion of such data facilitated an exploration of the variation of the strength of the relationship between outgroup contact and the OENE as a result of increased outgroup prejudice. Such an exploration is important given the counterintuitive prior findings of a small effect i.e. small influence of outgroup contact on the OENE (Stelter et al., 2022).

Hypotheses

It was therefore hypothesized that -

1. OENE validation checks will demonstrate the presence of an OENE. Specifically,
    a. Identification accuracy will be higher for in-group members when compared to outgroup members (Lee & Penrod, 2022; Meissner & Brigham, 2001; Wilson et al., 2013). More specifically, in-group members will exhibit higher hit rates when identifying in-group targets.
    b. Identification errors will be higher for outgroup members when compared to in-group members (Evans et al., 2009; Howard et al., 2019; Lee & Penrod, 2021; Meissner & Brigham, 2001; Wilson et al., 2013). More specifically, in-group members will exhibit higher false-alarm rates when identifying outgroup targets
    c. Overall identification accuracy will be higher for in-group members when compared to outgroup members (Howard et al., 2019; Jackiw et al., 2008; Lee & Penrod, 2022; Pica et al., 2015; Semplonius & Mondloch, 2013; Wilson et al., 2013). More specifically, In-group members will exhibit higher discriminability when identifying in-group targets
    d. The identification decision-making strategy will be more liberal for outgroup members when compared to in-group members i.e. the willingness to choose an outgroup member within the identification task will be greater (Jackiw et al., 2008; Lee & Penrod, 2022; Meissner & Brigham, 2001; Meissner et al., 2005; Slone et al., 2000; Wilson et al., 2013). More specifically, in-group members will exhibit higher response bias rates, signaling a lower willingness to choose in-group targets. This will be more apparent when participants are majority members making an identification decision for minority members (Meissner & Brigham, 2001).
2. For both quantity and quality of outgroup contact,
    a. Greater outgroup contact will be associated with an improvement in outgroup identification accuracy (Bukach et al., 2012; Hancock & Rhodes, 2008; Singh et al., 2021; Stelter et al., 2022; Walker & Hewstone, 2006b; Zhao et al., 2014b). More specifically, an increase in outgroup contact will be associated with a reduction in the OENE or identification difference scores.

b. Greater outgroup contact will be associated with an improvement in overall outgroup identification accuracy (Chiroro, 1994; Pezdek et al., 2003; Meissner & Brigham, 2001; Slone et al., 2000; Wright et al., 2003). More specifically, higher outgroup contact will be associated with increased outgroup discriminability.
c. Greater outgroup contact will be associated with an improvement in outgroup identification accuracy (Chiroro, 1994). More specifically, higher outgroup contact will be associated with higher outgroup hit rates
d. Greater outgroup contact will be associated with a lower willingness to choose outgroup targets i.e. a less liberal decision-making strategy. More specifically, higher outgroup contact will be associated with higher outgroup target response bias rates.
e. Greater outgroup contact will be associated with fewer outgroup identification errors (Chiroro, 1994). More specifically, higher outgroup contact will be associated with lower outgroup false alarm rates

3. For both implicit and explicit outgroup prejudice,
   a. Greater outgroup prejudice will be associated with poorer outgroup identification ability (Ma et al., 2011). More specifically, higher outgroup prejudice will be associated with larger identification difference scores. An increase in outgroup prejudice will be associated with a decrease in outgroup discrimination
   b. Greater outgroup prejudice will be associated with a reduction in outgroup overall identification accuracy (Walker & Hewstone, 2008). More specifically, higher outgroup prejudice will be associated with lower outgroup target discriminability
4. For both implicit and explicit outgroup prejudice,
   a. Greater outgroup prejudice will be associated with reduced outgroup contact or conversely, lower outgroup prejudice will be associated with higher outgroup contact (Brigham & Malpass, 1985, Pettigrew & Tropp, 2006; Walker & Hewstone, 2008).

CHAPTER 2: METHOD

Ethical Approval

Ethical approval was obtained from the University of Cape Town, Psychology Department (see Appendix A)

Search Strategies and Article Selection

The primary considerations for article selection were that the articles (a) contained an identification task, (b) manipulated or measured quantity of outgroup contact, (c) reported a quantity of outgroup contact-OENE correlation coefficient or sufficient data to calculate it, and (d) reported original experimental data. Thus, while quality of outgroup contact was recorded where available, the primary consideration for inclusion was the availability of quantity of outgroup contact data. Using this approach, observations could be made regarding (a) the frequency with which outgroup prejudice data is measured and reported alongside outgroup contact and OENE data, and (b) the frequency with which quality of outgroup contact is reported alongside quantity of outgroup contact data.

Articles were identified via a search of databases via platform. More specifically, 'ProQuest Dissertation & Theses Global', 'Google Scholar', 'PsycArticles', 'PsycINFO', 'Academic Search Premier', 'PubMed', 'Scopus', 'Web of Science', and 'Science Direct' were searched. The primary search string used included (a) historical derivations of terms describing the OENE, (b) variations of 'identification task' and 'recognition', and (c) variations and synonyms for out-group contact. A detailed list of search strings and Boolean operators used can be found in 'Appendix B'. Search parameters included English-only articles, a scan of key search terms in the abstract for a refined search, and publication date limiters of 01 January 1969[7] to 31 October 2022. Both published and unpublished works were considered.

The initial search yielded 354 articles (see Figure 2, the PRISMA Flowchart). After duplicates were removed, 174 articles remained. Abstracts of the remaining articles were manually scanned for relevancy. Articles which passed this first screening were sought for retrieval. The retrieved articles (n = 86) underwent a second screening process to assess eligibility. During this process the entirety of the article was considered. Articles that only

[7] The first empirical finding of the OENE was published in 1969. Thus, all articles since the year of the first publication up to the end of October 2022 were considered.

reported reaction time, in milliseconds, in lieu of traditional identification data or metrics were excluded.

With the primary inclusion focus being the quantity of outgroup contact-OENE relationship, articles which only reported outgroup identification data in addition to the outgroup quantity of contact-OENE relationship or the means to calculate it, were included. Manipulations of the target, or the to-be-remembered face's, ethnicity were limited to only within-participant manipulations. Thus, any between-participant manipulations of target ethnicity were excluded[8]. Articles which sampled two or more ethnic groups and reported collapsed results, i.e. same, or in-group, versus other, or outgroup, were not excluded from the analysis. Articles flagged as warranting exclusion were discussed and consensus was reached before final exclusion[9]. A list of excluded studies accompanied by reasons for exclusion is found in 'Appendix C' while a summary thereof can be found in 'Appendix D'. A total of 37 articles were retained.

A second stream of articles was identified via a manual reference list scan of the articles retrieved in the primary search (see Figure 3). The identified articles underwent the same screening process as listed above. A list of stream two excluded studies accompanied by reasons for exclusion can be found in 'Appendix E' while a summary thereof can be found in 'Appendix F'. The combination of search streams yielded a final sample of 42 articles.

Characteristics of the Final Sample

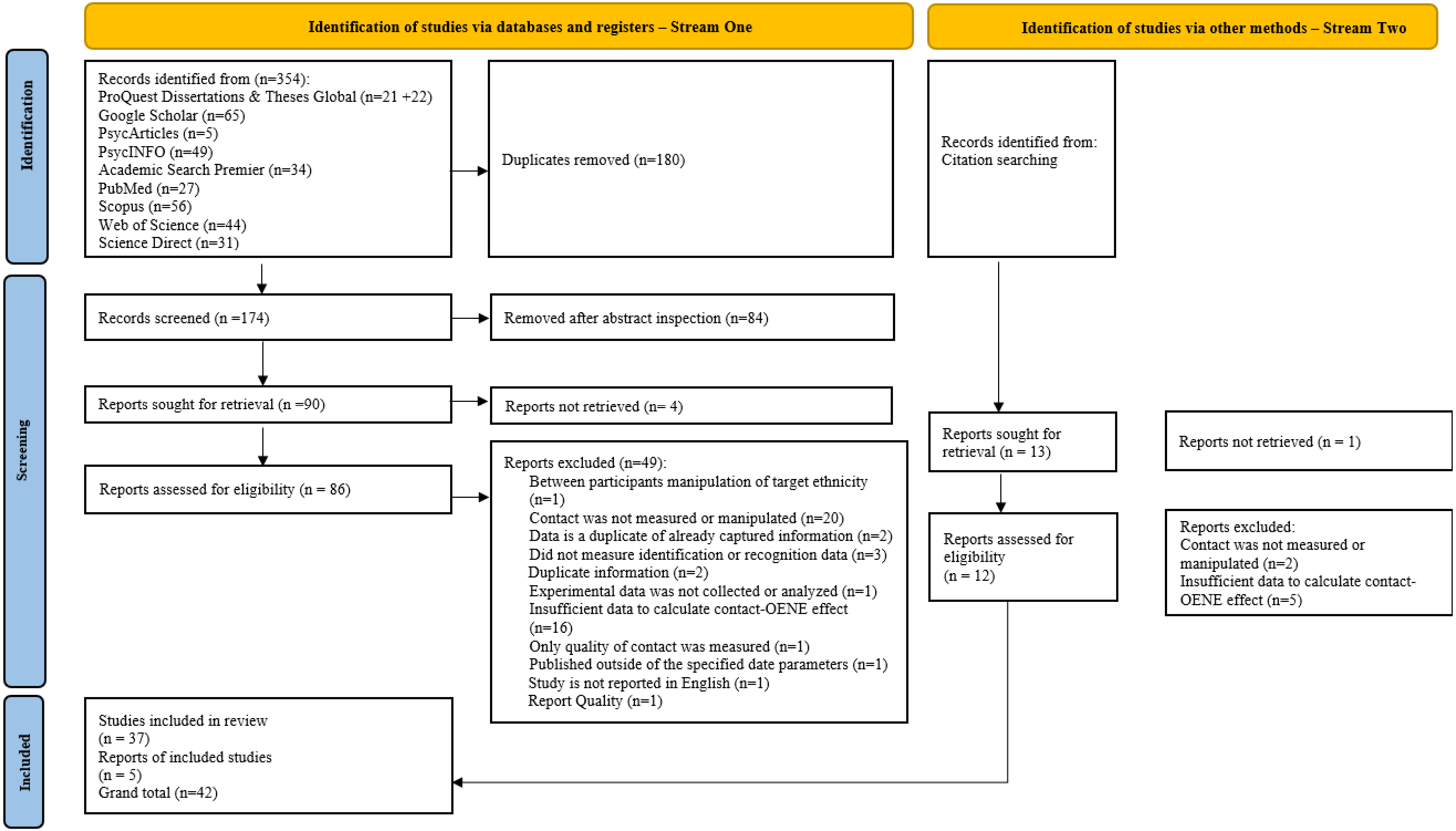
The final sample included 33 published articles, including traditional journals (n=31) and published conference papers (n=2), and 9 unpublished dissertations or theses. Publication dates ranged from 1976 to 2022. All included works had at least one study that met the stipulated inclusion criteria. Some of the included works however, had multiple studies that met the inclusion criteria. As a result, data from 8572 participants spanning 55 studies were extracted.

[8] Experimental studies of the OENE typically make use of within-participant manipulations, otherwise referred to as a repeated measure design. The decision to exclude between-participant, or independent sample designs, from the analysis stemmed from the need for design consistency. Calculations of OENE effect sizes (standardized mean difference) must account for the experimental design employed (Cooper et al., 2019). If the inherent research design is not factored into calculations, the variance of the aggregate effect will be incorrect. The repeated measures nature of OENE tasks is rarely factored into the analysis owing to the difficulty of back computing necessary in- and outgroup correlations. This analysis therefore ensured design consistency of the analyzed articles and this is reflective in OENE effect calculations.

[9] Articles flagged as potentially warranting exclusion were reviewed by the researcher's supervisor.

Figure 2

Overview of the Search, Screening and Selection Protocol for Included Articles



Both in-view identification, or in-view matching tasks, and delayed identification tasks were included in the sample. Conceptually however, both tasks are different. Delayed identification tasks include a memory component while in-view identifications do not. Data from such tasks should therefore not be analyzed together. A meaningful sub-analysis of in-view only tasks, or tasks without a memory component, was not permissible owing to the small sample size (n=1 articles/studies). Consequently, this data was excluded from the primary analysis. The primary analysis therefore focused on any face recognition or eyewitness identification task that had a memory component[10] i.e. had a delay between encoding and testing; and (a) made use of data from 8418 participants spanning 54 studies or (b) 41 articles.

Coding and Preparation

Data was coded and stored in a Microsoft Excel file. A coding template was specifically created for the analysis. The coded dataset included (a) base information such as author, year of publication, study number, sample size and gender etc.; (b) base identification data for in- and outgroup members i.e. hits, false-alarms, discrimination etc.; (c) effect sizes for all relationships of interest and base recognition data; and (d) theoretical and/or methodological moderator variables of interest such as time in the lifespan when contact occurred or length of target face encoding time etc. Supplemental data was used when both applicable and available. If the data was not explicitly reported in text, tables or supplemental analyses but, was available graphically, the necessary information was extracted from graphs using an internet-based application, WebPlotDigitizer (Rohatgi, 2022).

Each line in the coding sheet refers to a specific article/sample and all the pertinent information for the sample is coded within this row. Traditionally there would be no overlap of samples. This was not possible in this analysis owing to the multiple potential contact-OENE effects reported for a given sample, wherein for example, effects could be reported for amount of contact in both childhood and adulthood. In such instances, the base data was duplicated and the alternate contact-OENE effect and relevant individual means were slotted in. To account for this, variables coding the uniqueness of coded data per section were included. Depending on the specific analysis, data could be filtered accordingly while still retaining the complete set of contact-OENE effects.

[10] Eyewitness data such as lineup tasks were analyzed concurrently with face recognition tasks such Old-New tasks. Delayed matching tasks were included in the analysis.

Hits, false-alarms, prejudice and contact data were all converted to proportions. Proportions for outgroup prejudice and contact data facilitated meaningful comparisons across measures utilizing different scales and therefore different maximum scores. Similarly, the number of identified faces for hits or false alarms is dependent on the number of new or old target faces used and thus proportions facilitate meaningful comparisons across studies with varying numbers of target faces. Sample descriptive statistics such as participant's mean age and the associated standard deviation were not always reported. Thus, whether or not this data was reported was captured and when it was not available, a proxy sample age was captured based on the given sample description. In such instances university census data for undergraduate or postgraduate samples in the relevant years was searched to infer a suitable proximate age.

Different analyses, such as hits or the relationship between outgroup contact and prejudice, often make use of different or varying sample sizes. As analyses require complete data, and while not always explicitly acknowledged, listwise deletion is often employed to circumvent the problem of missing data. This is reflected in the degrees of freedom reported. In order to account for fluctuating sample samples when calculating effect sizes, each row in the dataset contained multiple sample sizes – i.e. one for each analysis.

Moderator Coding and Preparation

Moderators such as length of encoding time and delay between encoding and identification were standardized to common unit of measurement, namely, seconds and minutes respectively. If delay was not explicitly reported, an approximated delay was included based on the reported procedural information. A variable coding delay approximations was included and thus the data could be filtered to include all or original reported values only[11]. Sample country of testing and in- and outgroup target pairings were used to construct a variable coding the positionality of the sample relative to outgroup target faces i.e. majority members tested on minority outgroup faces etc. In addition, sample country of testing was used to construct a variable coding global positionality in terms of the global south and global north.

Period of time in which outgroup contact occurred was used to create a variable coding exposure during the critical period of outgroup contact (<=12) or late exposure[12]. This

[11] No delay, or immediate testing, was coded as 0.025 minutes (1.5 second delay).

[12] An associated variable for quality of out-group contact was not included in the analysis owing to limited usable data and therefore limited additive value (n=2 for the critical period, and n=2 for late life exposure)

was used in conjunction with quantity of out-group contact scores to construct a variable coding critical-high, critical-low, late exposure-high, and a late exposure-low contact. The type of task used to assess identification performance was used to construct a variable coding task complexity or cognitive load (low, high).

Outcome Measures or Effects

Base identification data, for OENE validation checks, used a standardized mean difference effect size which accounted for the repeated measures design of the sampled articles. The effect size used was Hedges g. Hedges g was chosen to correct for the potential bias introduced from smaller samples. Cohen's D, as seen in (1), was calculated by subtracting the mean outgroup value ($\bar{X}_2$) from the mean in-group value ($\bar{X}_1$) and dividing by the in-group standard deviation[13]. A positive effect size therefore indicates a higher in-group score.

$$\mathrm{d} = \frac{\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2}{\mathrm{s}_{\mathrm{x}_1}} \tag{1}$$

$$\mathrm{J} = \left(1 - \frac{3}{4 \times \mathrm{n} - 1}\right) \tag{2}$$

A correction factor was calculated (2), and multiplied with Cohen's D, to produce Hedge's g. To calculate variance, where possible, correlations for in- and outgroup scores were back computed (see (3), (4) and (5) for steps or (6) for full formula).

$$\mathrm{y} = \sqrt{\mathrm{n}}(\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2) \tag{3}$$

$$y_2 = \left(\frac{y}{t_D}\right)^2 - s_{x_1}^2 - s_{x_2}^2 \tag{4}$$

$$\mathrm{r}_{\mathrm{x}_1\mathrm{x}_2} = \frac{\mathrm{y}_2}{-2\mathrm{s}_{\mathrm{x}_1}\mathrm{s}_{\mathrm{x}_2}} \tag{5}$$

$$r_{x_1x_2} = \frac{\left(\frac{\sqrt{n}(\overline{x_1 - x_2})}{t_D}\right)^2 - s_{x_1}^2 - s_{x_2}^2}{-2s_{x_1}s_{x_2}} \tag{6}$$

A suitable *t* stat was required, or an appropriate *F* statistic which could be converted to a *t*. When this data was not available, and thus back computing an in- and outgroup

[13] Typically, pooled standard deviation is used as a denominator. As this is a repeated measures design, the standard deviation for in-group targets / group 1 was used.

correlation was not possible, a correlation of 0.5 was substituted into the variance formula (7)[14].

$$v_d = J^2 \times \left(\frac{2(1-r)}{n} + \frac{d^2}{2n}\right) \tag{7}$$

The outgroup contact-OENE, outgroup prejudice-OENE and outgroup prejudice-contact relationships were computed with a Fishers Z transformed Pearson's correlation coefficient as the primary effect size[15]. Standardization also facilitated the calculation of variance (8). While Fisher's Z transformed coefficients were used in meta-analytic models and reported, the aggregate effect per relationship of interest, was also transformed back into a Pearson's correlation coefficient and reported narratively.

$$\mathrm{V_Z} = \frac{1}{\mathrm{n} - 3} \tag{8}$$

If a pre-existing participant variable was used to infer contact such as geographic location, in other words a quasi-experimental design with high and low outgroup contact groups was assumed, a standardized mean difference score was calculated, transformed into a Pearson's correlation coefficient, and finally transformed using Fisher's Z transformation. The formula used to calculate Cohen's D used the pooled standard deviation as the denominator. In this way the between-participant nature of the design was accounted for. A positive effect indicated higher values for the high outgroup contact group. Hedge's g correction was applied. The conversion of standardized mean difference effect scores into Pearson's correlation coefficients accounted for whether or not the sample size was equal among the groups. See (9) for equal groups conversion and (10), (11) for unequal groups conversion (Aaron et al., 1998; Cooper et al., 2019).

$$\mathrm{r} = \frac{\mathrm{d}}{(4 + \mathrm{d}^2)} \tag{9}$$

$$\mathrm{y} = \frac{\mathrm{n_1 n_2}}{(\mathrm{n_1} + \mathrm{n_2})} \tag{10}$$

$$\mathrm{r} = \frac{\mathrm{d}}{\sqrt{\frac{\mathrm{d}^2 + (\mathrm{n_1} + \mathrm{n_2} - 2)}{\mathrm{y}}}} \tag{11}$$

[14] The variance formula requires Hedges g (denoted by *d; Cooper et al., 2019*)

[15] Meta-analytic procedures often assume a sampling distribution which is normal. The sampling distribution of the correlation, or raw, coefficients may not be approximately normally distributed. If samples are relatively small in size this could lead to a skewed distribution. Thus, to correct for this, coefficients were normalized via a Fishers Z transformation.

There are limitations to conversions between standardized mean difference and correlation effect sizes, particularly in terms of benchmarking the strength of the correlation (McGrath & Meyer, 2006). Correlations are impacted by base rates and as a result large *d* effects when converted do not always translate into equally high correlations, thus potentially underestimating the strength of the effect (McGrath & Meyer, 2006). A meta-analysis requires comparable effects and conversion to a standardized effect will always introduce some degree of bias. This potential bias introduced from conversion is argued to be preferable to the potentially greater bias that could be introduced when study data, using different effects, such as *d*, are omitted from the analysis.

Merged Prejudice Data

Outgroup prejudice data is not always captured and reported in studies assessing the outgroup contact-OENE relationships. If it is reported, explicit prejudice only could be reported, or vice versa. Thus, to facilitate the analysis of all the relationships of interest, prejudice data from Harvard's Project Implicit database was merged into the dataset (Xu et al., 2014a; Xu et al., 2014b). The Project Implicit database contains both implicit and explicit prejudice scores for participants sampled across the world. The data includes (a) participant characteristics such as age, ethnicity, gender; (b) the year prejudice was measured; and (c) the geographic location of the participants sampled. Project Implicit datasets were recovered for the years 2002 – 2022.

As noted previously, in the absence of reported sample descriptive statistics i.e. age, proxy descriptive statistics were inferred from the available sample description. Complete sample descriptive statistics were necessary for matching merged outgroup prejudice data to the relevant participant sample.

Outgroup prejudice data was matched to the year of publication. In the case of studies being published prior to 2002, and thus preceding the available Project Implicit datasets, outgroup prejudice data from the 2002 dataset was used. Outgroup prejudice data matched the ethnic-nationality and age of participants sampled. Merged outgroup prejudice data was age matched in five year intervals above and below the sample mean, i.e. a mean sample age of 21.05 utilized merged outgroup prejudice data from participants between the ages of 16 and 26[16]. Outgroup prejudice data was also matched to the geographic location of the

[16] When perfect age matched merged outgroup prejudice data was not available, the intervals around sample age were incrementally widened until a suitable sample was found. The degree of accuracy regarding the match was therefore coded.

sample. Geographic match specificity was greatest for samples within the United States. Such samples could be matched at the level of metropolitan statistical area. Countries outside of the United States were matched at the smallest level available, which was at the country level. The degree of match specificity for age, geographic location and ethnic-nationality were recorded.

Explicit outgroup prejudice scores in the Project Implicit dataset were coded counter intuitively. High scores signify lower outgroup prejudice, or more favourable outgroup attitudes. As there was also variation in the meaning of a higher outgroup prejudice score amongst reported data, all outgroup explicit prejudice scores were shifted to a consistent scale i.e. higher scores being indicative of lower explicit outgroup prejudice.

To account for variations in scales used, for both reported and Project Implicit data, all outgroup prejudice data were recorded as proportions i.e. observed score divided by scale maximum.

Implicit outgroup prejudice scores are interpreted according to both the reference category used and the sign, or direction, of the score. A positive score on a White-Black IAT using the reference category of ‘White equals good’, would be interpreted as having a positive bias or more favourable subconscious associations with a White ethnic-nationality. As reference categories may be inconsistent across IATs, all implicit outgroup prejudice data were transformed to ensure the reference category was always for in-group members. In this way, higher IAT scores are indicative a greater or more favourable in-group preference[17].

Ethnic-Nationality

Of particular importance to this analysis, is the delineation of ethnic-nationality for in and out-group target faces. The associated labels were dependent on the articles sampled and by no means are exhaustive. The location and description of participants sampled along with stimuli descriptions were used to create the ethnic-nationality labels. Faces were categorized as belonging to one of eight ethnic-nationalities or given the label ‘same’ or ‘other’ (see Table 1). In this way, results collapsed across participant samples could be easily differentiated in the analysis.

[17] Reported implicit IAT scores of exactly zero were recoded as .001 (n= 3 effect lines). This did not change the overall interpretation of the score and ensured all analyzed implicit values were on the same scale as having some degree of implicit prejudice i.e. a non-zero score.

Agreement and Risk of Bias

A sample of the retained articles was independently coded by a second coder (n=19, or 46.34% of articles sampled). Owing to the continuous nature of the coded data, level of agreement between coders was assessed via an intra-class correlation (ICC). The average ICC (2, 2) was 0.99 with a 95 percent confidence interval of .99 to 1.00 ($F(96, 96) = 181.34$, $p < .001$). This indicates an almost perfect level of agreement between coders. Owing to the high degree of agreement, the full sample was not independently assessed.

Table 1

Description of Ethnic-Nationality Classifications used

Label	(a) Self-identified as and/or (b) geographic and ancestral ties to -
White	Europe (excluding the European countries listed below)
Black	Africa
Hispanic	Mexico, Puerto Rico, Cuba (includes all Latin America and Caribbean), any country with Spanish as a primary language and/or Spanish culture
Middle East and North Africa (MENA)	All North African countries such as Egypt, Middle East, Turkey
Asian	East Asian countries i.e. China, Japan, Korea, Thailand, Cambodia, Vietnam, Malaysia, Laos
Indian	Indian sub-continent
Slavic	Ukraine, Russia, Serbia, Poland, Republic of Bosnia and Herzegovina, Serbia, Croatia, Belarus
Indigenous People of Canada (IPOC)	Indigenous Canadians i.e. First Nation
*Same	-
*Other	-

A risk of bias (RoB) assessment is not typically conducted for meta-analyses in this area of research. Without one, and therefore no means of knowing whether the included studies were methodologically sound or trustworthy, the associated meta-analytic model findings carry less evidentiary weight. A Cochrane's RoB assessment for a repeated measures design is available however, this assessment is tailored for use by health professionals (Higgins et al., 2023). It is therefore not suitable as is for studies using a face recognition or eyewitness methodology. A tailored RoB assessment was therefore developed for use within this analysis.

The developed RoB assessment comprises five domains (see Table 2). These domains test the quality of the design implemented, i.e. whether novel faces were used and thus whether memory for a given face was tested or whether there is the potential for carry-over effects from not utilizing a randomized or counterbalanced design when testing target faces etc. For each domain, articles received a score (see Table 2 & Appendix H). These scores were tallied, and the total score indicates the overall RoB for an article. A total RoB score lower than or equal to two indicates a low risk of bias, a total score of three or four indicates a medium risk of bias, and a total score greater than or equal to five indicates a high risk of bias.

In addition to the overall study classification, Figure 3 demonstrates a percentage score breakdown across studies sampled per domain. For each domain, the green band indicates the percentage of studies that scored 0 or who were classified as having low to no RoB. Conversely, for each domain, the yellow band indicates the percentage of studies that scored 1 or 2 in their respective domains, signaling the presence of some bias. Of the 54 studies included in the primary analysis, all but one were classified as having a low RoB (see Figure 3). Therefore, in terms of overall quality of the articles sampled, all articles were deemed to be of sufficient quality to warrant continued inclusion in the analysis.

While the quality of articles overall was deemed acceptable, it is worth noting that the majority of the studies sampled did not use novel faces within the identification task and did not implement a full cross-over design (see Figure 3). A full cross-over design is achieved when two groups of participants are sampled, both of different ethnic-nationalities i.e. ethnic-nationality A and B, and both samples are tested using target faces from ethnic-nationality A and B. Methodologically a full cross-over design is the gold standard for identification tasks assessing the OENE (Heesen, 2020; Lim & In 2021). Such designs facilitate a direct comparison between ethnic-nationalities sampled in terms of in- and outgroup identification performance. It is therefore concerning that the majority of cases do not adopt this design.

Figure 3

Overview of Findings for the Risk of Bias Assessment

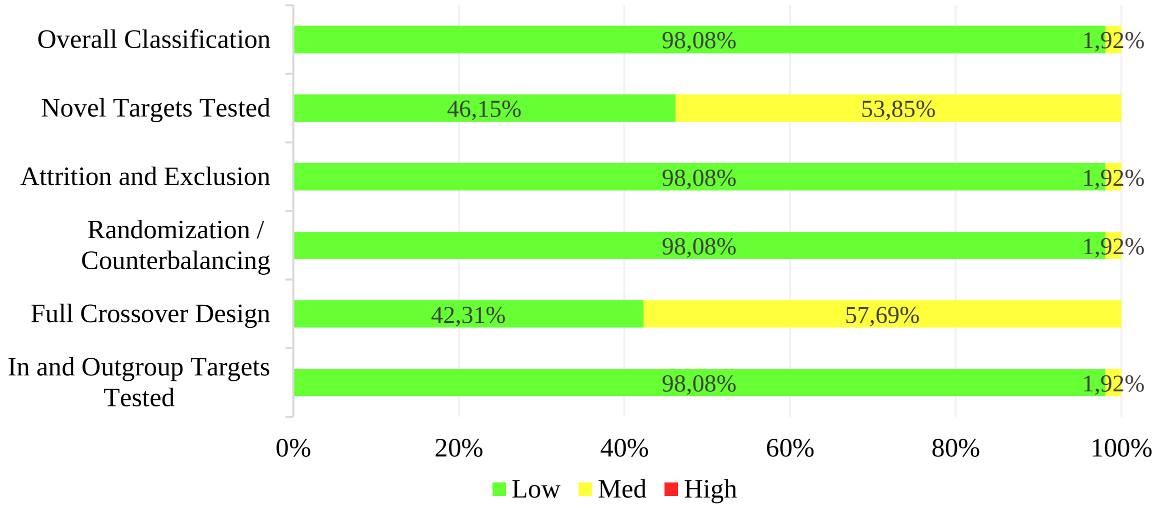


Table 2

Scoring Guide used in the Risk of Bias Assessment

Domain	Question	Point Allocation
Domain One	*Were in and out-group target faces tested?*	
	In- and out-group target faces were tested	0
	Only out-group target faces were tested	1
Domain Two	*Was a full cross-over design used?*	
	Yes	0
	No	1
Domain Three	*Was testing order randomized or counterbalanced for target ethnicity?*	
	If target faces were tested in blocks and those blocks were randomized or counterbalanced; or if target faces were mixed at testing	0
	Unclear OR if only out-group target faces were tested	1
	If target faces were tested in blocks and no randomization or counterbalancing was used	2
Domain Four	*Was participant attrition / excluded data reported?*	
	No one withdrew or were excluded; or the number of excluded participants reported alongside reasons for exclusion	0
	Number of excluded participants reported without reasons for exclusion; or reasons for exclusion reported without number excluded	1

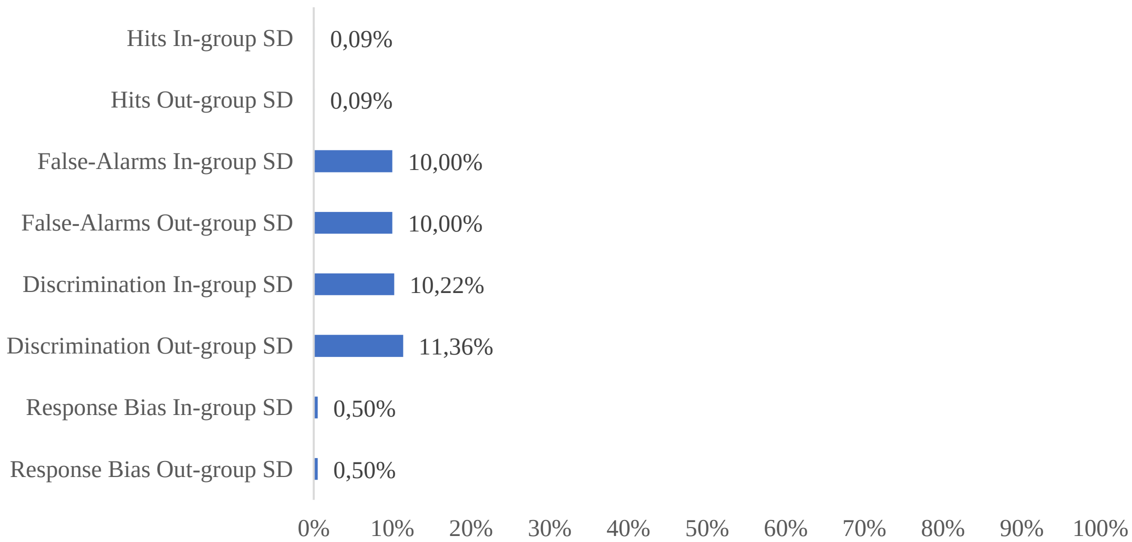
	Number of excluded participants and associated reasons for exclusion were not reported	2
Domain Five	*Were novel target faces used at testing? i.e. were the tested faces different in some way i.e. different pose, clothing, picture*	
	Yes	0
	No; or unclear	1

Missing Data and Imputation

To validate the presence of the OENE in the studies sampled, identification data, namely hits, false-alarms, discrimination and response bias, were coded. This data was then used to calculate a standardized mean difference effect score (Hedges g). The calculation of this effect required standard deviations to be known. Standard deviations were not however, always available (see Appendix I & Figure 4). This limited the number of calculable effects. The meta-analytic OENE validation check therefore made use of two concurrent analyses. Analysis one used listwise deletion to analyze only the reported data which was complete, in terms of both means and standard deviations. Analysis two used multiple imputation to impute the missing standard deviations, which were then used to calculate an effect size. In this way analysis two maximally made use of the full dataset and the potential bias introduced by only using complete reported data was controlled.

Figure 4

Percentage (%) of Missing Standard Deviations for Identification Data



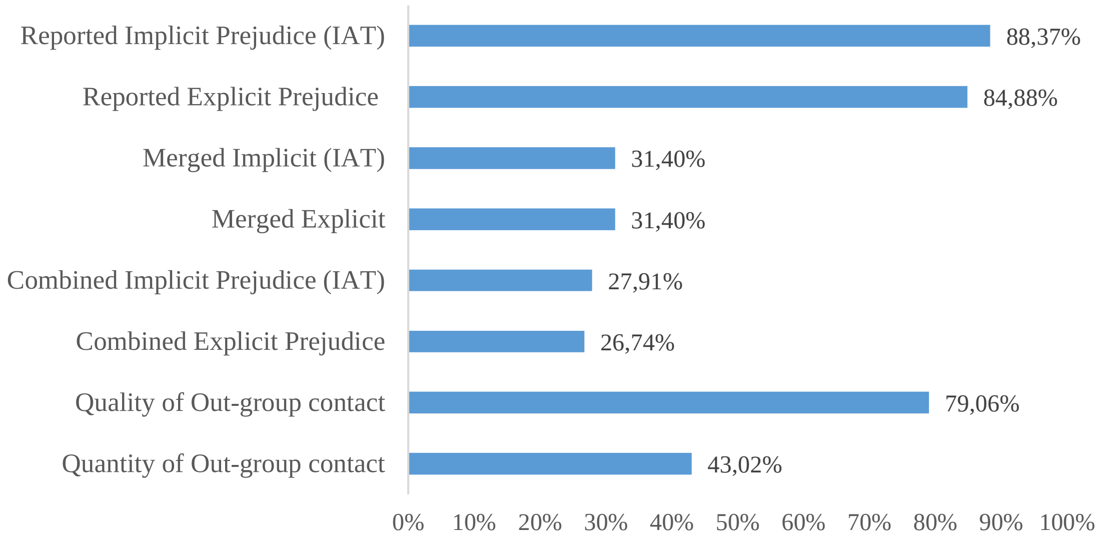
Multiple imputation was conducted in *R*, using the package *MICE* (Van Buuren & Groothuis-Oudshoorn, 2011). While the percentage of, or degree of, missing data is not the sole determinant for the efficacy of running multiple imputations, it is worth noting that the percentage of missing in-group standard deviations was relatively low and below the 40% suggested maximum guideline for effective imputations (Madley-Dowd et al., 2019). The highest percentage of missing data for standard deviations was only 10.22%. All imputations used predictive mean matching, 20 iterations, and all were thoroughly checked for plausibility before inclusion in the analysis (see Appendix J)

The outgroup contact-OENE-prejudice relationships were central to all analyses. As a result, all meta-analytic models, used relevant outgroup contact and prejudice variables listed in Figure 6 as moderators. For example, all outgroup contact and prejudice moderators were used in OENE validation checks, while a model of the quantity of outgroup contact-OENE relationship used the proportion of quality of outgroup contact and prejudice variables as moderators. Inspection of the core moderators revealed a high degree of missing data for the prejudice variables (see Figure 5). The analysis therefore used a combined variable wherein if reported values were present, reported outgroup prejudice values were used. For cases in which outgroup prejudice values were not measured or reported, the Project Implicit merged prejudice data was supplemented into the analysis.

Outgroup contact and prejudice scores, owing to their centrality in the analysis were completed via multiple imputation. The same methods used for imputation of standard deviations were used, i.e. predictive mean matching. All moderator imputations were thoroughly checked for reasonableness and plausibility (see Appendix K). Combined outgroup prejudice variables were below the suggested 40% guideline for imputing missing values (Madley-Dowd et al., 2019). Quantity of outgroup contact is on the cusp of this guideline (see Figure 5), whilst the percentage of missing values for quality of outgroup contact was exceedingly high. Thus, while deemed plausible for use, quality of outgroup contact as a moderator was assessed with tentative caution.

Figure 5

Percentage (%) of Incomplete Prejudice Data



Therefore, both implicit and explicit outgroup prejudice data included (a) reported scores; (b) merged or Project Implicit data; (c) combined scores; (d) merged scores that were completed via multiple imputation; and a (e) a second combined score that used reported values where available or values from the merged scores, that were completed via multiple imputation. Quality and quantity of out-group contact data included both a (a) reported score and (b) a score that was completed via multiple imputation.

Other methodological and/or theoretical moderators were used in the analysis. They were not however completed via multiple imputation if missing values were present.

Additional

Outgroup contact and prejudice variables that were completed via multiple imputation were used in a supplemental analysis wherein Pearson's correlations between identification data, out-group contact and prejudice were calculated. This analysis used only original or reported identification data and facilitated an additional comparison between the calculated correlations and the aggregate effects for the given relationship from the meta-analytic models.[18]

[18] The inclusion thereof facilitated an important point of reference for meta-analytic models with a small sample size.

Meta-Analysis Procedures

All data was analyzed in *R* using the *metafor* package (Viechtbauer, 2010). Multi-variate (hierarchical) modelling was used to account for the dependency introduced by the potential for multiple samples, such as White and Black participant samples, being present within a single study. Samples were therefore nested within the parent study.

A heterogeneity statistic, or percentage of heterogeneity score, was calculated (12). This statistic reflects how much of the observed variability in effect sizes is the result of heterogeneity as opposed to sampling error (Higgins et al., 2023). With higher scores indicating greater heterogeneity.

$$\mathrm{I}^2 = \left(\frac{\mathrm{Q} - \mathrm{df}}{\mathrm{Q}}\right) \times 100 \qquad (12)$$

Outliers and Influential Cases

Diagnostics were checked for all meta-analyses. Outliers were assessed via an inspection of the clustered studentized residuals. As a multi-variate model was used, this nested structure was maintained during diagnostic checks of models (Viechtbauer & Cheung, 2010). Studentized residuals greater than two were deemed outliers but, the totality of the residuals, more specifically the range of the residuals as a whole and in relation to one another, was also considered when assessing potential outliers. Outliers were also assessed in terms of their overlap with the aggregate effect's confidence interval. Influential effects were assessed via an examination of Cook's Distance. Both the raw and nested Cook's distances were assessed, with greater weight being given to influential points observed within the nested structure. Cook's distance's greater than 1 were inspected alongside values exceeding the sample specific threshold for highly influential cases[19] (Belsley et al., 1980). The totality of the relationships and positioning of values were also considered. Dfbeta's were also assessed and compared to a sample adjusted threshold[20] (Belsley et al., 1980). Additionally, a sensitivity analysis to assess changes in the aggregate effect was run in which suggested problematic cases were incrementally. Thus, all models have an original model and a reduced (redux) model which contains no outliers or influential cases.

---

[19] (4/n).
[20] $2/\sqrt{n}$

Publication Bias

Publication bias was inspected multiple ways. Egger's test is not suitable for complex or multi-variate meta-analytic models (Sterne & Egger, 2005; Van Aert et al., 2019). Therefore, standard error was included as a moderator to facilitate an equivalent regression test for funnel plot asymmetry (Sterne & Egger, 2005). Funnel plots were visually inspected. Trim and fill plots were not used as they are not compatible with multi-variate models. Significance values were plotted, and skewness was inspected (P curve analysis). Lastly, a limit or bias corrected effect size was calculated using PET-PEESE method(s) (Stanley, 2017).

All the above methods for detecting publication bias have their own inherent limitations, in other words, all might detect bias under a specific set of assumptions (Carter et al., 2019). Results were considered holistically. Both funnel plots and meta-regression or PET-PEESE method(s) are impacted by sample size. Both are not recommended for small samples of studies (Higgins et al., 2023). Funnel plots are not recommended for use when sample size of studies is fewer than 10, whilst PET-PEESE requires a sample of at least 20 studies. Smaller sample sizes are not uncommon in meta-analyses and compounding this issue is the consideration that PET-PEESE method(s) are not recommended when heterogeneity is high (Sladekov et al., 2023). This is another common issue for meta-analyses. The utility of funnel plots in regards to multi-variate models is debated. Interpretation is more challenging when effects could cluster together due to sampling dependency i.e. multiple samples within a single study which are accounted for within the meta-analytic model (Harrer et al., 2021). Funnel asymmetry also does not solely indicate the presence of publication bias as asymmetry could be the result of poor methodological quality (Sterne & Harbord, 2004). As per the RoB assessment however, the majority of studies sampled are methodologically sound. It is also worth noting that visual inspection of funnel plots is subjective and therefore inherently is subject to bias (Harrer et al., 2021).

CHAPTER 3: RESULTS

Validation

Validation of Merged Prejudice Data

Matched-to-sample outgroup implicit prejudice data was not significantly different from reported outgroup prejudice data and was therefore considered an acceptable proxy for missing outgroup implicit prejudice case (t (13.04) =-2.09, p = 0.056, see Appendix L&N). Matched-to-sample outgroup explicit prejudice data was, however, significantly different from reported outgroup prejudice data despite a high level of match-specificity[21] (t (34) =2.809, p=0.008, see Appendix L&M). While not fully validated as an acceptable proxy for reported data, all merged explicit outgroup values were analyzed with caution.

Overall Descriptive Statistics

Sample Descriptive statistics

The sampled participants encompassed a broad age range (10.56 – 72), with an average age of 23.97 (*SD* = 9.14). Participant age demographics were not explicitly reported and were thus inferred using proxy data in 18.67% of the sampled cases. Most of the sampled data tested identification memory on both male and female participants (93.33%). A sampling bias was evident, in which White participants were disproportionately sampled relative to other ethnic-nationalities (see Figure 6A). The three most sampled ethnic-nationality groups were White, Asian and Black respectively. The oversampling of White participants was mostly consistent across time (see Figure 6B). The most sampled country was the United States (44.49%), followed by Australia (17.57%) and Germany[22]. All of those are predominately White countries and align with the observation that the majority of the sample reside within the Global North (82.43%). With respect to sample positionality, the majority of participants were majority members who were tested on minority outgroup target faces (47.30%). A stimuli, or target, bias was evident within the sample. The most commonly used pairing of target faces were White-Black (35.16%), White-Asian (24.18%) and Same-Other[23] (10.99%).

[21] Only 33% of the sample was matched at the country level, and only four cases were matched outside of the five year interval for sample age. For those cases, data was matched within six years of publication.

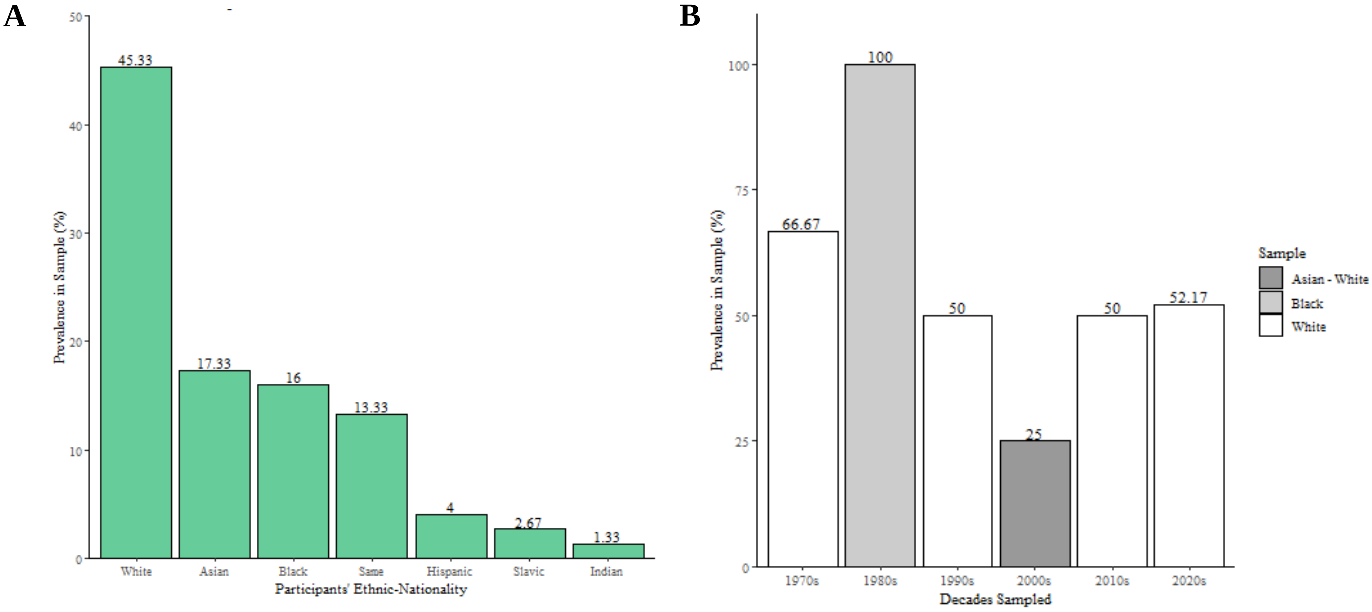
[22] See Appendix T

[23] Same-Other is used when data is collapsed across ethnicities sampled and/or multiple target faces were used (see Appendix U)

The most used identification task was an Old/New identification task (72%) whilst the least used identification tasks were delayed matching and lineup tasks (both 6.67%, see Appendix O).

Figure 6

Percentage (%) of Ethnic-Nationalities Sampled



Note. (6A) Prevalence of the sample bias in included articles. (6B) The most sampled Ethnic-Nationalities per decade are depicted. The 2000's were tied – with both White (25%) and Asian participants (25%) sampled most frequently.

Motivation was rarely manipulated within the sample via either task or instruction manipulations (74.67% of the sample had no such manipulation). Given the established contribution of socio-cognitive explanations of the OENE, this trend is unexpected and undercuts the importance of studies manipulating motivation. The most commonly used study, or encoding, instructions were the standard instructions informing participants of a pending identification test (84.75%, see Appendix P). Directed encoding, which refers to alerting the participants to pay attention to specific features or dimensions of the face during encoding resulting in a deeper level of face processing, was minimally represented within the sample (6.78%) The sample typically completed a fixed time identification task (90.63%, Self-paced 6.25% and Mixed 3.13%) with static, single-view target faces (96.88%). The average encoding time in seconds was short (*M* = 5.60s, *SD* =11.73s). The delay in minutes between encoding and test was short (*Mdn* = 2 mins, Mode = 0 mins). Delay was not explicitly reported and thus was inferred from procedural description in 60.66% of the sample. The average number of target faces present during study and average number of old or previously studied faces present during test were similar (*M* = 36, *M* = 35 respectively). The average number of new faces, or lures used during testing was 42, and the average total number of old and new faces during testing was 77. On average, the number of lures therefore exceeded the number of old target faces.[24]

Contact

Less than half of the sample measured and reported quality of outgroup contact alongside quantity of outgroup contact (40%). Given the theoretical importance of both quantity and quality of outgroup contact in effectively reducing outgroup prejudice and the OENE the observed trend to not measure both types of outgroup contact is concerning. Outgroup contact was mostly measured via self-report (85.33%) and the least used measure was an outgroup contact manipulation (4%, see Appendix Q). Manipulations of contact are often cost-prohibitive and time consuming when compared to self-report measures, so this trend aligns with expectations.

For both quantity and quality of outgroup contact, the most sampled period within the lifespan for level of outgroup contact was the entire life-span, followed by adulthood (see Table 3). The critical period of contact, which ends at 12, was underrepresented within the sample. Average reported quality of outgroup contact (*M* =.60, *SD* =.17) was higher than the average reported for quantity of outgroup contact (*M* =.51, *SD* =.16)[25]. The most sampled self-report measure of quantity of outgroup contact was that introduced by Hancock and Rhodes (2008), whilst the most sampled quality of outgroup contact self-report measure was Walker and Hewstone's *Individuating Experience Scale* (2006a; see Figure 7).

[24] Typically, the distribution of old and new target faces is equal i.e. 20 old, 20 new, 40 total faces that are tested during the identification task.

[25] Average values for outgroup contact variables completed via imputation followed the same trend (Quality, *M* =.51, *SD* =.21; Quantity *M* =.46, *SD* =.17)

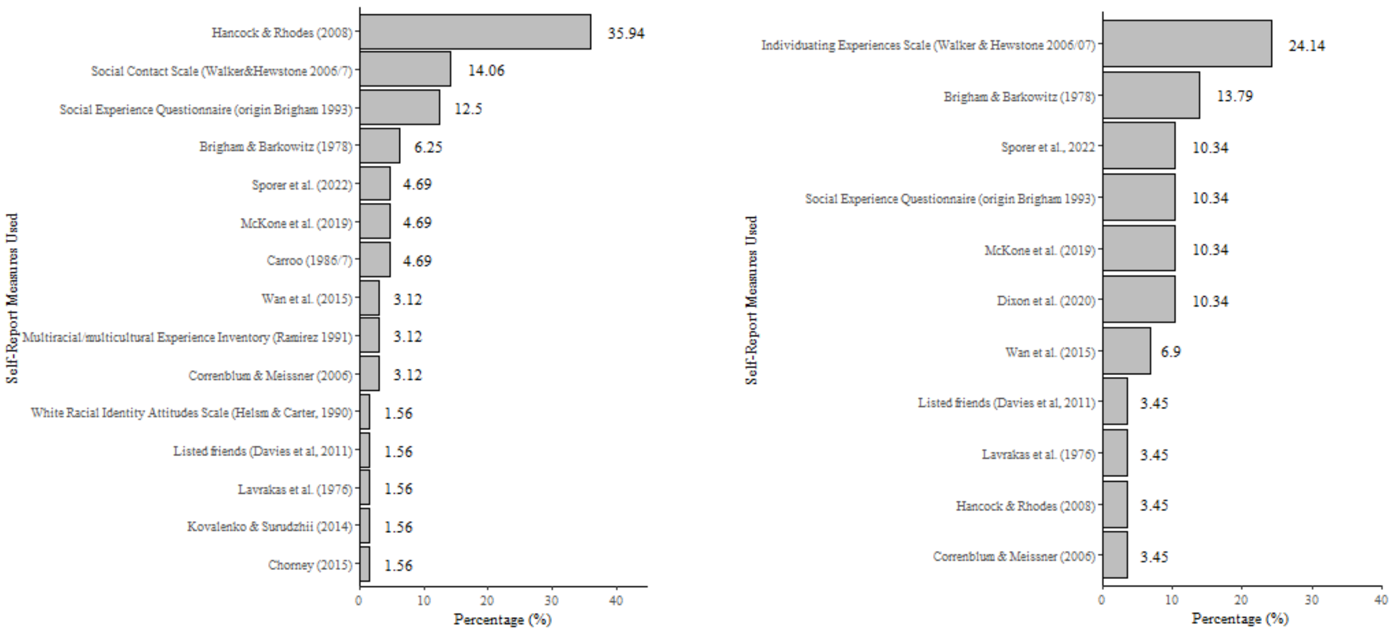
Table 3

Frequency (%) of Outgroup Contact per Time Band Sampled

Quantity of Outgroup Contact		Quality of Outgroup Contact	
Timespan	Percentage (%)	Timespan	Percentage (%)
Life span	66.86	Life span	94.23
Adult (18+)	11.63	Adult (18+)	1.92
Teenage (13-17)	8.14	Childhood (0-12)	1.92
Upper childhood to teen (10-17)	4.65	Upper childhood to teen (10-17)	1.92
Childhood (0-12)	4.07		
Middle childhood (6 -12)	4.07		
Present moment/manipulation	0.58		

Figure 7

Frequency (%) of Self-report Outgroup Contact Measures used



Note. (7A) Frequency of quantity of outgroup contact self-report measures in the sample. (7B) Frequency of quality of outgroup contact self-report measures in the sample.

Prejudice

Studies that reported an effect size for the outgroup contact-OENE relationship rarely measured and reported outgroup prejudice. Explicit outgroup prejudice was not measured in 72% of the sample, whilst implicit outgroup prejudice was not measured in 95.33% of the sample. On average, participants exhibited a low level of explicit outgroup prejudice and a neutral-slight average implicit outgroup prejudice (see Appendix R). The most sampled self-report measure for explicit outgroup prejudice was Brigham's (1993) *Attitude towards Blacks Scale* (Figure 8).

Figure 8

Frequency (%) of Self-report Explicit Outgroup Prejudice Measures

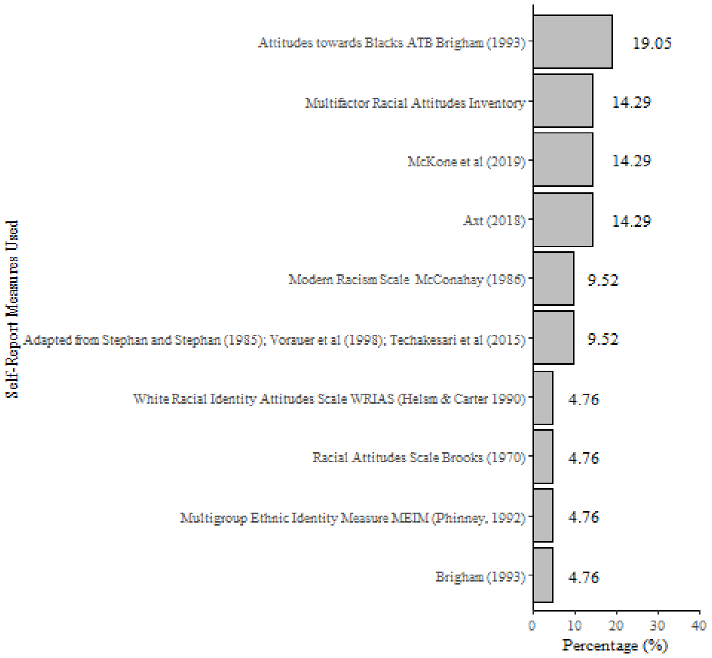
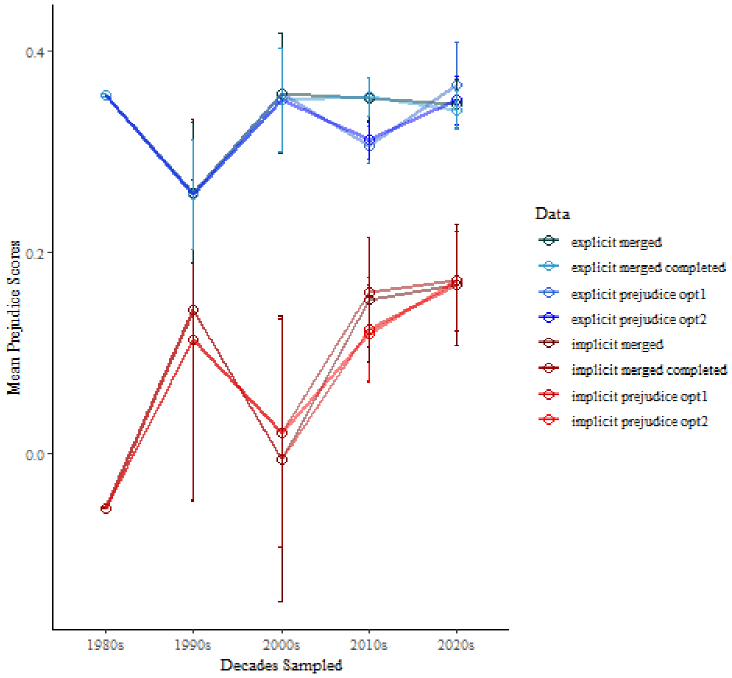


Figure 9

Line Graph Depicting Average Outgroup Prejudice Across Decades Sampled



Note. Explicit outgroup prejudice was reordered to facilitate an easier comparative analysis. For both types of outgroup prejudice, higher values indicate greater levels of outgroup prejudice. Graphed prejudice[26] values in the order they appear in the legend are as follows, (a) Harvard's Project Implicit Database values; (b) Harvard's Project Implicit Database values completed via imputation; (c) an original outgroup prejudice value when reported or a substitute Harvard Project Implicit Database value when missing; and (d) a combination prejudice score as before but, the substituted value i.e. Harvard's Project Implicit Database values were first completed via imputation before being used. 95% Confidence intervals are depicted

All average implicit prejudice values follow the same trends, affirming the suitability of data completed via multiple imputation (see Figure 9[27]). The largest mean difference across outgroup implicit prejudice indices, was noted in the 2010s. This mean change is relatively small (0.04). All sample implicit outgroup prejudice values fell within the neutral to slight range yet, there were descriptive changes in outgroup prejudice across time (see Footnote 26). Average implicit prejudice increased into the 1990s before decreasing in the 2000s. From the 2000s through to the 2020s, an upward trend in implicit prejudice was observed within the sample. The highest average implicit prejudice score, across all indices, was observed in the 2020s (M = .17 (slight), SD = .17)

All explicit prejudice values follow the same trends, which (a) affirm the validity of the data completed via multiple imputation (see Figure 9); and (b) which demonstrate the sample tended towards a low degree of explicit outgroup prejudice (all values < .40, maximum scale value is 1). The largest mean difference across explicit prejudice indices was noted in the 2010s and again is relatively small (0.05). Average explicit outgroup prejudice decreased from the 1980s to the 1990s before increasing in the 2000s. In the 2010s, explicit prejudice decreased slightly on average and remained relatively consistent into the 2020s.

---

[26] Graphed values are included to describe the sample analyzed. As depicted via the 95% CIs, mean outgroup prejudice changes are likely the result of random fluctuations. This is not intended for extrapolation beyond the sample. While the sample included studies from the 1970s, this output was not graphed as (a) no original outgroup prejudice data was reported, (b) a substituted value from the Harvard Project Implicit Database was used and as the database records began in 2002, the values used therefore had a sizable year match difference and thus may not be reflective of the true outgroup prejudice for this timespan within the sample (n=3). Additionally, see Appendix G for more detail on implicit outgroup prejudice interpretation i.e. the convention used.

[27] For descriptive statistics of outgroup prejudice across time, including mean change, see Appendix S

The highest average explicit prejudice score, across all indices, was observed in the 1980s (*M* = .36, *SD* =.00).

Sample average implicit and explicit outgroup prejudice values follow divergent trends between the 1980s and the 2000s. It is speculated that the divergence observed in the sample within the 1990s is likely the result of historic world events such as the end of Apartheid, the end of the 'Cold War' i.e. the collapse of the Soviet Union and the increase in telecommunications i.e. the release of internet connectivity to the broader public. Such contextual events could account for the trending decrease in explicit prejudice i.e. a hesitancy in forthrightness during this period whilst average implicit prejudice increased. Similarly, the divergent trends noted in the 2000s could be speculated to be the result of contextual world events such as the September 11th collapse of the World Trade Centers (Cashin, 2010). The pattern of 'othering' in the wake of this act of terrorism could account for the trending increase in explicit outgroup prejudice observed within this time period.

OENE

As a validation check, the aggregate strength of the OENE within the sample was tested (see Appendix V-AK)[28]. The sample data exhibited classic patterns of an OENE for hits, false-alarms, response bias and discrimination (see Table 4). This is supporting evidence in favour of hypotheses 1a-d, and the validity of findings from the meta-analytic models exploring the OENE-outgroup contact and OENE-outgroup prejudice relationships.

All models had large and statistically significant effect size heterogeneity which supported the need for the accompanying moderator analyses. Across the examined identification indices[29], outgroup contact and outgroup prejudice significantly moderated the strength of the observed OENE. This lends support to the interconnected nature of these relationships.

[28] Full model and moderator analysis is supplied within the appendix
[29] Identification indices include hits, false alarms, response bias and discriminability/discrimination.

Table 4

Summary of OENE Aggregate Effects

												95% Cis		Heterogeneity			
Outcome	Hypothesis	Data	Number of Excluded Effects Post Diagnostics	Number of Excluded Articles Post Diagnostics	Number of effects	Number of studies (k)	Aggregate Effect - Hedge's g	Expected Direction	Interpretation	p	Sig	LB	UB	I^2	Significant Cochrane's Q	Between Study Heterogeneity (τ^2)	Within Study Heterogeneity
Hits	Hits will be higher for in-group relative to out-group members (positive difference score) indicating the presence of the OENE	Original Redux	3	2	29	11	0.13	Yes	Negligible	0.059	-	-0.01	0.27	93.02%	Yes	0.00	0.13
		Imputed Redux	3	2	32	13	0.11	Yes	Negligible	0.099	-	-0.02	0.25	92.77%	Yes	0.00	0.14
		Original All	-	-	32	13	0.16	Yes	Negligible	0.343	-	-0.17	0.50	94.81%	Yes	0.28	0.16
		Imputed All	-	-	35	15	0.14	Yes	Negligible	0.353	-	-0.15	0.42	94.55%	Yes	0.21	0.17
False-Alarms	False-Alarms will be higher for out-group relative to in-group members (negative difference score) indicating the presence of the OENE	Original Redux	8	2	19	9	-0.57	Yes	Medium to Large	<.0001	***	-0.79	-0.35	86.61%	Yes	0.08	0.04
		Imputed Redux	8	2	22	11	-0.50	Yes	Medium	<.0001	***	-0.70	-0.30	87.20%	Yes	0.08	0.05
		Original All	-	-	27	11	-0.62	Yes	Medium to Large	0.002	**	-1.01	-0.24	95.98%	Yes	0.28	0.24
		Imputed All	-	-	30	13	-0.55	Yes	Medium to Large	0.001	**	-0.88	-0.22	95.67%	Yes	0.23	0.23
Discrimination	Discrimination will be higher for in-group relative to out-group members (positive difference score) indicating the presence of the OENE	Original Redux	26	6	52	19	0.33	Yes	Small to Medium	<.0001	***	0.18	0.47	86.06%	Yes	0.05	0.07
		Imputed Redux	20	9	68	22	0.40	Yes	Small to Medium	<.0001	***	0.28	0.53	86.65%	Yes	0.05	0.06
		Original All	-	-	78	25	0.52	Yes	Medium to Large	<.0001	***	0.28	0.76	92.80%	Yes	0.20	0.29
		Imputed All	-	-	88	31	0.69	Yes	Medium to Large	<.0001	***	0.39	0.99	93.95%	Yes	0.36	0.63
Response Bias	Response Bias will be higher for in-group relative to out-group members (positive difference score) indicating the presence of the OENE	Original Redux	8	2	7	6	0.44	Yes	Small to Medium	<.0001	***	0.32	0.55	60.40%	Yes	0.01	0.00
		Imputed Redux	13	3	17	10	0.46	Yes	Small to Medium	<.0001	***	0.25	0.66	89.05%	Yes	0.00	0.17
		Original All	-	-	15	8	0.63	Yes	Medium to Large	0.027	*	0.07	1.19	97.70%	Yes	0.55	0.10
		Imputed All	-	-	30	13	0.47	Yes	Small to Medium	0.004	**	0.15	0.79	96.30%	Yes	0.22	0.20

Note. Original data indicates data that had complete SDs. Original-redux indicates a subsample of values after the removal of outliers/influential points. Imputed indicates data which contains SDs estimated via imputation. Level of significance is indicated via *<.05, **<.01, ***<.001.

Primary Meta-Analyses

For meta-analytic models containing influential cases and/or outliers, the aggregate effect for the model without such cases was reported i.e. the '-redux' model. For models without such cases, the aggregate effect for the full sample i.e. the '-all' model was narratively reported. Aggregate effects for all meta-analytic models can be found in a summary table at the end of each analysis i.e. at the end of each relationship of interest. Only significant moderators were reported[30]. Significant moderators across all models can be found within the Appendices. Moderators that are significant across all models carry greater evidentiary weight.

OENE-Contact

Quantity of Outgroup Contact – Identification Difference Scores

Identification difference scores were derived from discrimination (68.09%) and Cambridge Face Matching Test (CFMT[31], 31.91%) values. The aggregate effect for the relationship between outgroup contact and identification difference scores was statistically significant and slight/small in strength (r = -.11, p<.001). Heterogeneity was moderate and significant within the sample supporting the need for a moderator analysis (I^2 = 38.79%). Greater levels of outgroup contact are associated with a significant reduction in the size of the OENE, or size of the identification difference scores however, heterogeneity indicates the observed effect size varies substantially across populations. Therefore, there is only tentative support for Hypothesis 2a. Publication bias was present within the sample[32].

Moderators

Explicit Prejudice (merged completed)[33]. When favourable explicit outgroup attitudes increase, or conversely when explicit outgroup prejudice decreases, the relationship between outgroup contact and identification difference scores is strengthened. In this way, when there is lower explicit outgroup prejudice, greater outgroup contact is associated with a

---

[30] All levels of a categorical moderator may not be significant. In such instances, only significant levels of categorical moderators were reported narratively.

[31] The CFMT task differs from traditional identification tests in which the participant views a previously seen or 'old' face once during the study i.e. encoding and once during testing. In this task participants are 'trained' on three target faces i.e. multiple iterations of study and test are used, with varying facial viewing angles – which in theory should provide richer information and therefore a deeper encoding of the target faces. As the test progresses, identifications become incrementally more difficult as targets are displayed with greater visual 'noise' or distortions.

[32] See Appendix AL for additional publication bias plots and/or checks. See Appendix AM for diagnostics.

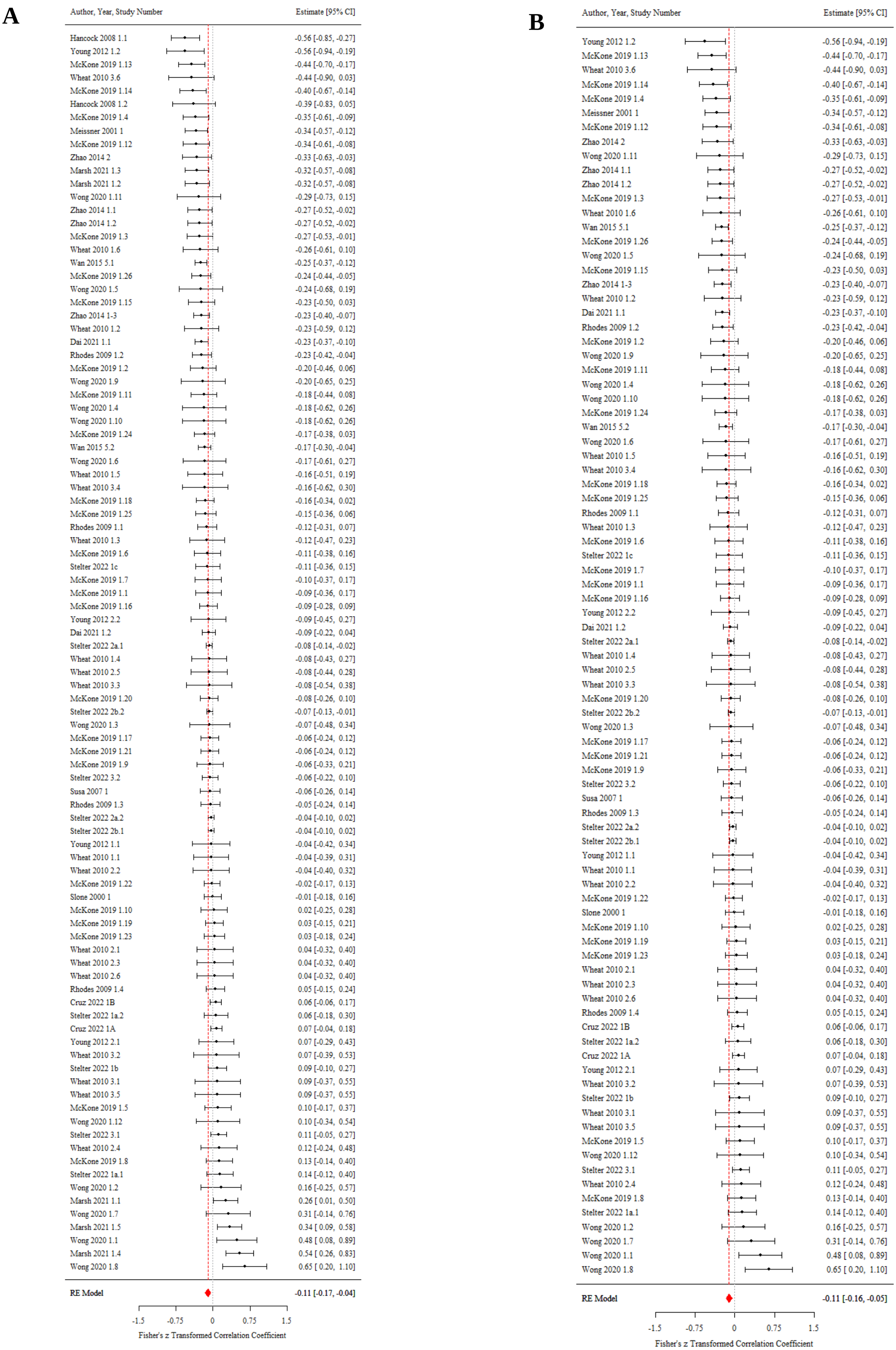
[33] Harvard Project Implicit Database values wherein missing values were estimated via imputation.

Figure 10

Forest Plots for Quantity of Outgroup Contact – Outgroup Identification Difference Scores

A

B

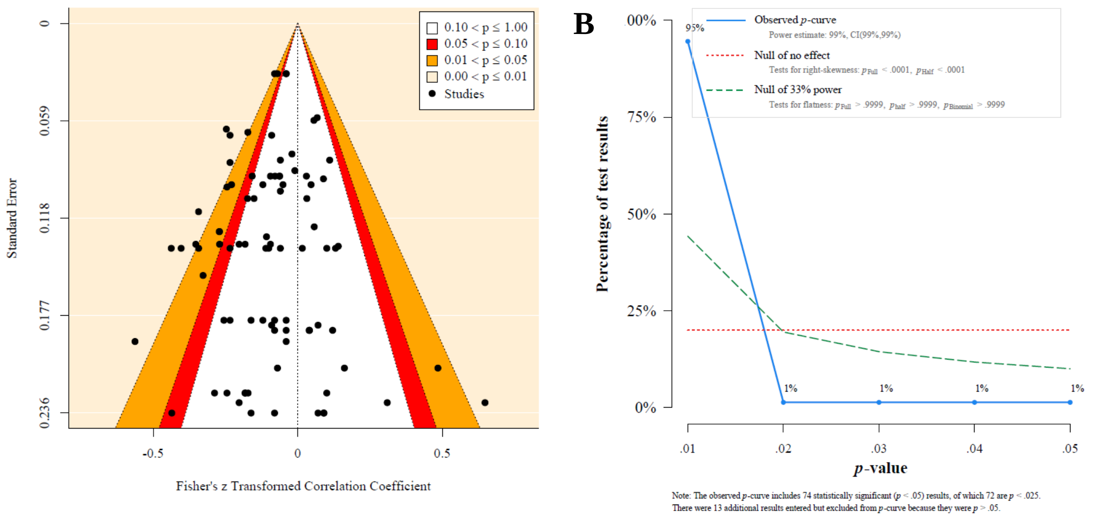


Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (10A) Forest plot for all suitable effects. (10B) Forest plot after influential cases and/or outliers have been removed i.e. '-redux' model.

Figure 11

Diagnostic Plots for Quantity of Outgroup Contact – Outgroup Identification Difference Scores after Outliers and Influential Cases were Removed

A



Note. (11A) Funnel Plot testing for publication bias. (11B) P-curve Analysis testing for publication bias.

- greater reduction in the OENE (Original-all: -.59, z=-2.50, p<.05). The direction of the estimate is therefore consistent with expectations.

Quantity of outgroup contact in the critical period. When outgroup contact occurs within the critical period, up to 12 years of age, outgroup contact is associated with a greater reduction in the size of the OENE (for both models, Original-all & Original-redux: -.19, z=-3.25, p<.01). When outgroup contact occurs outside of the critical period, i.e. later life, the relationship between outgroup contact and identification difference scores is weakened. Outgroup contact within this period is associated with lower gains in the reduction of the OENE (both models: .06, z=1.97, p<05). Results are in the expected direction.

Quantity of outgroup contact (level) in or out of the critical period. Consistent with expectations, when quantity of outgroup contact is high and this contact occurs within the critical period, outgroup contact is associated with a greater reduction in the OENE (both models: -.21, z=-2.52, $p<.05$).

Quantity of outgroup contact time bands. When outgroup contact occurs during adulthood, the inverse relationship of outgroup contact-OENE is strengthened. Contact during adulthood is therefore associated with a reduction in the OENE or difference scores[34] (Original-all: -.15, z=-2.81, $p<.01$; Original-redux: -.16, z=-3.75, $p<.001$). When outgroup contact occurs during middle childhood i.e. between ages 6-12, as opposed to occurring during adulthood, outgroup contact is associated with an even greater reduction in the observed OENE (Original-all: -.21, z=-2.94, $p<.01$; Original-redux: - .20, z=-3.23, $p<.01$). When outgroup contact occurs across the life span as opposed to occurring during adulthood, the association between outgroup contact and an effective reduction in the size of the OENE is weakened (Original-redux: .09, z= 2.20**,** $p<.05$). This is counter to the expected pattern of results, wherein contact across the life span would be expected to be more beneficial in OENE reduction as life span encompasses contact within the critical period (see footnote 34).

Encoding Instructions. When encoding instructions are directed, a deeper level of face-processing ensues. In such cases, the association between outgroup contact and a reduction in the size of the OENE is strengthened (Original-all: -.56, z=-2.36, $p<.05$; Original-redux: -.59, z=-2.88, $p<.01$). This is consistent with the expected pattern of results. By comparison when the encoding instructions are standard, i.e. the boiler plate warning of a later memory test, the association between outgroup contact and a reduction in the OENE is weakened (Original-all: .53, z=2.22, $p<.05$; Original-redux: .51, z=2.46, $p<.05$). This finding is counter to expectations of a weaker negative relationship, relative to directed encoding, in which outgroup contact would still be associated with a reduction in the OENE.

Type of encoding. Similar to the above-mentioned trends, this moderator concerns the level of face processing that occurs during the study or encoding phase. As such, when the level of processing is basic, outgroup contact is associated with a reduction in the OENE (Original-redux: -.06, z=-2.03, $p<.05$) however, the association between outgroup and a

[34] The meta-analysis found that outgroup contact was associated with a smaller OENE or fewer outgroup identification errors. The moderator analysis confirms such findings with all specific time bands* reducing outgroup identification errors. The greatest error reductions are associated with outgroup contact occurring earlier in life – the earliest time band in this sample was 6-12 and this had the greatest reduction in outgroup errors.

*Life span contact is relatively difficult to self-report and objectively assess. This finding is counterintuitive to what would be expected however, this difficulty of assessment should be considered when more focused time bands all follow the same pattern of any outgroup contact being beneficial.

reduction in OENE is even greater when a deep level of face-processing occurs (Original-all: -.42, z=-3.92, p<.001, Original-redux: -.35, z=-3.03, p<.01)

Task/Cognitive demands. When task and/or cognitive demands are high, the inverse relationship between outgroup contact and identification difference scores are strengthened. Namely, outgroup contact is associated with a greater reduction in the observed OENE when cognitive or task difficulty increases (Original-all: -.09, z=-2.33, p<.05; Original-redux: -.09, z=-2.65, p<.01).

Sample country's positionality. Countries in the 'Global North' tend to be less integrated and or less diverse which therefore presumably has a knock-on effect re opportunity for outgroup contact or exposure. When a country belongs to the 'Global North', outgroup contact is associated with an even greater reduction in the observed OENE or difference scores (Original-all: -.10, z=-3.08, p<.01; Original-redux: -.10, z=-3.64, p<.001).

This follows expectations as contact would be more beneficial in cases where a prior history thereof may be limited.

Motivation. In the absence of any motivation manipulations or instructions, outgroup contact is associated with a significant reduction in the size of the OENE i.e. a lower number of outgroup identification errors[35] (Original-all: -.08, z=-2.10, p<.05: Original-redux: -.09, z=-2.99, p<.01). This pattern is consistent with expectations.

Task. When the identification task is an Alternative Forced Choice (AFC[36]) task, outgroup contact is associated with the greatest reduction in the observed OENE across all task types (Original-all: -.50, z=-3.21, p<.01). When the identification task used is either a CFMT, delayed matching or an old-new task as opposed to an AFC task, outgroup contact is associated with a smaller reduction in OENE.

Face format at encoding. When target faces are viewed from multiple viewpoints, outgroup contact is associated with a greater reduction in the observed OENE (Original-redux: -.23, z=-2.09, p<.05). This is consistent with expectations as greater viewing angles of

[35] None of the tasks in the sample included a motivation to individuate instruction or made use of a task that encouraged motivation. No motivation was therefore the reference category.

[36] AFC tasks are a specific type of Old-New task. As with Old-New tasks, participants completing an AFC task are shown (i) previously viewed or 'old' targets and (ii) novel or 'new' targets. Participants are asked to identify previously seen targets i.e. to make an identification. Unlike traditional Old-New tasks which administer target faces sequentially and do not constrain the participants identification decisions, AFC tasks impose identification constraints and present target faces side by side or simultaneously, enabling a comparison of target faces. There are numerous configurations such as 2AFC or 3AFC for example. In a 2AFC task, participants are shown two faces and they must identify which of the displayed faces was previously viewed.

a face facilitate better memory recall and by extension outgroup identification, with the caveat that participants are motivated to use such information.

Positionality of participants relative to outgroup members. When participants are majority members who are being tested on minority outgroup members, the opportunity in daily life for outgroup exposure is lower, and thus memory and identification performance for outgroup targets is poor. Thus, consistent with expectations, in such cases i.e. majority members tested on minority outgroup members, greater outgroup contact is associated with a greater reduction in the OENE (Original-redux: -.07, z=-2.44, p<.05).

Publication. Counter to expectations, non-published works have a stronger association between outgroup contact and a reduction in OENE (Original-redux: -.14, z=-2.41, p<.05). This is counter to expectations, as the expectation would be that published works find a stronger, often significant, relationship between contact and the OENE.

Moderator Summary

All significant moderators, across models, are reported in Appendix AAF.

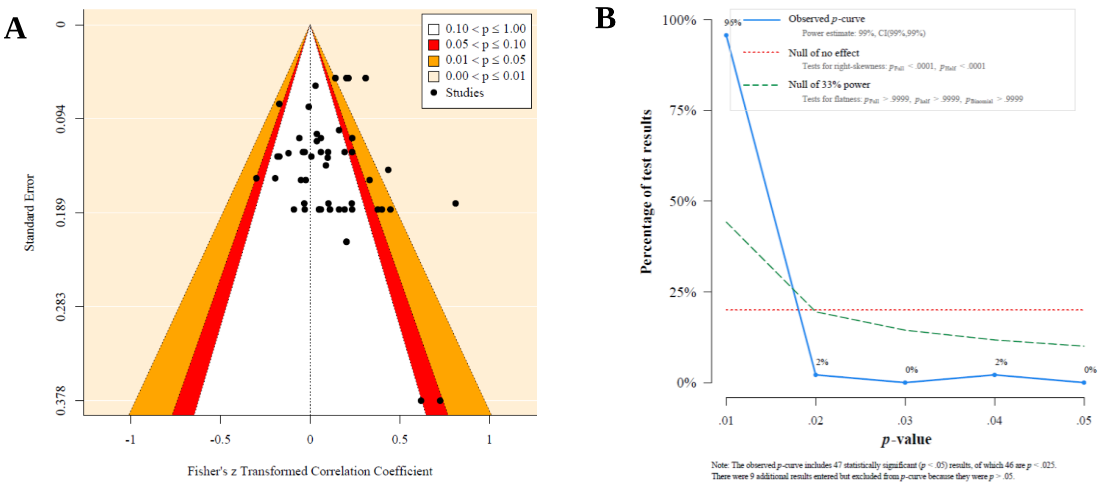
Quantity of Outgroup Contact – Outgroup Discrimination

The aggregate effect for the relationship between quantity of outgroup contact and outgroup discrimination was small/slight and statistically significant (r = .10, p<.01). Heterogeneity was moderate and significant supporting the need for a moderator analysis (I^2 = 56.3%). Higher outgroup contact is associated with higher outgroup discrimination or overall identification accuracy however, owing to the significant heterogeneity there is only tentative support for Hypothesis 2b. Publication bias was not present within the sample[37].

[37] See Appendix AN for additional publication bias plots and/or checks. See Appendix AO for diagnostics.

Figure 12

Diagnostic Plots for Quantity of Outgroup Contact – Outgroup Discrimination (d-prime) after Outliers and Influential Cases were Removed



Note. (12A) Funnel Plot testing for publication bias. (12B) P-curve Analysis testing for publication bias.

Moderators

Type of encoding. When a basic-level of face processing is used during the study phase of an identification task, outgroup contact is associated with greater gains in outgroup discrimination (Original-all: .13, z=3.23, p<.001; Original-redux: .13, z=3.16, p<.01).

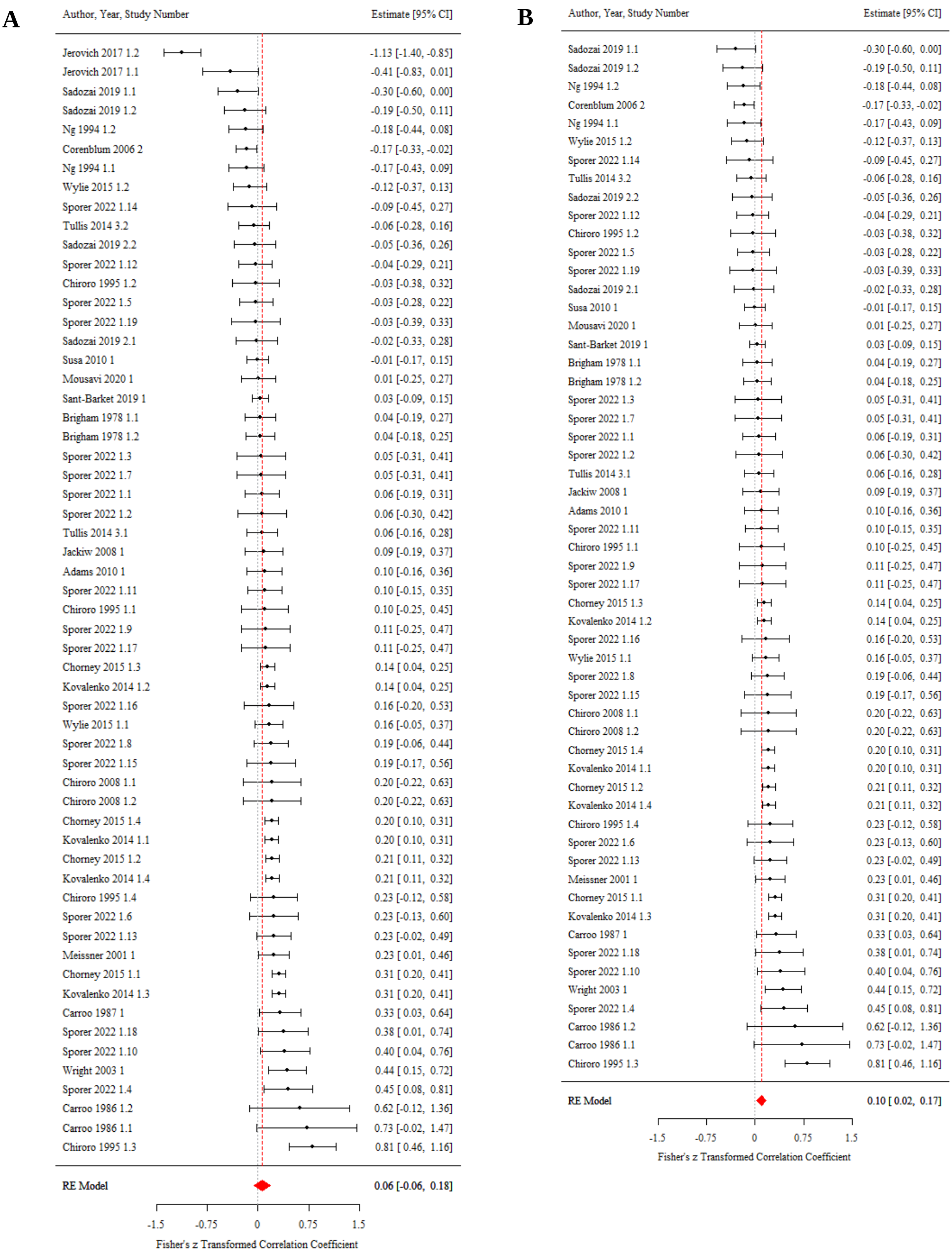
By comparison, the expected pattern of results is not observed for a combination of basic and deep face processing. A combination variable including deep processing should follow the above trend of strengthening the relationship between outgroup contact and outgroup discrimination. Instead, when either basic or deep face processing was used, the association between outgroup contact and outgroup discrimination was weakened (Original-all: -1.02, z=-5.85, p<.001)

Sample country's positionality. When sample countries are located within the 'Global South', the association between outgroup contact and outgroup discrimination is strengthened. Outgroup contact is therefore associated with greater outgroup discrimination (Original-all: .25, z=2.14, p<.05; Original-redux: .23, z=2.40, p<.05).

Motivation. When an identification task does not include any motivation manipulations or instructions, outgroup contact is associated with greater gains in outgroup discrimination (Original-all: .11, z=2.96, p<.01; Original-redux: .11, z=2.91, p<.01). Counter to expectations, a combination of motivation and non-motivation conditions weakens the relationship between outgroup contact and outgroup discrimination (Original-all: -1.00, z=-5.75, p<.001).

Figure 13

Forest Plots for Quantity of Outgroup Contact – Outgroup Discrimination (d-prime)



Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (13A) Forest plot for all suitable effects. (13B) Forest plot after influential cases and/or outliers have been removed i.e. '-redux' model.

Length of delay (categorical). When a brief, 2–15-minute delay is used between encoding and testing, the relationship between outgroup contact and outgroup discrimination is strengthened (Original-redux: .19, z=2.46, $p<.05$). Thus, even with a longer delay outgroup contact is associated with greater outgroup discrimination.

Positionality of participants relative to outgroup members. When participants are majority members who are tested on minority outgroup members, outgroup contact is associated with greater outgroup discrimination (Original-redux: .13, z=2.15, $p<.05$). Such participants display a greater OENE and despite this, the relationship between outgroup contact and outgroup discrimination is strengthened (Chien et al., 2018; Wong et al., 2020).

Task/Cognitive demands. When an identification task has a high task and or cognitive demand, outgroup contact is associated with greater gains in outgroup discrimination (Original-redux: .11, z=2.83, $p<.01$)

Moderator Summary

All significant moderators, across models, are reported in Appendix AAG.

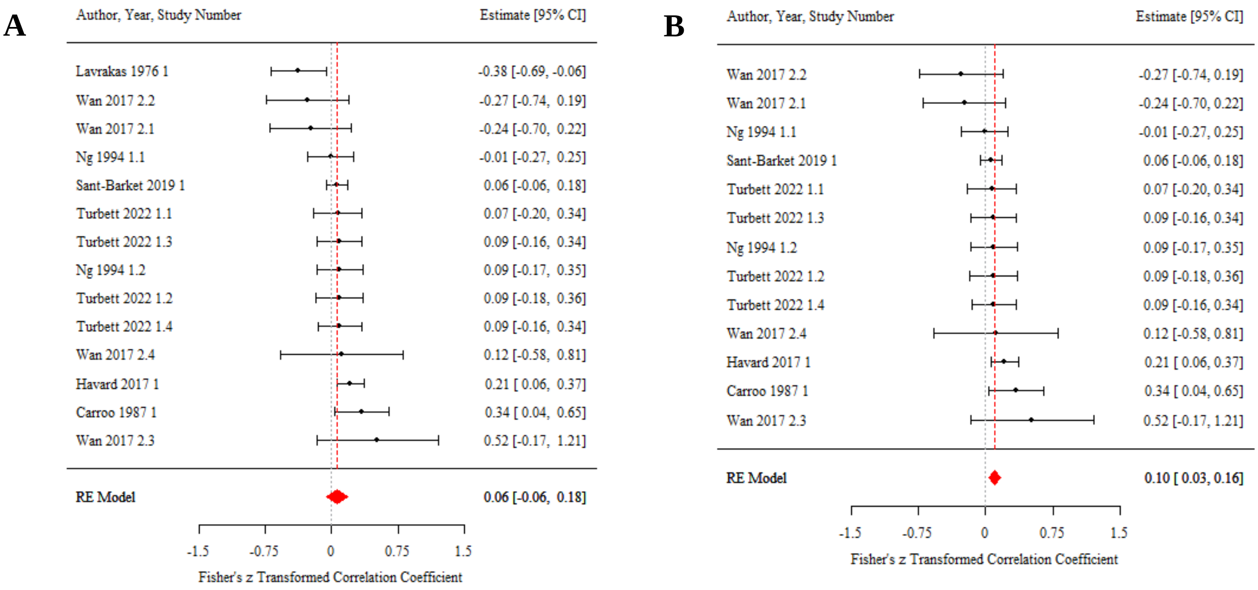
Quantity of Outgroup Contact – Outgroup Hits

The aggregate effect for the relationship between outgroup contact and outgroup hits was slight/small and statistically significant (r=.10, $p<.01$). Heterogeneity was not important and not statistically significant. Higher levels of outgroup contact are associated with higher outgroup hits, or correct identifications. As heterogeneity suggests effects are consistent across populations, hypothesis 2c was therefore supported. Publication bias was not present within the sample[38].

[38] See Appendix AP for additional publication bias plots and/or checks. See Appendix AQ for diagnostics.

Figure 14

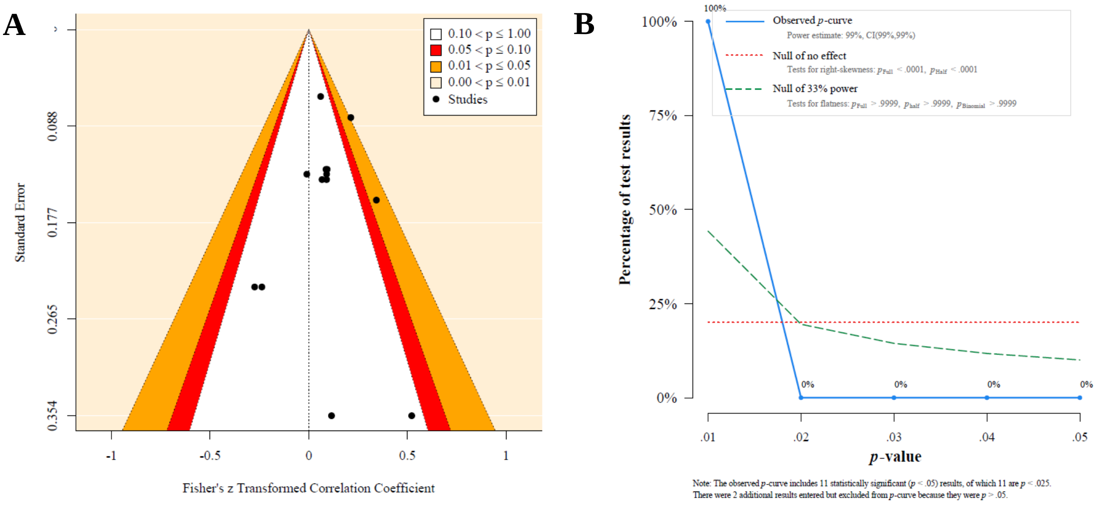
Forest Plots for Quantity of Outgroup Contact- Outgroup Correct Identifications (Hits)



Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (14A) Forest plot for all suitable effects. (14B) Forest plot after influential cases and/or outliers have been removed i.e. '-redux' model.

Figure 15

Diagnostic Plots for Quantity of Outgroup Contact – Outgroup Correct Identifications (Hits) after Outliers and Influential Cases were Removed



Note. (15A) Funnel Plot testing for publication bias. (15B) P-curve Analysis testing for publication bias.

Moderators

Implicit Prejudice (combined option 1)[39]. When participants have higher implicit outgroup prejudice, the relationship between outgroup contact and outgroup hits is weakened (Original-all: -.42, z=-2.19, p<.05). The presence of implicit prejudice therefore limits the associated gains of outgroup contact with outgroup hits. A large proportion of total variance was explained via the inclusion of implicit prejudice, therefore underscoring the importance of including prejudice in the outgroup contact-outgroup identification relationships (80.80%).

Implicit Prejudice (merged)40. The same pattern of results was observed (Original-all: -.42, z=-2.19, p<.05).

Length of encoding (categorical). When encoding time is brief, 0-2 seconds, outgroup contact is no longer associated with trending gains in outgroup hits (Original-all: -.38, z=-2.35, p<.05). A longer encoding time by comparison, strengthens the relationship between outgroup contact and outgroup hits thereby associating outgroup contact with greater outgroup hits. With the longest encoding of greater than 9 seconds per face, outgroup contact is associated with the largest gains in outgroup hits (Original-all: .59, z=3.29, p<.01).

Lures. A higher number of lures, or new faces, used during testing weakens the relationship between outgroup contact and outgroup hits (-.01; z=-2.48, p<.01). Therefore, when more new faces are used at test, the associated gains from outgroup contact with outgroup hits are reduced. This is an important moderator accounting for 48.53% of the total variance

Positionality of participants relative to outgroup members. When minority members are tested on majority outgroup members, outgroup contact is associated with greater outgroup hits (Original-all: .25, z=2.06, p<.05, 20.87%).

Face format at encoding. When study faces are viewed and therefore studies from multiple viewpoints, outgroup contact is associated with greater gains in outgroup hits (Original-redux: .21, z=2.64, p<.01).

---

[39] Original reported values were used when available, when no such data was available a substitute value from Harvard's Project Implicit Database was used.

[40] Values from Harvard's Project Implicit Database.

Sample country's positionality. When a sample country is classified as belonging to the 'Global North', outgroup contact is associated with greater outgroup hits (Original-redux: .10, z=2.44, $p<.05$)

Moderator Summary

All significant moderators, across models, are reported in Appendix AAH.

Quantity of Outgroup Contact – Outgroup False-alarms

The aggregate effect for the relationship between outgroup contact and false alarm rates was slight/small, non-significant and counter to expectations yet this may be the result of a small sample (r=.06, $p>.05$, n = 4 effects). Therefore, hypothesis 2e was not supported (see Appendix AT). Effect size heterogeneity was non-significant, and the model yielded no significant moderators.

Quantity of Outgroup Contact – Outgroup Response Bias

The aggregate effect for the relationship between outgroup contact and outgroup response bias was slight/small, non-significant and counter to expectations (r=-.19, $p>.05$). This sample was also small (n = 3 effects). Hypothesis 2d was therefore not supported. Effect size heterogeneity was considerable and statistically significant (I^2 = 78.39%, see Appendix AU). Implicit outgroup prejudice and sample age were significant moderators (Appendix AU).

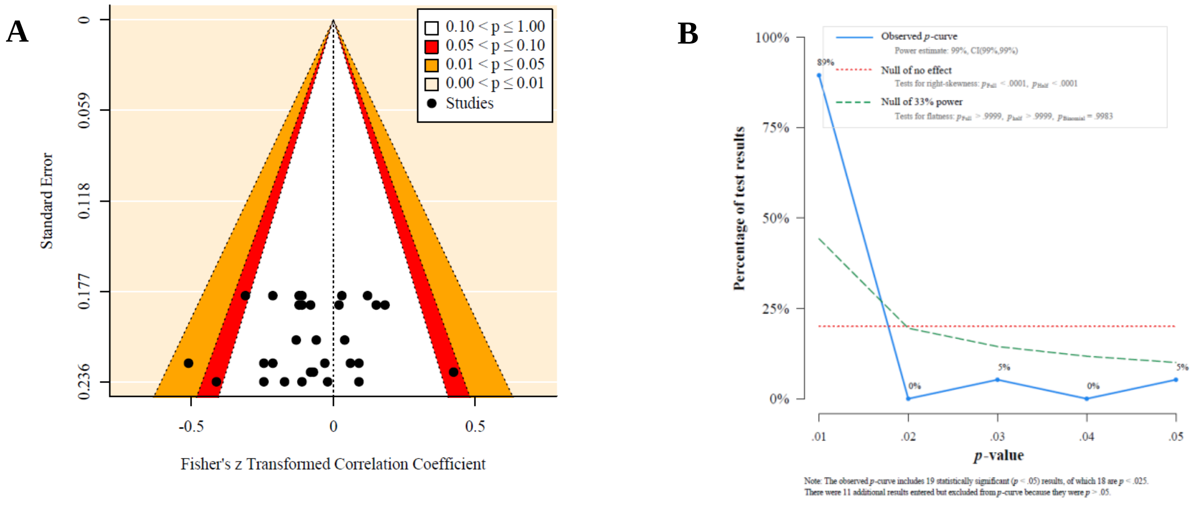
Quality of Outgroup Contact – Identification Difference Scores

The aggregate effect for the relationship between quality of outgroup contact and identification difference scores was slight/small and non-significant (r = -.07, $p>.05$). Heterogeneity was not a problem and was non-significant ($p>.05$). As quality of outgroup contact increases, it is associated with a reduction in the OENE or identification difference scores. Owing to non-significant heterogeneity, Hypothesis 2a was therefore supported for quality of outgroup contact. Publication bias was not present in the sample[41]

[41] See Appendix AR for additional publication bias plots and/or checks. See Appendix AS for diagnostics.

Figure 16

Diagnostic Plots for Quality of Outgroup Contact – Identification Difference Scores after Outliers and Influential Cases were Removed



Note. (16A) Funnel Plot testing for publication bias. (16B) P-curve Analysis testing for publication bias.

Moderators

Task/Cognitive demands. When cognitive and/or task demands are high within the context of the identification test, greater outgroup quality of contact is associated with a greater reduction in the OENE or size of the difference scores (Original-all: -.04, z=-2.59, p<.01).

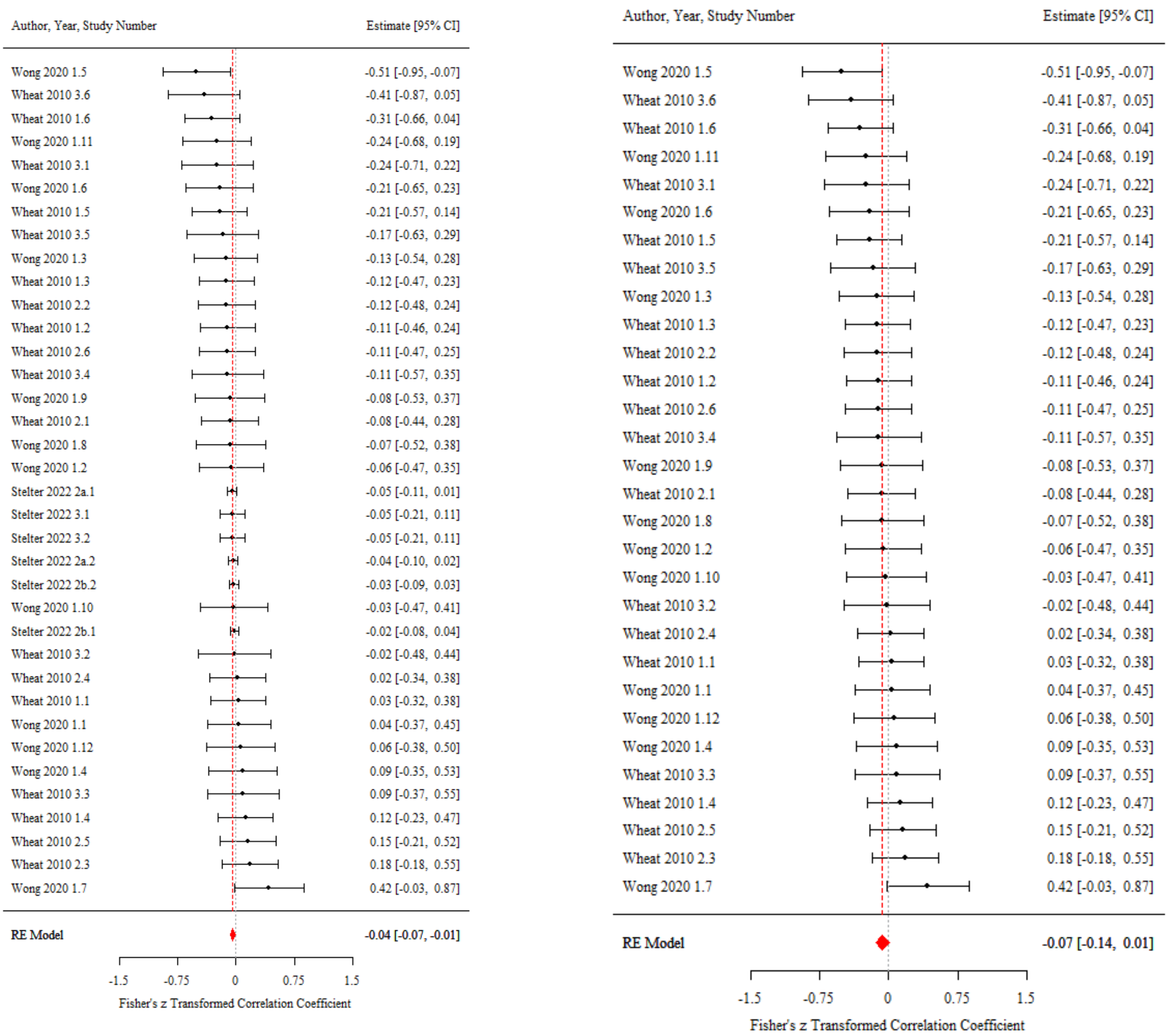
Sample country's positionality. When a sample country belongs to the 'Global North', higher outgroup quality of contact is associated with a greater reduction in the observed OENE (Original-all: -.04, z=-2.79, p<.01).

Motivation. Even when the identification task does not contain any motivation manipulations or instructions, greater outgroup quality of contact is associated a greater reduction in the OENE (Original-all: -.04, z=-2.60, p<.01).

Positionality of participants relative to outgroup members. When participants are majority members who are tested on minority outgroup members, greater outgroup quality of contact is associated with a greater reduction in the OENE (Original-redux: -.14, z=-2.28, p<.05)

Figure 17

Forest Plots for Quality of Outgroup Contact – Outgroup Identification Difference Scores



Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (17A) Forest plot for all suitable effects. (17B) Forest plot after influential cases and/or outliers have been removed i.e. '-redux' model.

Moderator Summary

All significant moderators, across models, are reported in Appendix AAI.

Quality of Outgroup Contact – Outgroup Discrimination

The aggregate effect for the relationship between outgroup quality of contact and outgroup discrimination was slight/small, non-significant and counter to expectations as higher quality of contact should be associated with an increase in overall outgroup identification accuracy (*r*=-.12, p>.05, see Appendix AV). The results may be the result of a very small sample (n = 2 effects). Hypothesis 2b was therefore not supported for quality of outgroup contact. Effect size heterogeneity was non-significant.

OENE – Outgroup Contact Summary

The results of all meta-analytic models on the OENE-outgroup contact relationship are summarized in Table 5.

Table 5

Summary of Aggregate Effects for OENE – Outgroup Contact

													95% Cis		Heterogeneity			
Outcome	Hypothesis	Data	Number of Excluded Effects Post Diagnostics	Number of Excluded Articles Post Diagnostics	Number of effects	Number of studies (k)	Aggregate Effect - **Fishers Z**	Aggregate Effect - Pearson's r	Expected Direction	Strength Interpretation	p	Significance Level	LB	UB	I²	Significant Cochrane's Q	Between Study Heterogeneity (τ²)	Within Study Heterogeneity
Quantity of Outgroup contact - Difference Scores	Higher outgroup contact decreases the observed OENE	Original Redux	7	2	87	14	-0.11	-0.11	Yes	Slight	0.000	***	-0.16	-0.05	38.79%	Yes	0.01	0.00
		Original All	-	-	94	16	-0.11	-0.11	Yes	Slight	0.001	**	-0.17	-0.04	52.91%	Yes	0.01	0.01
Quantity of Outgroup contact - Outgroup Discriminability	Higher outgroup contact increases outgroup discrimination	Original Redux	2	1	56	20	0.10	0.10	Yes	Slight	0.010	**	0.02	0.17	56.35%	Yes	0.02	0.00
		Original All	-	-	58	21	0.06	0.06	Yes	Slight	0.305	-	-0.06	0.18	73.14%	Yes	0.06	0.00
Quantity of Outgroup contact - Outgroup Hits	Higher outgroup contact increases outgroup hits	Original Redux	1	1	13	6	0.10	0.10	Yes	Slight	0.005	**	0.03	0.16	0%	No	0.00	0.00
		Original All	-	-	14	7	0.06	0.06	Yes	Slight	0.343	-	-0.06	0.18	34.68%	No	0.02	0.00
Quantity of Outgroup contact - Outgroup False Alarms	Higher outgroup contact decreases outgroup false alarms	Original Redux	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Original All	-	-	4	3	0.06	0.06	No	Slight	0.195	-	-0.03	0.15	0%	No	0.00	0.00
Quantity of Outgroup contact - Outgroup Response Bias	Higher outgroup contact increases outgroup response bias	Original Redux	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Original All	-	-	3	2	-0.19	-0.18	No	Slight	0.163	-	-0.45	0.08	78.39%	Yes	0.00	0.04
Quality of Outgroup contact - Difference Scores	Higher quality of outgroup contact decreases the observed OENE	Original Redux	6	1	30	2	-0.07	-0.07	Yes	Slight	0.070	-	-0.14	0.01	0%	No	0.00	0.00
		Original All	-	-	36	3	-0.04	-0.04	Yes	Slight	0.003	**	-0.07	-0.01	0%	No	0.00	0.00
Quality of Outgroup contact - Outgroup Discriminability	Higher quality of outgroup contact increases outgroup discrimination	Original Redux	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Original All	-	-	2	1	-0.12	-0.12	No	Slight	0.314	-	-0.35	0.11	0%	No	0.00	0.00

Note. Level of significance is indicated via *<.05, **<.01, ***<.001.

OENE-Prejudice

Implicit Outgroup Prejudice – Identification Difference Scores

The aggregate effect for the relationship between implicit prejudice and identification difference scores was small/slight and non-significant (*r*= .07, p>.05). This may be due to the small sample size (n=18 effects from a single study). Increases in implicit prejudice were associated with a larger OENE, or greater difference scores. Hypothesis 3a was therefore only descriptively supported. Effect size heterogeneity was non-significant (p>.05). Quantity of outgroup contact was a significant moderator (see Appendix AW)

Explicit Outgroup Prejudice – Identification Difference Scores

The aggregate effect for the relationship between explicit prejudice and identification difference scores was slight/small and non-significant (*r*=-.01, p>.05). Lower explicit outgroup prejudice is associated with a decrease in the OENE, or the size of the identification difference scores. Hypothesis 3a was therefore only descriptively supported for explicit outgroup prejudice. Effect size heterogeneity was non-significant however significant moderators were observed namely, (a) quantity of outgroup contact; (b) outgroup contact within the critical period; (c) Implicit prejudice; (d) Task; and (e) sample age were significant moderators (see Appendix AX). Publication bias was not present in the sample[42]

Explicit Outgroup Prejudice – Outgroup Discrimination

The aggregate effect for the relationship between explicit prejudice and outgroup discrimination was slight/small and non-significant (*r*=.03, p>.05). Effect size heterogeneity was considerable and statistically significant supporting the need for a moderator analysis (I^2 = 73.36%, p<.05). Lower explicit outgroup prejudice, or conversely higher favourable outgroup attitudes, are associated with increased outgroup discrimination i.e. greater overall accuracy for outgroup targets. While the aggregate estimate descriptively supports Hypothesis 3b, the non-significance thereof and considerable heterogeneity lend only tentative descriptive support for the hypothesis (see Appendix AAA). No significant moderators were found.

---

[42] See Appendix AY for additional publication bias plots and/or checks. See Appendix AZ for diagnostics.

OENE – Outgroup Prejudice Summary

The results of all meta-analytic models on the OENE-outgroup prejudice relationship are summarized in Table 6.

Table 6

Summary of Aggregate Effects for OENE – Outgroup Prejudice

													95% Cis			Heterogeneity		
Outcome	Hypothesis	Data	Number of Excluded Effects Post Diagnostics	Number of Excluded Articles Post Diagnostics	Number of effects	Number of studies (k)	Aggregate Effect - Fishers Z	Aggregate Effect - Pearson's r	Expected Direction	Strength Interpretation	p	Significance Level	LB	UB	I²	Significant Cochrane's Q	Between Study Heterogeneity (τ^2)	Within Study Heterogeneity
Implicit outgroup prejudice - Difference scores	Higher implicit outgroup prejudice increases the OENE	Original All	-	-	18	1	0.07	0.07	Yes	Slight	0.110	-	-0.02	0.16	0	No	0.00	0.00
Explicit outgroup prejudice - Difference scores	Lower explicit outgroup prejudice will be associated with a smaller OENE	Original Redux	26	1	26	4	-0.01	-0.01	Yes	Slight	0.709	-	-0.05	0.04	32.41	No	0.00	0.00
		Original All	-	-	52	5	-0.04	-0.04	Yes	Slight	0.088	-	-0.10	0.01	55.48	Yes	0.00	0.02
Explicit outgroup prejudice - outgroup discrimination	Lower outgroup prejudice will be associated with increased outgroup discrimination	Original All	-	-	12	7	0.03	0.03	Yes	Slight	0.535	-	-0.07	0.13	73.36	Yes	0.00	0.02

Note. Level of significance is indicated via *<.05, **<.01, ***<.001.

Outgroup Contact- Outgroup Prejudice

Implicit Outgroup Prejudice – Quantity of Outgroup Contact

The aggregate effect for the relationship between implicit prejudice and contact was slight/small and non-significant (r=-.13, $p>.05$). Effect size heterogeneity was non-significant ($p>.05$). Higher levels of implicit outgroup prejudice are associated with a reduction in quantity of outgroup contact. Owing to the non-significance of the aggregate effect, Hypothesis 4a was therefore only descriptively supported. No publication bias was present in the sample. Fixed or self-paced encoding, motivation and task were significant moderators (see Appendix AAB).

Explicit Outgroup Prejudice – Quantity of Outgroup Contact

The aggregate effect for the relationship between explicit prejudice and contact was slight/small and non-significant (r=.05, $p>.05$). Effect size heterogeneity was considerable and significant (I^2 =86.9%). Lower explicit outgroup prejudice was associated with increased outgroup contact however, owing to non-significance of the aggregate effect and considerable heterogeneity, Hypothesis 4a was descriptively only tentatively supported. Publication bias was not present in the sample (see Appendix AAC).

Outgroup Contact – Outgroup Prejudice Summary

The results of all meta-analytic models on the outgroup contact-outgroup prejudice relationship are summarized in Table 7.

Table 7

Summary of Aggregate Effects for Outgroup Contact – Outgroup Prejudice

Outcome	Hypothesis	Data	Number of Excluded Effects Post Diagnostics	Number of Excluded Articles Post Diagnostics	Number of effects	Number of studies (k)	Aggregate Effect - Fishers Z	Aggregate Effect - Pearson's r	Expected Direction	Strength Interpretation	p	Significance Level	95% Cis LB	95% Cis UB	Heterogeneity I²	Heterogeneity Significant Cochrane's Q	Heterogeneity Between Study Heterogeneity (τ²)	Heterogeneity Within Study Heterogeneity
Implicit outgroup prejudice - quantity of outgroup contact	Higher implicit prejudice will be associated with a reduction in outgroup contact	Original All	-	-	20	3	-0.13	-0.13	Yes	Slight	0.120	-	-0.29	0.03	0	No	0.01	0.00
Explicit outgroup prejudice - quantity of outgroup contact	Higher favourable outgroup attitudes (lower explicit prejudice) is associated with increased outgroup contact	Original All	-	-	55	7	0.05	0.05	Yes	Slight	0.686	-	-0.21	0.32	86.9	Yes	0.11	0.02

Note. Level of significance is indicated via *<.05001.

Table 8

Aggregate Effect for Implicit Outgroup Prejudice – Explicit Outgroup Prejudice

Outcome	Hypothesis	Data	Number of Excluded Effects Post Diagnostics	Number of Excluded Articles Post Diagnostics	Number of effects	Number of studies (k)	Aggregate Effect - Fishers Z	Aggregate Effect - Pearson's r	Expected Direction	Strength Interpretation	p	Significance Level	95% Cis LB	95% Cis UB	Heterogeneity I²	Heterogeneity Significant Cochrane's Q	Heterogeneity Between Study Heterogeneity (τ²)	Heterogeneity Within Study Heterogeneity
Implicit - Explicit Prejudice	Higher implicit prejudice will be associated with lower favourable outgroup attitudes	Original All	-	-	20	3	-0.10	-0.10	Yes	Slight	0.055	-	-0.20	0.00	22.97	No	0.00	0.01

Note. Level of significance is indicated via *<.05, **<.01, ***<.001. Explicit prejudice was coded as higher scores being indicative of more favourable outgroup attitudes i.e. lower explicit outgroup prejudice. Therefore, the inverse relationship observed signifies that as explicit outgroup prejudice increases there is an associated increase in implicit outgroup prejudice.

Outgroup Prejudice

Implicit Outgroup Prejudice – Explicit Outgroup Prejudice

The aggregate effect for the relationship between implicit and explicit prejudice was slight/small and non-significant (r=-.10, p>.05). Effect size heterogeneity was non-significant (p>.05). Higher implicit outgroup prejudice is associated with higher explicit outgroup prejudice however, this association was non-significant. Publication bias was not present within the sample and (see Appendix AAD[43])

Implicit Outgroup Prejudice– Explicit Outgroup Prejudice Summary

The results of all meta-analytic models on the outgroup implicit-explicit prejudice relationship are summarized in Table 8.

Correlations

Data was not available on all combinations of outgroup prejudice, outgroup contact and outgroup identification performance. Therefore, to explore all possible relationships and to facilitate a point of reference for meta-analytic models with small samples, correlations were calculated to assess trends (see Table 9).

[43] See Appendix AAE for Cooks Distance/diagnostics

Table 9

Matrix Containing both Correlations and Aggregate Effects for Outgroup Prejudice, Outgroup Contact and Outgroup Identification Performance

	Quantity outgroup contact	Quality outgroup contact	Implicit outgroup prejudice	Explicit outgroup prejudice	Outgroup hits	Outgroup false alarms	Outgroup response bias	Outgroup discrimination
Quantity outgroup contact	-		-0.13 -	0.05 *	0.10 **	*0.06*	*-0.18*	0.10 **
Quality outgroup contact	0.21 **	-						*-0.12*
Implicit outgroup prejudice	-0.30 ***	-0.39 ***	-	-0.10 -				
Explicit outgroup prejudice	0.27 ***	0.05	-0.15	-				
Outgroup hits	0.19 -	0.22	-0.19	*-0.03*	-			
Outgroup false alarms	-0.58 ***	-0.20	0.44 *	*0.04*	-0.65 ***	-		
Outgroup response bias	0.38 *	0.28	-0.38 *	*-0.12*	-0.16	-0.49 **	-	
Outgroup discrimination	0.32 **	0.27 *	-0.27 *	0.26 *	0.81 ***	-0.84 ***	0.19 -	-

Note. Bottom left triangle represents calculated correlation coefficients using completed outgroup contact and completed outgroup prejudice values (combined option 2).

Stars for p indicate *<.05, **<.01, ***<.001.

Italics represent a relationship counter to expectations. The top right triangle includes meta-analytic aggregate effects as a point of comparison

CHAPTER 4: DISCUSSION

This research evaluated fifty-three years of empirical findings on the outgroup contact, OENE and outgroup prejudice relationships. Forty-one articles, containing fifty-four studies and 8418 participants were analyzed. Participant base performance on in- and outgroup target faces was assessed using hit rates (target accuracy), false alarm rates (errors in identifications), discriminability (overall target accuracy[44]), and response bias (How likely an individual is to choose a target or make an identification[45]). The sample demonstrated classic patterns of an OENE with participants having (a) greater target accuracy for in-group members (Lee & Penrod, 2022; Meissner & Brigham, 2001; Wilson et al., 2013); (b) a greater number of identification errors for outgroup members (Evans et al., 2009; Howard et al., 2019; Lee & Penrod, 2021; Meissner & Brigham, 2001;Wilson et al., 2013); (c) higher overall target accuracy for in-group members (Howard et al., 2019; Jackiw et al., 2008; Lee & Penrod, 2022; Pica et al., 2015; Semplonius & Mondloch, 2013; Wilson et al., 2013); and (d) a more conservative decision-strategy i.e. less willingness to choose an in-group target (Jackiw et al., 2008; Lee & Penrod, 2022; Meissner & Brigham, 2001; Meissner et al., 2005; Slone et al., 2000; Wilson et al., 2013). Hit rates, or target accuracy was the only non-significant effect and this was likely due to a limited sample (n = 3 effects from 2 articles). The size of the OENE for the above-mentioned identification indices differed to prior research (Lee & Penrod; 2021; Meissner & Brigham, 2001). While prior research observed a (a) small to medium effect size for hit rates, false alarm rates and response bias; and (b) observed the largest OENE effect for discriminability i.e. overall target accuracy (medium to large); the largest OENE effect size observed in this sample was false alarms (medium to large). Across OENE base performance analyses, outgroup contact and prejudice were found to significantly moderate the OENE effect size i.e. poorer identification performance for outgroup members. Quality and quantity of outgroup contact significantly reduced the size of the OENE (Singh et al., 2021; Stelter et al., 2022; Walker & Hewstone, 2006b). Outgroup contact later in life significantly increased the OENE, lending support to the presence of a critical period in which outgroup contact must occur to ensure native or in-group level identification performance for outgroup members (McKone et al., 2019). Higher outgroup prejudice, both implicit and explicit prejudice, significantly increased the observed OENE (Ma et al., 2011).

[44] A standardized composite score which accounts for both target accuracy and mistakes.
[45] A standardized composite score.

Contact-OENE

Outgroup contact is an important theoretical driver of the OENE for both perceptual expertise and socio-cognitive models of the OENE (Singh et al., 2021; Young et al., 2012; Zhou et al., 2014). Prior studies have however consistently demonstrated a small relationship between outgroup contact and the OENE (Singh et al., 2022). The findings of this research were consistent with the wider literature in that a small i.e. slight relationship was observed between outgroup contact and the OENE[46]. Such findings are at odds with not only the theoretical importance placed on outgroup contact in accounting for the emergence of the OENE but are also at odds with the associated gains of outgroup contact namely, (a) a reduction in the observed OENE; and/or (b) an increase in outgroup identification performance (Chiroro, 1994; Bukach et al., 2012; Hancock & Rhodes, 2008; Meissner & Brigham, 2001; Pezdek et al., 2003; Singh et al., 2021; Slone et al., 2000; Stelter et al., 2022; Walker & Hewstone, 2006b; Wright et al., 2003; Zhao et al., 2014b) .Thus, beyond an aggregation of findings to date, this research sought to unpack the counterintuitive findings of a slight/small relationship between outgroup contact and the OENE. This was achieved, in part, via accounting for the influence of outgroup prejudice on the outgroup contact-OENE relationship. Explorations of the influence of outgroup prejudice have been limited via the limited number of studies which measure and report both outgroup prejudice and the outgroup contact-OENE relationship. Within the sample, 72% of coded data did not measure implicit outgroup prejudice whilst 93.33% did not measure explicit outgroup prejudice. To circumvent such limitations, this research used matched-to-sample outgroup prejudice data. Based on the findings of the analysis, it is argued that the outgroup contact-OENE relationship is complex and that such complexities should be apparent within the analysis. More specifically, by accounting for the influence of outgroup prejudice, the time span in which outgroup contact occurred and motivation for example, the face-value counterintuitive finding of a weak relationship between outgroup contact and the OENE can be better understood.

Quantity of Contact – Difference Scores

The aggregate estimate for the relationship between quantity of outgroup contact and identification difference scores (r=-.11, 95%, p<.001, CI [-.16, -.05]) was consistent with the results of the Singh et al. (2021) meta-analysis (r=-.15). Higher levels of outgroup contact

---

[46] Higher outgroup contact was associated with a reduction in the OENE within the sample.

were significantly associated with a reduction in the OENE. The associated benefits of outgroup contact, namely a trending reduction in the OENE, persisted when participants were not motivated to individuate outgroup faces thereby strengthening the support for the beneficial influence of outgroup contact. Majority members typically exhibit a lower degree of outgroup contact when compared to minority members. Majority members, or participants with a lower predicted level of outgroup contact, were associated with a greater reduction in the OENE when outgroup contact increased. This was further supported by the significance of sample positionality i.e. Global North samples as a moderator[47].

The observed small/slight relationship was further explained by the moderating influence of outgroup prejudice and time span in which outgroup contact occurred. Low explicit outgroup prejudice significantly strengthened this relationship, thereby increasing the associated identification gains of outgroup contact. Similarly, outgroup contact that occurred below the age of 12, the upper limit of the theorized critical period for beneficial outgroup contact, and high levels of outgroup contact within the critical period are both significantly associated with strengthening the trending identification gains of greater outgroup contact. This lends further support to the claims regarding the importance of outgroup contact within the critical period (McKone et al., 2019). In regard to the specific time period in which outgroup contact occurred, a wider trend was noted within the data wherein for both quantity and quality of outgroup contact, lifetime contact was oversampled to the detriment of other time bands. Considering the importance of critical time period, future studies should utilize a more varied assessment of outgroup contact which includes 0-12 years of age.

Quantity of Outgroup Contact – Discrimination

The aggregate estimate for the relationship between quantity of outgroup contact and outgroup discrimination (r=.10, p<.01, 95% CI [.02, .17]) was consistent with the results of the Meissner and Brigham (2001) meta-analysis (r=.13). Higher levels of outgroup contact were significantly associated with higher outgroup overall identification accuracy. This relationship persisted when no motivation manipulation was in use, again affirming the importance of outgroup contact in reducing the OENE. Variations in the observed small/slight relationship were accounted for when the positionality of the participants was considered. For majority members who were tested on minority outgroup members, thus

[47] Sample positionality i.e. belonging to the Global North or South accounted for 51.98% of the variance in the meta-analytic model. Therefore, it is an important explanatory variable for fluctuations in the outgroup contact – identification difference scores relationship

members with supposedly lower historical outgroup contact, outgroup contact was associated with greater outgroup discrimination gains.

Quantity of Outgroup Contact – Hits

The slight/small relationship between outgroup quantity of contact and outgroup hits was significant (*r*=.10, p<.01, 95% CI [.03, .16]). Greater levels of outgroup contact were associated with greater correct outgroup identifications, or hits. The strength of the observed relationship was largely explained by accounting for the moderating influence of outgroup prejudice which accounted for 80.80% of the variance explained in the meta-analytic model. Despite the associated identification gains of outgroup contact, in the presence of implicit outgroup prejudice, such gains are lost. For majority members who are tested on minority outgroup faces, thus with presumed lower historical outgroup contact, outgroup contact was associated with larger outgroup identification gains (20.87% variance explained).

Quality of Outgroup Contact – Difference Scores

Quality of outgroup contact was understudied and/or underreported within the sample, with only 40% of the sample including measurements of both outgroup quantity and quality of contact. Both types of outgroup contact should be assessed simultaneously given the theoretical importance of both quantity and quality of contact.

Greater quality of outgroup contact was associated with a reduction the OENE (*r*=-.07, p>.05, 95% CI [-.14, .01]). The associated identification gains of greater quality of outgroup contact persisted when no motivation instructions or manipulations were used thus strengthening the importance of quality of outgroup contact. Sample positionality was again an important consideration in explaining the small/slight observed relationship. Majority members who were tested on minority outgroup faces, had higher associated identification gains for quality of outgroup contact.

Quantity of Outgroup Contact – Response Bias

While the aggregate estimate for the relationship between quantity of outgroup contact and outgroup response bias was not in the expected direction or significant[48], the calculated correlation coefficient was (*r*=.38, p<.05). This finding was likely the result of a small sample (n= 3 effects from 2 studies). The calculated correlation coefficient, which utilized imputation to estimate missing proportions of outgroup contact, supported the

---

[48] Meta-analytic estimate for contact – response bias was *r* =-.19, p>.05.

hypothesized pattern of results. Namely, that higher levels of outgroup contact are significantly associated with higher levels of outgroup response bias, i.e. a shift towards a conservative outgroup decision strategy. The amount of implicit outgroup prejudice significantly moderated the outgroup contact – response bias relationship. When implicit outgroup prejudice increases, the associated gains of outgroup contact on outgroup identification performance are lost i.e. the outgroup decision-making strategy reverts to a liberal criterion which fits the classic pattern of an OENE. Accounting for implicit prejudice explained 100% of the variation in the meta-analytic model and thus underscores the importance of accounting for outgroup prejudice in the outgroup contact-response bias relationship.

Quantity of Outgroup Contact – False Alarms

The aggregate estimate for the relationship between quantity of outgroup contact and outgroup false-alarms was not in the expected direction, nor was it significant[49]. This was likely the result of a small sample (n = 4 effects from 3 studies) which biased the aggregate findings. Conversely, the calculated correlation coefficient was significant and noted higher levels of outgroup contact being associated with a reduction in identification errors, or false-alarms (r=-.58, $p<.001$).

Quality of Outgroup Contact – Discrimination

The aggregate estimate for quality of outgroup contact and outgroup discrimination suffered from the inherent bias of a small sample (n=2 effects from 1 study). Consequently, the estimate was not significant or in the expected direction[50] whilst the calculated correlation coefficient was (r =.27, $p<.05$). For the calculated correlation, higher quality of outgroup contact was associated with gains in outgroup overall identification accuracy.

Conclusion for Contact-OENE

The totality of the outgroup contact-OENE analyses suggest that the importance of future studies measuring and accounting for the influence of outgroup prejudice within their analyses. It is evident that outgroup prejudice plays an important role in accounting for variation within the strength of the observed relationship. This research used Harvard's Project Implicit Database to explore the influence of prejudice, of which the prejudice data was matched to geographic location, age, ethnicity, and time period, and whilst the matched

[49] The meta-analytic estimate for outgroup contact – false alarms was r=.06, $p>.05$
[50] The meta-analytic estimate for outgroup contact – discrimination was r = -.12, $p>.05$

to sample data was deemed suitable for use in the analysis, proxy data cannot match the correctness of reported/actual data. Matched to sample prejudice data is the closest available approximation of non-measured outgroup prejudice. While it is the closest approximation, this data was limited in terms of (a) the degree of geographic match specificity, the smallest matched geographic unit was a metropolitan statistical tract within the United States while data for all other countries were only available at a country level; and (b) the degree of match specificity for studies published prior to 2002 or the start of the Harvard Project Implicit Database. Studies published prior to 2002 therefore had poorer approximations of outgroup prejudice owing to a larger year match difference. Stronger assertions regarding the moderating influence of outgroup prejudice therefore require growth of the extant literature.

While there are numerous pathways through which the outgroup contact-OENE relationship may be explored, more sample studies than not exhibited a preference for analyzing identification difference scores at the expense of other identification metrics[51]. Identification difference scores are typically computed by taking the difference of the outgroup identification performance from that of in-group identification performance. These calculations often use discrimination values or overall identification accuracy. While theoretically sound as (a) discrimination values are a more accurate standardized assessment of overall accuracy which considers not only correct identifications but also the amount of identification errors and (b) difference scores facilitate easier interpretations, nonetheless correlations with other identification metrics should not be discarded entirely. It is argued that such correlations should be included as a supplemental analysis in order to strengthen conclusions drawn regarding the strength and consistency of the outgroup contact-OENE relationship.

OENE-Prejudice

Explicit Outgroup Prejudice – Discrimination

Higher levels of favourable outgroup attitudes, or conversely low explicit outgroup prejudice, was associated with higher outgroup discrimination or overall identification accuracy (r=.03, p>.05, 95% CIs [-.07, .13]). The observed relationship was small and consistent with the findings of the Meissner and Brigham (2001) meta-analysis which found a small relationship between favourable outgroup attitudes and outgroup discriminability (r=-

[51] Identification metrics refer to outgroup hits, outgroup false alarms, and outgroup discrimination.

.01). The calculated correlation coefficient indicated a more substantial relationship (*r*=.26, p<.05).

Explicit Outgroup Prejudice – Difference Scores

More favourable outgroup attitudes, or lower explicit outgroup prejudice, was associated with a reduction in the OENE (*r*=-.01, p>.05, 95% CIs [-.05, .04]). This slight/small relationship was significantly influenced by the level of outgroup contact. When there is increased outgroup contact, the associated gains of lower explicit outgroup prejudice on identification accuracy are strengthened. The importance of outgroup contact was underscored by the amount of variance explained by outgroup contact within the meta-analytic model (original study data: 30.49%, using missing contact values estimated via imputation: 27.15%). Quality of contact was equally as important. Higher levels of quality of contact were associated with greater reductions in the observed OENE when the sample had low explicit outgroup prejudice. Quality of outgroup contact explained 22.77% of the variance in the meta-analytic model.

The most important moderator, in terms of variance explained, was the time span in which outgroup contact occurred. Consistent with the theorized critical period of contact, high levels of outgroup contact within the critical period were associated with significant reductions in the OENE when explicit outgroup prejudice was low. Implicit prejudice was also an important moderator of the explicit prejudice – identification difference scores relationship. As implicit outgroup prejudice increased, the associated outgroup identification gains from low explicit prejudice were lost.

Implicit Outgroup Prejudice – Difference Scores

The aggregate estimate for the relationship between implicit outgroup prejudice and identification difference scores was slight and non-significant (*r*=.07, p>.05, 95% CIs [-.02, .16]). This finding is consistent with prior research (Ma et al, 2011). Variations in the aggregate effect were a result of the influence of quantity of outgroup contact. When quantity of outgroup contact was high, higher implicit outgroup prejudice was associated with a larger OENE. Quantity of outgroup contact accounted for 10.02% of the variation in the meta-analytic model.

Conclusion for OENE-Prejudice

Data from the OENE-prejudice relationships again affirms the interconnected nature of the contact-OENE-prejudice relationships and the importance of accounting for outgroup contact when assessing the OENE-prejudice relationship. Within the limited sample that reported a correlation coefficient for the outgroup contact-outgroup prejudice relationship, a wider preference was observed wherein explicit prejudice was favoured over implicit prejudice in assessments of the outgroup contact-outgroup prejudice relationship. It is argued that both types of outgroup prejudice should be assessed simultaneously in order to strengthen assertions regarding the role of outgroup prejudice on outgroup identification performance.

Prejudice-Contact

Explicit Outgroup Prejudice – Quantity of Contact

More favourable outgroup attitudes, or lower explicit outgroup prejudice was associated with an increase in quantity of outgroup contact (r=.05, p>.05, 95% CIs [-.21, .32]). This finding was consistent with other meta-analyses in terms of the direction of the relationship (Hseih et al., 2022; Meissner & Brigham, 2001; Zhou, 2018). Prior research has however observed a larger effect size for the explicit prejudice – outgroup contact relationship. The calculated correlation coefficient (r=.27, p<.001) is therefore more aligned with the wider literature findings.

Implicit Outgroup Prejudice – Quantity of Contact

The aggregate estimate for the relationship between implicit outgroup prejudice and quantity of outgroup contact was slight/small (r=-.13, p>.05, 95% CIs [-.29, .03]). As implicit outgroup prejudice increases, the quantity of outgroup contact reduces. This finding is consistent with the Pettigrew and Tropp (2006) meta-analysis. The calculated correlation coefficient demonstrated a stronger relationship between implicit prejudice and contact (r=-.30, p<.001)

Overall Summary and Limitations

All results speak to the importance of acknowledging the tri-directional relationship between outgroup contact, outgroup prejudice and the OENE. While prior meta-analyses have mainly focused on the constituent parts of these relationships, namely studied the relationships of interest separately, this meta-analysis is novel in that it addressed all relationships concurrently alongside matched-to-sample prejudice data. As a result of the findings of this analysis, it is argued that as a baseline going forward, any assessments of two out of the three core variables must include and/or account for the influence of the remaining core variable. Only via a more nuanced assessment of such relationships will greater clarity be reached regarding the causal mechanisms of the OENE. In addition, future studies should develop a theoretical model for this three-way relationship which explores both the likelihood of casual connections and the mechanisms through which they are casually connected.

Motivation was rarely manipulated in the sampled studies namely, only 25.33% included a motivation manipulation or instruction. This limits assertions that can be drawn regarding the core relationships and the presence of motivation. The presence of motivation could for example explain variation in the small outgroup contact-OENE relationship by strengthening the associated identification gains of outgroup contact to a much greater degree than a no motivation condition. It is thus important that future studies include motivation manipulations. Outgroup contact did however attenuate the observed OENE when no motivation condition was present, thus lending greater credence to socio-cognitive, or hybrid, theoretical models of the OENE.

While not always explicitly reported and/or included within the study design, attention checks should be used to validate participants are actively attending the study and test faces. There is a presumed interplay between attention and motivation, which consequently could have a knock-on effect on the observed OENE (Engelmann et al., 2009). Future studies could explore this further via manipulations of both attention and motivation. Such data would be beneficial in unpacking the exact casual mechanisms that underlie the emergence of the OENE.

A major limitation of this analysis was the inability to simultaneously run a meta-regression with multiple moderators. Such an analysis requires complete data. With varying patterns of incomplete data, moderators could only be assessed individually in order to maximally preserve the coded data. While valuable insights can still be gained from this style of analysis, it is inherently limited in its ability to calculate the unique contribution of strongly correlated moderators.

Another intrinsic limitation of working with self-report outgroup contact data, which comprised the majority of the sampled data, is the associated difficulty of retroactively self-reporting level of contact at various stages of the life-span (Holtman et al., 2005). Retroactively self-reporting contact between the ages of five to twelve, twelve to eighteen and eighteen plus is a difficult task with greater unreliability expected in the self-report for the earliest age group. With any self-report data, participants may under or over report either from task difficulty or from the need to appear socially acceptable in their responses. Explicit self-report measures of explicit outgroup prejudice will by extension be equally vulnerable to the potential unreliability introduced via self-report data.

When critically reflecting on the studies included in the analysis, several systematic limitations, that impact on the quality of results, become apparent. Noticeably there is poor representation of both target-face and participant ethnic-nationalities other than White and Black, which limits the generalizability of the findings. Secondly, outgroup contact is rarely observed naturalistically and/or is rarely manipulated experimentally. Measurements of outgroup contact are most frequently conducted on a post-hoc basis via methods of self-report. Consequently, there is a need for more diverse measurements of outgroup contact. Greater variability in measurements would strengthen evidence on the outgroup contact, outgroup identification and outgroup prejudice relationships.

The presence of outgroup identification deficits, or the OENE, highlights the critical need for specialized training for law enforcement officers to minimize identification errors. Additionally, it is imperative to educate jurors about the potential biases and dangers associated with outgroup identifications to ensure fair and accurate judicial outcomes – particularly with regard to the extent of the evidentiary weight attributed to outgroup identifications.

REFERENCES

References marked with an asterisk (*) indicate articles that were included in the meta-analysis

Aaron, B., Komrey, J., & Ferron, J. (1998, November). Equating “r”-based and “d”-based Effect Size Indices: Problems with a Commonly Recommended Formula. Paper presented at the Annual Meeting of Florida Educational Research Association, Orlando, Florida.

*Adams, R. B., Jr., Pauker, K., & Weisbuch, M. (2010). Looking the other way: The role of gaze direction in the cross-race memory effect. Journal of Experimental Social Psychology, 46(2), 478. https://www.proquest.com/scholarly-journals/looking-other-way-role-gaze-direction-cross-race/docview/214785705/se-2 https://uct.primo.exlibrisgroup.com/discovery/openurl?institution=27UCT_INST&vid=27UCT_INST:27UCT&?genre=article&issn=00221031&title=Looking+the+other+way%3A+The+role+of+gaze+direction+in+the+cross-race+memory+effect&volume=46&issue=2&date=2010&atitle=Looking+the+other+way%3A+The+role+of+gaze+direction+in+the+cross-race+memory+effect&spage=478&sid=ProQ%3Aibss&author=Adams

Agadullina, E. R., & Lovakov, A. V. (2018). Are people more prejudiced towards groups that are perceived as coherent? A meta-analysis of the relationship between out-group entitativity and prejudice. The British Journal of Social Psychology. 57(4), 703–731. https://doi.org/10.1111/bjso.12256

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons.

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

Blanton, J. J., & Burrows, C. N. (2015). Implications of the Implicit Association Test D-Transformation for Psychological Assessment. Assessment (Odessa, Fla.), 22(4), 429–440. https://doi.org/10.1177/1073191114551382

Brigham, J. C. (1993). College students’ racial attitudes. *Journal of Applied Social Psychology*, *23*(23), 1933-1967.

*Brigham, J. C., & Barkowitz, P. (1978). Do 'They all look alike'? The effect of race, sex, experience, and attitudes on the ability to recognize faces. Journal of Applied Social Psychology, 8(4), 306-318. https://doi.org/10.1111/j.1559-1816.1978.tb00786.x

Brigham, J. C., & Malpass, R. S. (1985). The Role of Experience and Contact in the Recognition of Faces Of Own- and Other-Race Persons. Journal of Social Issues, 41(3), 139–155. https://doi.org/10.1111/j.1540-4560.1985.tb01133.x

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

*Carroo, A. W. (1986). Other race recognition: a comparison of black American and African subjects. Percept Mot 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

Carter, E., Schonbrodt, F., Gervais, W., & Hilgard, J. (2019). Correcting for Bias in Psychology: A Comparison for Meta-Analytic Methods. Advances in Methods and Practices in Psychological Science 2019, 2(2), 115–144. doi/10.1177/2515245919847196

Cashin, S. (2010). To Be Muslim or "Muslim-Looking" in America: A Comparative Exploration of Racial and Religious Prejudice in the 21st Century (Unpublished thesis). Georgetown University Law Center, Georgetown. https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=2713&context=facpub

*Cavazos, J. G. (2020). Does Race Matter? Testing Parameters for Collaboration Benefits on Face Identification Accuracy (Publication Number 28390883) [Ph.D., The University of Texas at Dallas]. ProQuest Dissertations & Theses A&I. Ann Arbor. https://www.proquest.com/dissertations-theses/does-race-matter-testing-parameters-collaboration/docview/2489261080/se-2?accountid=14500 https://uct.primo.exlibrisgroup.com/discovery/openurl?institution=27UCT_INST&vid=27UCT_INST:27UCT&?genre=dissertations&issn=&title=Does+Race+Matter%3F+Testing+Parameters+for+Collaboration+Benefits+on+Face+Identification+Accuracy&volume=&issue=&date=2020&atitle=&spage=&sid=ProQuest+Dissertations+%26+Theses+A%26I&author=Cavazos

Chien, S., Tai, C., & Yang, S. (2018). The development of the own-race advantage in school-age children: A morphing face paradigm. PloS One, 13(4), e0195020–e0195020. https://doi.org/10.1371/journal.pone.0195020

Chiroro, P. (1994). Individual differences in recognition memory for faces (Unpublished doctoral dissertation). Durham University, Durham.

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

*Chorney, M. (2015). The Factors of Face Perception and Recognition of Distant Ethnic Groups. Procedia - Social and Behavioral Sciences, 190, 234-240. https://doi.org/https://doi.org/10.1016/j.sbspro.2015.04.941

Cooper, H., & Dent, A. (2011). Ethical Issues in the Conduct and Reporting of Meta-Analysis. In A.T Panter & S.K Sterba (Eds.), Handbook of Ethics in Quantitative Methodology (pp. 437–464). Routledge. https://doi.org/10.4324/9780203840023-28

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). The handbook of research synthesis and meta-analysis (3rd edition.). Russell Sage Foundation.

*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/https://doi.org/10.1016/j.jecp.2005.09.001

Cortiz, A., Peña, J., Jacobs, P., Rico, B., Hahn, J., & Lowe, R. (2023, November 1). Census Bureau Releases 2020 Census Population for More Than 200 New Detailed Race and Ethnicity Groups. Census.gov. Retrieved February 3, 2024, from https://www.census.gov/library/stories/2023/09/2020-census-dhc-a-race-overview.html

*Cruz, F., Palma, T. A., Bansemer, E., Correll, J., Fonseca, S., Gonçalves, P., & Santos, A. S. (2022). Do individuation instructions reduce the cross-race effect? A registered replication of Hugenberg, Miller, and Claypool (2007). Journal of Experimental Social Psychology, 104, 104423. https://doi.org/https://doi.org/10.1016/j.jesp.2022.104423

*Dai, J. (2021). Race Bias in Adolescent Face Recognition: The Role of Friendship Diversity, Ethnic-Racial Identity, Ethnic-Racial Socialization, and Perceived Discrimination (Publication Number 29049938) [Ph.D., The Pennsylvania State University]. ProQuest Dissertations & Theses A&I. Ann Arbor.

https://www.proquest.com/dissertations-theses/race-bias-adolescent-face-recognition-role/docview/2655150911/se-2?accountid=14500 https://uct.primo.exlibrisgroup.com/discovery/openurl?institution=27UCT_INST&vid=27UCT_INST:27UCT&?genre=dissertations&issn=&title=Race+Bias+in+Adolescent+Face+Recognition%3A+The+Role+of+Friendship+Diversity%2C+Ethnic-Racial+Identity%2C+Ethnic-Racial+Socialization%2C+and+Perceived+Discrimination&volume=&issue=&date=2021&atitle=&spage=&sid=ProQuest+Dissertations+%26+Theses+A%26I&author=Dai

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic Out-Group Faces Are Biased in the Prejudiced Mind. Psychological Science., 19(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and Explicit Prejudice and Interracial Interaction. Journal of Personality and Social Psychology, 82(1), 62–68. https://doi.org/10.1037/0022-3514.82.1.62

Durrheim, K., Quayle, M., & Dixon, J. (2016). The Struggle for the Nature of “Prejudice”: “Prejudice” Expression as Identity Performance. Political Psychology. 37(1), 17–35. https://doi.org/10.1111/pops.12310

Evans, J. R., Marcon, J. L., & Meissner, C. A. (2009). Cross-racial lineup identification: assessing the potential benefits of context reinstatement. Psychology, Crime & Law, 15(1), 19–28. https://doi.org/10.1080/10683160802047030

Ferguson, D. P., Rhodes, G., Lee, K., & Sriram, N. (2001). “They all look alike to me”: Prejudice and cross-race face recognition. The British Journal of Psychology, 92(4), 567–577. https://doi.org/10.1348/000712601162347

*Finklea, K. M. (2008). Investigating inappropriate cue utilization in the own -race bias (Publication Number 3307561) [Ph.D., University of California, San Diego]. ProQuest Dissertations & Theses A&I. Ann Arbor. https://www.proquest.com/dissertations-theses/investigating-inappropriate-cue-utilization-own/docview/89152764/se-2?accountid=14500 https://uct.primo.exlibrisgroup.com/discovery/openurl?institution=27UCT_INST&vid=27UCT_INST:27UCT&?genre=dissertations&issn=&title=Investigating+inappropriate+cue+utilization+in+the+own+-race+bias&volume=&issue=&date=2008&atitle=&spage=&sid=ProQuest+Dissertations+%26+Theses+A%26I&author=Finklea

Garrett, B. L., Liu, A., Kafadar, K., Yaffe, J., & Dodson, C. S. (2020). Factoring the Role of Eyewitness Evidence in the Courtroom. *Journal of Empirical Legal Studies.*, *17*(3), 556–579. https://doi.org/10.1111/jels.12259

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. Journal of Personality and Social Psychology, 74(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

*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

Haney, C. (2002, January). The Psychological Impact of Incarceration: Implications for Post-Prison Adjustment. Paper presented at the 'From Prison to Home' Conference, National Institutes of Health Natcher Conference Center, United States. https://aspe.hhs.gov/sites/default/files/migrated_legacy_files/42351/Haney.pdf

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D.D. (2021). Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton, FL and London: Chapman & Hall/CRC Press. ISBN 978-0-367-61007-4.

*Havard, C., Memon, A., & Humphries, J. E. (2017). The own-race bias in child and adolescent witnesses: Evidence from video line-ups [Article]. International Journal of Police Science and Management, 19(4), 261-272. https://doi.org/10.1177/1461355717731579

Heesen, P. (2020, September 21). Crossover trials: what are they and what are their advantages and limitations? - Students 4 Best Evidence. Students 4 Best Evidence. https://s4be.cochrane.org/blog/2020/09/07/crossover-trials-what-are-they-and-what-are-their-advantages-and-limitations/

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2023). Cochrane Handbook for Systematic Reviews of Interventions (version 6.4). Cochrane. www.training.cochrane.org/handbook

Holtman, Z., Louw, J., Tredoux, C., & Carney, T. (2005). Prejudice and Social Contact in South Africa: A Study of Integrated Schools Ten Years after Apartheid. *South African Journal of Psychology.*; 35(3):473-493. doi:10.1177/008124630503500306

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2013). Same faces, different labels: Generating the cross-race effect in face memory with social category information. Memory & Cognition, 41(7), 1021–1031. https://doi.org/10.3758/s13421-013-0316-7

Howard, S., Thomas, A. K., & Sommers, S. R. (2019). “They all still look the same to me”: Navon processing and the cross-race effect. Journal of Cognitive Psychology (Hove, England), 31(8), 839–851. https://doi.org/10.1080/20445911.2019.1679156

Hsieh, W., Faulkner, N., & Wickes, R. (2022). What reduces prejudice in the real world? A meta-analysis of prejudice reduction field experiments. *British Journal of Social Psychology*, *61*(3), 689–710. https://doi.org/10.1111/bjso.12509

Hugenberg, K., & Sacco, D. F. (2008). Social Categorization and Stereotyping: How Social Categorization Biases Person Perception and Face Memory. Social and Personality Psychology Compass, 2(2), 1052–1072. https://doi.org/10.1111/j.1751-9004.2008.00090.x

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

Innocence Project. (n.d.). DNA exonerations in the United States. https://innocenceproject.org/dna-exonerations-in-the-united-states/

*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 Comportment, 40(1), 52-57. https://doi.org/10.1037/0008-400x.40.1.52

*Jerovich, A. (2017). Implicit attitudes and the other race effect Western Sydney University (Australia)].

Johnson, S. L. (1984). Cross-racial identification errors in criminal cases. Cornell Law Review, 69(5), 934–987.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007a). The Other-Race Effect Develops during Infancy: Evidence of Perceptual Narrowing. Psychological Science, 18(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kelly, D., Liu, S., Ge, L., Quinn, P., Slater, A., Lee, K., Liu, Q., & Pascalis, O. (2007b). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. Infancy. 11(1). https://doi.org/10.1080/15250000709336871

*Kovalenko, A., & Surudzhii, M. (2014). Cross-race Effect: The Role of Social and Individual Factors in Face Recognition Process. Procedia - Social and Behavioral Sciences, 114, 136-140. https://doi.org/https://doi.org/10.1016/j.sbspro.2013.12.673

Kramer, R. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition.*, *172*, 46–58. https://doi.org/10.1016/j.cognition.2017.12.005

Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. The British Journal of Psychology, 107(2), 374–388. https://doi.org/10.1111/bjop.12147

*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

Lee, J., & Penrod, S. (2022). Three-level meta-analysis of the other-race bias in facial identification. *Applied Cognitive Psychology*, *36*(5), 1106-1130.

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

Lim, C., & In, J. (2021). Considerations for crossover design in clinical study. Korean Journal Anesthesiology, 74(4), 293–299. https://doi.org/10.4097/kja.21165

Ma, Y., Yang, S., & Han, S. (2011). Attitudes influence implicit racial face categorization in a perceptual task. Group Processes & Intergroup Relations, 14(6), 887–899. https://doi.org/10.1177/1368430211409962

Madley-Dowd, H. R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. Journal of Clinical Epidemiology. 110, 63–73. https://doi.org/10.1016/j.jclinepi.2019.02.016

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

*Marsh, B. U. (2021). The Cost of Racial Salience on Face Memory: How the Cross-Race Effect is Moderated by Racial Ambiguity and the Race of the Perceiver and the Perceived. Journal of Applied Research in Memory and Cognition, 10(1), 13-23. https://doi.org/https://doi.org/10.1016/j.jarmac.2020.09.008

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods, 11*(4), 386–401. https://doi-org.ezproxy.uct.ac.za/10.1037/1082-989X.11.4.386

*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 (Nature Publisher Group), 9, 1-13. https://doi.org/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 Number 3021544) [Ph.D., The Florida State University]. ProQuest Dissertations & Theses A&I. Ann Arbor. https://www.proquest.com/dissertations-theses/influence-perceptual-learning-memory-faces/docview/250826761/se-2?accountid=14500 https://uct.primo.exlibrisgroup.com/discovery/openurl?institution=27UCT_INST&vid=27UCT_INST:27UCT&?genre=dissertations&issn=&title=The+influence+of+perceptual+learning+in+memory+for+faces%3A+A+process+-tracing+approach+to+investigating+the+own+-race+bias&volume=&issue=&date=2001&atitle=&spage=&sid=ProQuest+Dissertations+%26+Theses+A%26I&author=Meissner

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

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: a dual-process approach. Applied Cognitive Psychology, 19(5), 545–567. https://doi.org/10.1002/acp.1097

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own- and other-race faces. Visual Cognition, 21(9-10), 1287–1305. https://doi.org/10.1080/13506285.2013.832451

*Mousavi, S. M., & Oruc, I. (2020). Tuning of face expertise with a racially heterogeneous face-diet. Visual Cognition, 28(9), 523-539. https://doi.org/10.1080/13506285.2020.1836696

*Ng, W., & Lindsay, R. (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

Norris, R. J., Weintraub, J. N., Acker, J. R., Redlich, A. D., & Bonventre, C. L. (2020). The criminal costs of wrongful convictions. Criminology & Public Policy, 19(2), 367–388. https://doi.org/10.1111/1745-9133.12463

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

Pettigrew, T. F. (2009). Secondary Transfer Effect of Contact: Do Intergroup Contact Effects Spread to Noncontacted Outgroups? *Social Psychology (Göttingen, Germany)*, *40*(2), 55–65. https://doi.org/10.1027/1864-9335.40.2.55

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. European Journal of Social Psychology, 38(6), 922–934. https://doi.org/10.1002/ejsp.504

Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition memory: More evidence for the cross-race effect. Journal of Applied Psychology, 88(4), 760–763. https://doi.org/10.1037/0021-9010.88.4.760

Pica, E., Warren, A., Ross, D., & Kehn, A. (2015). Choosing Your Words and Pictures Wisely: When Do Individuation Instructions Reduce the Cross-Race Effect?: Individuation and the cross-race effect. Applied Cognitive Psychology, 29, 360–368. https://doi.org/10.1002/acp.3112

Reicher, S. (2007). Rethinking the Paradigm of Prejudice. South African Journal of Psychology = Suid-Afrikaanse Tydskrif Vir Sielkunde. 37(4), 820–834. https://doi.org/10.1177/008124630703700410

*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

Rohatgi, A. (2022). WebPlotDigitizer. https://apps.automeris.io/wpd/

*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. Q J Exp Psychol (Hove), 72(3), 557-569. https://doi.org/10.1177/1747021818760482

Sangrigoli, S., & De Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. Journal of Child Psychology and Psychiatry and Allied Disciplines. 45(7), 1219–1227. https://doi.org/10.1111/j.1469-7610.2004.00319.x

*Sant-Barket, S. M. (2019). Perceiver Contributors to Facial Recognition: How Might Racial (Self) Awareness Facilitate or Inhibit Cross-Racial Identification? Columbia University].

Schieferdecker, D., & Wessler, H. (2017). Bridging Segregation Via Media Exposure? Ingroup Identification, Outgroup Distance, and Low Direct Contact Reduce Outgroup

Appearance in Media Repertoires. Journal of Communication. 67(6), 993–1014. https://doi.org/10.1111/jcom.12338

Schmid, C. (2003). Immigration and Asian and Hispanic minorities in the New South: An exploration of history, attitudes and demographic trends. Sociological Spectrum, 23(2), 129–157. https://doi.org/10.1080/02732170309212

Semplonius, T., & Mondloch, C. (2013). Increasing attentional competition and uncertainty: An ecological approach to the cross-race effect. Journal of Vision (Charlottesville, Va.), 13(9), 986–986. https://doi.org/10.1167/13.9.986

Seutloali, G. M. (2015). Own-race bias in facial recognition amongst black, coloured and white participants [Honours dissertation, University of Cape Town]. UCT Psychology Repository. http://www.psychology.uct.ac.za/sites/default/files/image_tool/images/117/Maseutloali.Seutloali.pdf

Singh, B., Mellinger, C., Earls, H. A., Tran, J., Bardsley, B., & Correll, J. (2021). Does Cross-Race Contact Improve Cross-Race Face Perception? A Meta-Analysis of the Cross-Race Deficit and Contact. Personality & Social Psychology Bulletin, 14616722110244–1461672211024463. https://doi.org/10.1177/01461672211024463

Sladekova, M., Webb, L. E. A., & Field, A. P. (2023). Estimating the change in meta-analytic effect size estimates after the application of publication bias adjustment methods. Psychological Methods, 28(3), 664–686. https://doi.org/10.1037/met0000470

*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

Smith, E., & Hattery, A. J. (2011). Race, Wrongful Conviction & Exoneration. Journal of African American Studies. 15(1), 74–94. https://doi.org/10.1007/s12111-010-9130-5

Southall, R. (2023). The middle class and suburbia: desegregation towards non-racialism in South Africa? *Journal of Contemporary African Studies*, *41*(1), 60–74. https://doi.org/10.1080/02589001.2022.2083589

Sporer, S. L. (2001). The Cross Race Effect: Beyond Recognition of Faces in the Laboratory. *Psychology, Public Policy, and Law*, *7*(1), 170–200. https://doi.org/10.1037/1076-8971.7.1.170

*Sporer, S. L., Gehrke, J., & McClure, K. A. (2022). Patrolling the borders: Experience with multiple out-group members is associated with the own-ethnicity effect. Journal of

Police and Criminal Psychology, 37(2), 433-446. https://doi.org/10.1007/s11896-020-09387-x

Stanley, T. D. (2017). Limitations of PET-PEESE and Other Meta-Analysis Methods. Social Psychological & Personality Science, 8(5), 581–591. https://doi.org/10.1177/1948550617693062

*Stelter, M., Simon, D., Calanchini, J., Christ, O., & Degner, J. (2022). Real-life outgroup exposure, self-reported outgroup contact and the other-race effect [Article]. British journal of psychology, 1. https://doi.org/10.1111/bjop.12600

Sterne, J., & Egger, M. (2005). Regression Methods to Detect Publication and Other Bias in Meta-Analysis. In H. Rothstein, A. Sutton & M. Bornstein (Eds.), Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments (pp. 100-110). John Wiley & Sons.

Sterne, J., & Harbord, R. (2004). Funnel Plots in Meta-analysis. *The Stata Journal.*, *4*(2), 127–141. https://doi.org/10.1177/1536867X0400400204

Stevenson. (2006). Confronting mass imprisonment and restoring fairness to collateral review of criminal cases. Harvard Civil Rights-Civil Liberties Law Review., 41.

Struve, C. T. (2018). The federal rules of inmate appeals. Arizona State Law Journal, 50(1), 247-316.

*Susa, K. J. (2007). Examining the cross-race effect in face recognition from a temporal perspective (Publication Number 1449753) [M.A., The University of Texas at El Paso]. ProQuest Dissertations & Theses A&I. Ann Arbor. https://www.proquest.com/dissertations-theses/examining-cross-race-effect-face-recognition/docview/304761919/se-2?accountid=14500 https://uct.primo.exlibrisgroup.com/discovery/openurl?institution=27UCT_INST&vid=27UCT_INST:27UCT&?genre=dissertations&issn=&title=Examining+the+cross-race+effect+in+face+recognition+from+a+temporal+perspective&volume=&issue=&date=2007&atitle=&spage=&sid=ProQuest+Dissertations+%26+Theses+A%26I&author=Susa

*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/https://doi.org/10.1521/soco.2010.28.4.523

*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. Mem Cognit, 42(6), 863-875. https://doi.org/10.3758/s13421-014-0409-y

*Turbett, K., Jeffery, L., Bell, J., Digges, A., Zheng, Y., Hsiao, J., & Palermo, R. (2022). Serial dependence of facial identity for own- and other-race faces [Article]. Quarterly Journal of Experimental Psychology, 75(9), 1711-1726. https://doi.org/10.1177/17470218211059430

Turner, R. N., Hewstone, M., & Voci, A. (2007a). Reducing Explicit and Implicit Outgroup Prejudice Via Direct and Extended Contact: The Mediating Role of Self-Disclosure and Intergroup Anxiety. Journal of Personality and Social Psychology, 93(3), 369–388. https://doi.org/10.1037/0022-3514.93.3.369

Turner, R. N., Hewstone, M., Voci, A., Paolini, S., & Christ, O. (2007b). Reducing prejudice via direct and extended cross-group friendship. European Review of Social Psychology, 18(1), 212–255. https://doi.org/10.1080/10463280701680297

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

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. Quarterly Journal of Experimental Psychology (2006), 69(10), 1996–2019. https://doi.org/10.1080/17470218.2014.990392

Van Aert, R. C., Wicherts, J. M., & Van Assen, M. A. L. (2019). Publication bias examined in meta-analyses from psychology and medicine : A meta-meta-analysis. PloS One, 14(4), e0215052–e0215052. https://doi.org/10.1371/journal.pone.0215052

van Buuren, S & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. doi:10.18637/jss

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

Viechtbauer, W., & Cheung, M. (2010). Outlier and influence diagnostics for meta-analyses. *Research synthesis methods*, *1*(2), 112-125. Doi: 10.1002/jrsm.11

Voci, A., & Hewstone, M. (2003). Intergroup Contact and Prejudice Toward Immigrants in Italy: The Mediational Role of Anxiety and the Moderational Role of Group Salience. Group Processes & Intergroup Relations, 6(1), 37–54. https://doi.org/10.1177/1368430203006001011

Wagner, U., van Dick, R., Pettigrew, T. F., & Christ, O. (2003). Ethnic Prejudice in East and West Germany: The Explanatory Power of Intergroup Contact. Group Processes & Intergroup Relations, 6(1), 22–36. https://doi.org/10.1177/1368430203006001010

Walker, P. M., & Hewstone, M. (2006a). 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., & Hewstone, M. (2006b). A developmental investigation of other-race contact and the own-race face effect. British Journal of Developmental Psychology, 24(3), 451–463. https://doi.org/10.1348/026151005X51239

Walker, P. M., & Hewstone, M. (2008). The influence of social factors and implicit racial bias on a generalized own-race effect. Applied Cognitive Psychology, 22(4), 441–453. https://doi.org/10.1002/acp.1382

*Wan, L., Crookes, K., Dawel, A., Pidcock, M., Hall, A., & McKone, E. (2017). Face-blind for other-race faces: Individual differences in other-race recognition impairments. Journal of Experimental Psychology: General, 146(1), 102-122. https://doi.org/10.1037/xge000024910.1037/xge0000249.supp (Supplemental)

*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

Wang, J., Cao, R., Chakravarthula, P., Li, X., & Wang, S. (2024). A Critical Period for Developing Face Recognition. *Patterns*, 5(2), 100895–100895. https://doi.org/10.1016/j.patter.2023.100895

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior.*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness Evidence. Psychological Science in the Public Interest : a Journal of the American Psychological Society., 7(2), 45–75. https://doi.org/10.1111/j.1529-1006.2006.00027.x

Welnhofer, K. (n.d.). William Bernard Jackson - National registry of exonerations pre 1989. University of Michigan Law School https://www.law.umich.edu/special/exoneration/Pages/casedetailpre1989.aspx?caseid=159.

*Wheat, E. (2010). The other race effect: the role of experience and social attitudes on face recognition [Masters, University of Richmond].

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. Social Issues and Policy Review, 7(1), 83–113. https://doi.org/10.1111/j.1751-2409.2012.01044.x

Witter, T., & Tredoux, C. (2020). The own-group bias in face processing: the effect of training on recognition performance [Doctoral Dissertation, University of Cape Town]. UCT Repository. http://hdl.handle.net/11427/33070

*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. 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

*Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. Applied Cognitive Psychology, 17(3), 365-373. https://doi.org/10.1002/acp.898

*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 and law, 21(5), 490-506. https://doi.org/https://doi.org/10.1080/1068316X.2014.989173

Xiao, W. S., Quinn, P. C., Pascalis, O., & Lee, K. (2014). Own- and other-race face scanning in infants: Implications for perceptual narrowing. Developmental Psychobiology, 56(2), 262–273. https://doi.org/10.1002/dev.21196

Xu, K., Nosek, B., & Greenwald, A. G. (2014a). Psychology data from the Race Implicit Association Test on the Project Implicit Demo website. Journal of Open Psychology Data, 2(1), e3. http://doi.org/10.5334/jopd.ac

Xu, K., Nosek, B., & Greenwald, A. G. (2014b). Race IAT 2002-2013 [dataset] *Open Science Framework*, Available at: https://osf.io/52qxl/

*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.ezproxy.uct.ac.za/10.1177/1088868311418987

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

*Zhao, M., Hayward, W. G., & Bülthoff, I. (2014a). Face format at encoding affects the other-race effect in face memory. Journal of Vision, 14(9). http://ezproxy.uct.ac.za/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-36187-001&site=ehost-live

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

Zhou, G., Pu, X., Young, S. G., & Tse, C.S. (2014). Effects of divided attention and social categorization on the own-race bias in face recognition. *Visual Cognition*, *22*(9–10), 1296–1310. https://doi.org/10.1080/13506285.2014.998324

Zhou, X., Elshiekh, A., & Moulson, M. C. (2019). Lifetime perceptual experience shapes face memory for own- and other-race faces. Visual Cognition, 27(9-10), 687–700. https://doi.org/10.1080/13506285.2019.1638478

APPENDICES

Appendix A

Ethical Approval

UNIVERSITY OF CAPE TOWN

Department of Psychology

University of Cape Town Rondebosch 7701 South Africa
Telephone (021) 650 3417
Fax No. (021) 650 4104

13 December 2021

Jade MacDonnell
Department of Psychology
University of Cape Town
Rondebosch 7701

Dear Jade

I am pleased to inform you that ethical clearance has been given by an Ethics Review Committee of the Faculty of Humanities for your study, *Does prejudice moderate the relationship between contact and the own group bias? A meta-analysis of the own group bias*. The reference number is PSY2021-056.

I wish you all the best for your study.

Yours sincerely

Lauren Wild (PhD)
Associate Professor
Chair: Ethics Review Committee

Appendix B

Search Strings used in Article Acquisition

Database	Search String 1	Boolean Operator	Search String 2	Boolean Operator	Search String 3	Parameters
ProQuest Dissertations and Theses	("cross race effect" OR "cross race" OR "own group" OR "own group bias" OR "other race bias" OR "own race bias" OR "other race effect" OR "other race")	AND	("identification" OR "recognition" OR "recognition memory" OR "identification performance" OR "face recognition" OR "memory" OR "recognition task")	AND	("interracial contact" OR "contact" OR "other race contact" OR "out group contact" OR "self-reported contact" OR "self report contact" OR "interracial experience" OR "other group contact" OR "own group contact" OR "out group contact" OR "contact measure")	Abstract and summary text search 01-01-1969 to 31-10-2022 English
Google Scholar	("cross race effect" AND "identification task" AND "contact" AND "prejudice")	-	-	-	-	Incognito Mode to ensure reproducibility 1969-2022
PsycArticles	("cross race effect" OR "cross race" OR "own group" OR "own group bias" OR "other race bias" OR "own race bias" OR "other race effect" OR "other race")	AND	("identification" OR "recognition" OR "recognition memory" OR "identification performance" OR "face recognition" OR "memory" OR "recognition task")	AND	("interracial contact" OR "contact" OR "other race contact" OR "out group contact" OR "self-reported contact" OR "self report contact" OR "interracial experience" OR "other group contact" OR "own group contact" OR "out group contact" OR "contact measure")	Abstract text search 01-01-1969 to 31-10-2022 English Human population group
PsycINFO	("cross race effect" OR "cross race" OR "own group" OR "own group bias" OR "other race bias" OR "own race bias" OR "other race effect" OR "other race")	AND	("identification" OR "recognition" OR "recognition memory" OR "identification performance" OR "face recognition" OR	AND	("interracial contact" OR "contact" OR "other race contact" OR "out group contact" OR "self-reported contact" OR "self report contact" OR "interracial experience" OR "other group contact" OR "own group contact" OR "out group contact" OR "contact measure")	Abstract text search 01-01-1969 to 31-10-2022 English Human population group

Database	Search String 1	Boolean Operator	Search String 2	Boolean Operator	Search String 3	Parameters
			"memory" OR "recognition task")			
Academic Search Premier	("cross race effect" OR "cross race" OR "own group" OR "own group bias" OR "other race bias" OR "own race bias" OR "other race effect" OR "other race")	AND	("identification" OR "recognition" OR "recognition memory" OR "identification performance" OR "face recognition" OR "memory" OR "recognition task")	AND	("interracial contact" OR "contact" OR "other race contact" OR "out group contact" OR "self-reported contact" OR "self report contact" OR "interracial experience" OR "other group contact" OR "own group contact" OR "out group contact" OR "contact measure")	Abstract text search 01-01-1969 to 31-10-2022 English Human population group
PubMed	("cross race effect" OR "cross race" OR "own group" OR "own group bias" OR "other race bias" OR "own race bias" OR "other race effect" OR "other race")	AND	("identification" OR "recognition" OR "recognition memory" OR "identification performance" OR "face recognition" OR "memory" OR "recognition task")	AND	("interracial contact" OR "contact" OR "other race contact" OR "out group contact" OR "self-reported contact" OR "self report contact" OR "interracial experience" OR "other group contact" OR "own group contact" OR "out group contact" OR "contact measure")	Abstract and title text search English Human population group 1969-2022
Scopus	("cross race effect" OR "cross race" OR "own group" OR "own group bias" OR "other race bias" OR "own race bias" OR "other race effect" OR "other race")	AND	("identification" OR "recognition" OR "recognition memory" OR "identification performance" OR "face recognition" OR "memory" OR "recognition task")	AND	("interracial contact" OR "contact" OR "other race contact" OR "out group contact" OR "self-reported contact" OR "self report contact" OR "interracial experience" OR "other group contact" OR "own group contact" OR "out group contact" OR "contact measure")	Abstract text search English 1969-2022
Web of Science	("cross race effect" OR "cross race" OR "own group" OR "own group bias" OR "other race	AND	("identification" OR "recognition" OR "recognition memory" OR "identification	AND	("interracial contact" OR "contact" OR "other race contact" OR "out group contact" OR "self-reported contact" OR "self report contact" OR "interracial	01-01-1969 to 31-10-2022

Database	Search String 1	Boolean Operator	Search String 2	Boolean Operator	Search String 3	Parameters
	bias" OR "own race bias" OR "other race effect" OR "other race")		performance" OR "face recognition" OR "memory" OR "recognition task")		experience" OR "other group contact" OR "own group contact" OR "out group contact" OR "contact measure")	
Science Direct	“cross race effect”	AND	(“identification” OR “recognition”)	AND	“contact”	Subject Areas: Psychology, Neuroscience, Social Science, Arts & Humanities

Appendix C

Articles Excluded from Stream One

The following are excluded articles from stream one alongside a reason for exclusion. Stream one refers to the traditional method of sourcing articles i.e. all articles were found using a search of specific key words and were checked for relevancy.

ID	Original Number	Author	Main Reason for Exclusion	Specifics if needed -
1	2	Anderson, 1999	Insufficient data to calculate contact-OENE effect[52]	Uni-ethnic and multi-ethnic homes could be a proxy for high and low contact groups but standard error was not reported (graphically or otherwise). No effects or suitably useful statistics were reported.
2	4	Brunet,2022	Published outside of the specified date parameters.	
3	5	Busche, 2012	Reporting quality	Observed scores discussed where not specified. No mention of recognition metrics.
4	8	Chiroro, 1994	Data is a duplicate of already captured information	Duplicate of 1995 article
5	13	Correll, 2021	Insufficient data to calculate contact-OENE effect	
6	14	Crookes, 2013	Insufficient data to calculate contact-OENE effect	
7	16	Da Silva, 2008	Insufficient data to calculate contact-OENE effect	

[52]This reason, broadly speaking, is used when contact has been either measured or manipulated; a contact-OENE effect was not reported; and, there is insufficient data to calculate this effect. Insufficient data would also include reporting unstandardized beta coefficients only, in a model with multiple predictors.

ID	Original Number	Author	Main Reason for Exclusion	Specifics if needed -
8	18	Davis, 2016	Did not measure identification or recognition data	Analyzed response times
9	20	Flores, 2015	Contact was not measured or manipulated	
10	21	Furl, 2022	Contact was not measured or manipulated	Participant contact was not measured. Compared computational algorithms for different levels of contact
11	22	Gier, 2019	Contact was not measured or manipulated	Looked at own age experience only
12	23	Goodman, 2007	Insufficient data to calculate contact-OENE effect	
13	24	Guarino, 2019	Did not measure identification or recognition data	Analyzed familiarity ratings
14	25	Gwinn, 2014	Duplicate information	Duplicate (PHD) of article in the pool. Both were excluded
15	26	Gwinn, 2015	Did not measure identification or recognition data	Analyzed a constructed confidence / familiarity rating(from yes-no and confidence)/
16	29	Hourihan, 2012	Contact was not measured or manipulated	
17	30	Hu, 2014	Contact was not measured or manipulated	
18	31	Hugenberg, 2007	Contact was not measured or manipulated	
19	32	Hutchinson, 2013	Contact was not measured or manipulated	
20	33	Jackiw, 2007	Data is a duplicate of already captured information	Duplicate of 2008 article
21	36	Knuycky, 2009	Contact was not measured or manipulated	
22	37	Knuycky, 2014	Contact was not measured or manipulated	
23	39	Lawrie, 2010	Insufficient data to calculate contact-OENE effect	
24	40	Leffers, 2021	Insufficient data to calculate contact-OENE effect	

ID	Original Number	Author	Main Reason for Exclusion	Specifics if needed -
25	41	Ma, 2011	Contact was not measured or manipulated	
26	42	MacLin, 2003	Experimental data was not collected or analyzed	A review or general article
27	43	Marcon, 2007	Duplicate information	Duplicate (thesis) of article in the pool. Both were excluded
28	44	Marcon, 2010	Contact was not measured or manipulated	
29	46	Marsh, 2021b	Insufficient data to calculate contact-OENE effect	
30	47	McDonnell, 2014	Contact was not measured or manipulated	
31	48	McKone, 2007	Insufficient data to calculate contact-OENE effect	Pertains to experiment three
32	50	Megias, 2018	Insufficient data to calculate contact-OENE effect	Analyzed proportion correct scores and Reaction Times
33	52	Mendes, 2020	Study is not reported in English	General review only which also would be problematic
34	55	Pica, 2015	Insufficient data to calculate contact-OENE effect	
35	56	Proverbio, 2019	Contact was not measured or manipulated	Reaction times were analyzed
36	57	Ray, 2010	Contact was not measured or manipulated	
37	58	Reggev, 2020	Insufficient data to calculate contact-OENE effect	
38	60	Rodriguez, 2008	Contact was not measured or manipulated	
39	61	Rose, 2013	Contact was not measured or manipulated	
40	64	Shriver, 2010	Contact was not measured or manipulated	
41	65	Silvas, 2018	Insufficient data to calculate contact-OENE effect	
42	66	Smith, 2004	Contact was not measured or manipulated	

ID	Original Number	Author	Main Reason for Exclusion	Specifics if needed -
43	71	Tuettenberg, 2019	Only quality of contact was measured[53]	
44	74	Van Bavel, 2012	Contact was not measured or manipulated	Social exclusion manipulation but without a needed "high" contact group – conditions were alone vs. excluded
45	75	Vidal, 2018	Between participants manipulation of target ethnicity	
46	76	Walker, 2006	Insufficient data to calculate contact-OENE effect	
47	77	Walker, 2006b	Insufficient data to calculate contact-OENE effect	Only quality aka individuation contact was reported
48	80	Wiese, 2018	Insufficient data to calculate contact-OENE effect	
49	81	Wilson, 2010	Contact was not measured or manipulated	

[53] Only studies that measured or manipulated contact and from which a contact-OENE effect could be reported or calculated where used. Quality of contact where reported, was reported from studies that met this criterion.

Appendix D

Summary of Reasons for Exclusion from Stream One

Reason	Count
Between participants manipulation of target ethnicity	1
Contact was not measured or manipulated	20
Data is a duplicate of already captured information	2
Did not measure identification or recognition data	3
Duplicate information	2
Experimental data was not collected or analyzed	1
Insufficient data to calculate contact-OENE effect	16
Only quality of contact was measured	1
Published outside of the specified date parameters	1
Study is not reported in English	1
Reporting Quality	1

Appendix E

Articles Excluded from Stream Two

The following are excluded articles from stream two alongside a reason for exclusion. Stream two refers articles that were sourced via an exploration of reference lists.

ID	Original Number	Author	Main Reason for Exclusion	Specifics if needed -
1	1	Baldwin, 2012	Contact was not measured or manipulated	Motivation manipulation
2	3	Sangrigoli, 2004	Contact was not measured or manipulated	
3	5	DeGutis, 2013	Insufficient data to calculate contact-OENE effect	
4	8	Cross, 1971	Insufficient data to calculate contact-OENE effect	
5	9	Malpass & Kravitz, 1969	Insufficient data to calculate contact-OENE effect	
6	11	Wiese, 2012	Insufficient data to calculate contact-OENE effect	
7	13	Sporer, 2010	Insufficient data to calculate contact-OENE effect	Unpublished manuscript

Appendix F

Summary of Reasons for Exclusion from Stream Two

Reason	Count
Contact was not measured or manipulated	2
Insufficient data to calculate contact-OENE effect	5

Appendix G

Interpretation of Outgroup Implicit Prejudice Scores (IAT)

Range		Interpretation in terms of preference
0.00	±0.14	Neutral, no or little
±0.15	±0.34	Slight
±0.35	±63	Moderate
±64		Strong

Note. Adapted from 'Implications of the Implicit Association Test D-Transformation for Psychological Assessment', Blanton et al. (2015). Sage Journals[54].

[54] https://doi-org.ezproxy.uct.ac.za/10.1177/1073191114551382

Appendix H

Risk of Bias Assessment

						Domain One (max 1)	Domain Two (max 1)	Domain Three (max 2)	Domain Four (max 2)	Domain Five (max 1)	Total Risk of Bias (max 7)	
Article ID	Citation	Study Number	Published	Type	Category	In - and Outgroup Targets Tested	Full Cross-over design used	Randomized Order	Attrition and Exclusion	Novel Targets at Testing	Score	Classification
1	Chiroro & Valentine, 1995	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
2	Adams et al., 2010	1	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
3	Brigham & Barkowitz, 1978	1	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
4	Carroo, 1986	1	Yes	Journal Article	Recognition	1	1	1	0	1	4	Medium Risk of Bias
5	Cavazos, 2020	2	No	Dissertation	In-View Matching	0	0	0	0	0	0	Low Risk of Bias
6	Chiroro et al., 2008	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
7	Chorney, 2015	1	Yes	Conference Paper	Recognition	0	1	0	0	1	2	Low Risk of Bias
8	Corenblum & Meissner, 2006	2	Yes	Journal Article	Recognition	0	1	0	0	0	1	Low Risk of Bias
9	Cruz et al., 2022	1A	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
9	Cruz et al., 2022	1B	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
10	Finklea, 2008	2B	No	Dissertation	Recognition	0	0	0	0	1	1	Low Risk of Bias

11	Hancock & Rhodes, 2008	1	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
12	Havard et al., 2017	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
13	Jackiw et al., 2008	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
14	Jerovich, 2017	1	No	Dissertation	Recognition	0	0	0	1	1	2	Low Risk of Bias
15	Kovalenko & Surudzhii, 2014	1	Yes	Conference Paper	Recognition	0	1	0	0	1	2	Low Risk of Bias
16	Marsh, 2021	1	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
17	McKone et al., 2019	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
18	Meissner, 2001	1	No	Dissertation	Recognition	0	0	0	0	0	0	Low Risk of Bias
19	Mousavi & Oruc, 2020	1	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
20	Ng & Lindsay, 1994	1	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
21	Rhodes et al., 2009	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
22	Sadozai et al., 2019	1	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
22	Sadozai et al., 2019	2	Yes	Journal Article	Recognition	0	1	0	0	0	1	Low Risk of Bias
23	Sant-Barket, 2019	1	No	Dissertation	Recognition	0	1	0	0	0	1	Low Risk of Bias
24	Sporer et al., 2022	1	Yes	Journal Article	Recognition	0	1	0	0	0	1	Low Risk of Bias
25	Stelter et al., 2022	1a	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
25	Stelter et al., 2022	1b	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias

25	Stelter et al., 2022	1c	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
25	Stelter et al., 2022	2a	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
25	Stelter et al., 2022	2b	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
25	Stelter et al., 2022	3	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
26	Susa et al., 2010	1	No	Dissertation	Recognition	0	1	0	0	0	1	Low Risk of Bias
27	Susa, 2007	1	No	Dissertation	Recognition	0	1	0	0	0	1	Low Risk of Bias
28	Tullis et al., 2014	3	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
29	Turbett et al., 2022	1	Yes	Journal Article	Recognition	0	1	0	0	0	1	Low Risk of Bias
30	Wan et al., 2015	5	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
31	Wan et al., 2017	2	Yes	Journal Article	Recognition	0	1	0	0	0	1	Low Risk of Bias
32	Wong et al., 2020	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
33	Wright et al., 2003	1	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
34	Wylie et al., 2015	1	Yes	Journal Article	Recognition	0	1	0	0	0	1	Low Risk of Bias
35	Zhao et al., 2014	1-3	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias
36	Zhao et al., 2014b	1	Yes	Journal Article	Recognition	0	0	0	0	0	0	Low Risk of Bias
36	Zhao et al., 2014b	2	Yes	Journal Article	Recognition	0	1	0	0	0	1	Low Risk of Bias
37	Lavrakas et al., 1976	1	Yes	Journal Article	Recognition	0	0	0	0	1	1	Low Risk of Bias

38	Young & Hugenberg , 2012	1	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
38	Young & Hugenberg , 2012	2	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
39	Wheat, 2010	1	No	Dissertation	Recognition	0	1	0	0	0	1	Low Risk of Bias
39	Wheat, 2010	2	No	Dissertation	Recognition	0	1	0	0	0	1	Low Risk of Bias
39	Wheat, 2010	3	No	Dissertation	Recognition	0	1	0	0	0	1	Low Risk of Bias
40	Slone et al., 2000	1	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
41	Carroo, 1987	1	Yes	Journal Article	Recognition	0	1	0	0	1	2	Low Risk of Bias
42	Dai, 2021	1	No	Dissertation	Recognition	0	0	0	0	0	0	Low Risk of Bias

Note. Cavazos (2020) was excluded from the final sample for being an in-view matching task and therefore not testing memory for target faces.

Appendix I

List of Studies for which Standard Deviations were Imputed for OENE

This table includes articles for which standard deviations could not be retrieved. A complete set of data was needed to calculate the OENE. As such data was analyzed (a) without incomplete data – via listwise deletion and (ii) was analyzed after standard deviations were imputed – which allowed the calculation of an effect.

Hits - SDs	FA's - SDs	d' – SDs	RB - SDs
Ng & Lindsey, 1994	Ng & Lindsey, 1994	Ng & Lindsey, 1994	Ng & Lindsey, 1994
Slone et al., 2000	Slone et al., 2000		
		Brigham & Barkowitz, 1978	
		Finklea, 2008	
		Hancock & Rhodes, 2008	
		Wylie et al., 2015	Wylie et al., 2015
		Young & Hugenberg, 2012	
			Chiroro & Valentine, 1995
			Marsh, 2021
			Tullis et al., 2014

Note. Hits = correct identifications. FA's = false alarms or incorrect foil/filler identifications. d'=discriminability or overall identification accuracy. This metric considers both hits and false alarms. RB = response bias or how willing a participant was in terms of making an identification decision.

Appendix J

Standard Deviation Imputation for Identification Data

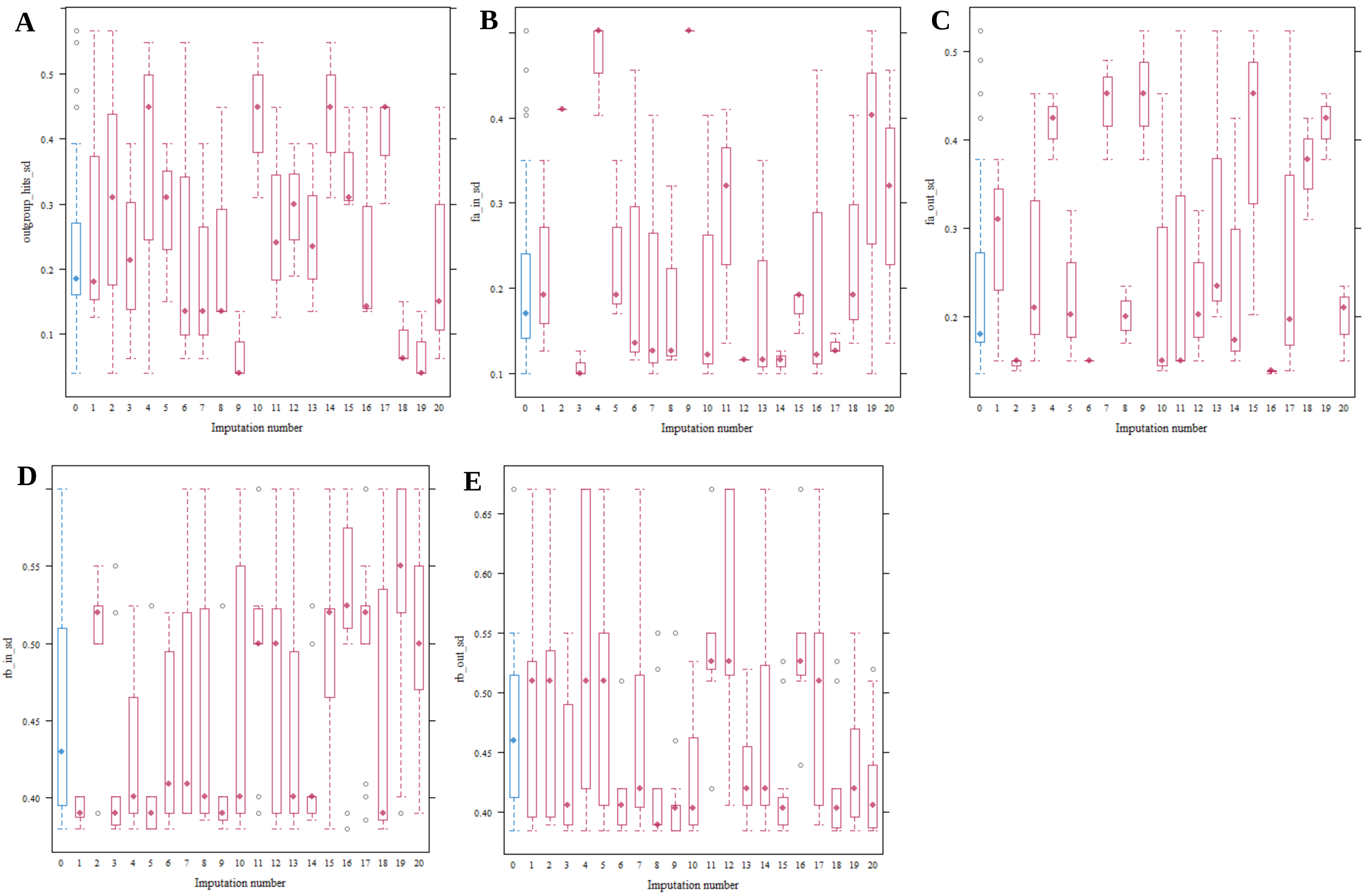
Each identification metric, namely hits, false-alarms, discrimination and response bias, was filtered to retain only complete cases. The percentage of missing data was assessed - results reported in the main body of the report. A subset of numeric only predictors was used for imputation. These predictors included all the identification metrics listed above alongside their associated effects (standardized mean difference score). This was necessary as imputation considers the relationships between variables when determining the imputed value. Predictive mean matching was used for all imputations and all imputations used 20 iterations. This means 20 potential values were imputed for each missing cell, the results of which were pooled to determine the actual or final imputed value. In this way the results are typically deemed reasonable approximations of the actual missing value, if this value could be known. After each imputation the results were inspected for plausibility, or reasonableness, as an additional check.

For strip, bw and density plots, blue is used to indicate the original data. Red is used to indicate the imputed values or the spread of the data which includes imputed values. For the strip plots, ideally the position of the imputed or red values should match or overlap with the observed (blue) data. With regards to bw plots, the distribution of the data is depicted via box and whisker plots to facilitate an easier comparison of the original versus distributed data. For density plots, imputed data should approximate the shape of the observed data. For convergence plots, the lines or iterations should converge and/or follow a similar pattern.

For all diagnostic checks, the imputed data was deemed to be a reasonable approximation and thus valid for use.

Figure J1

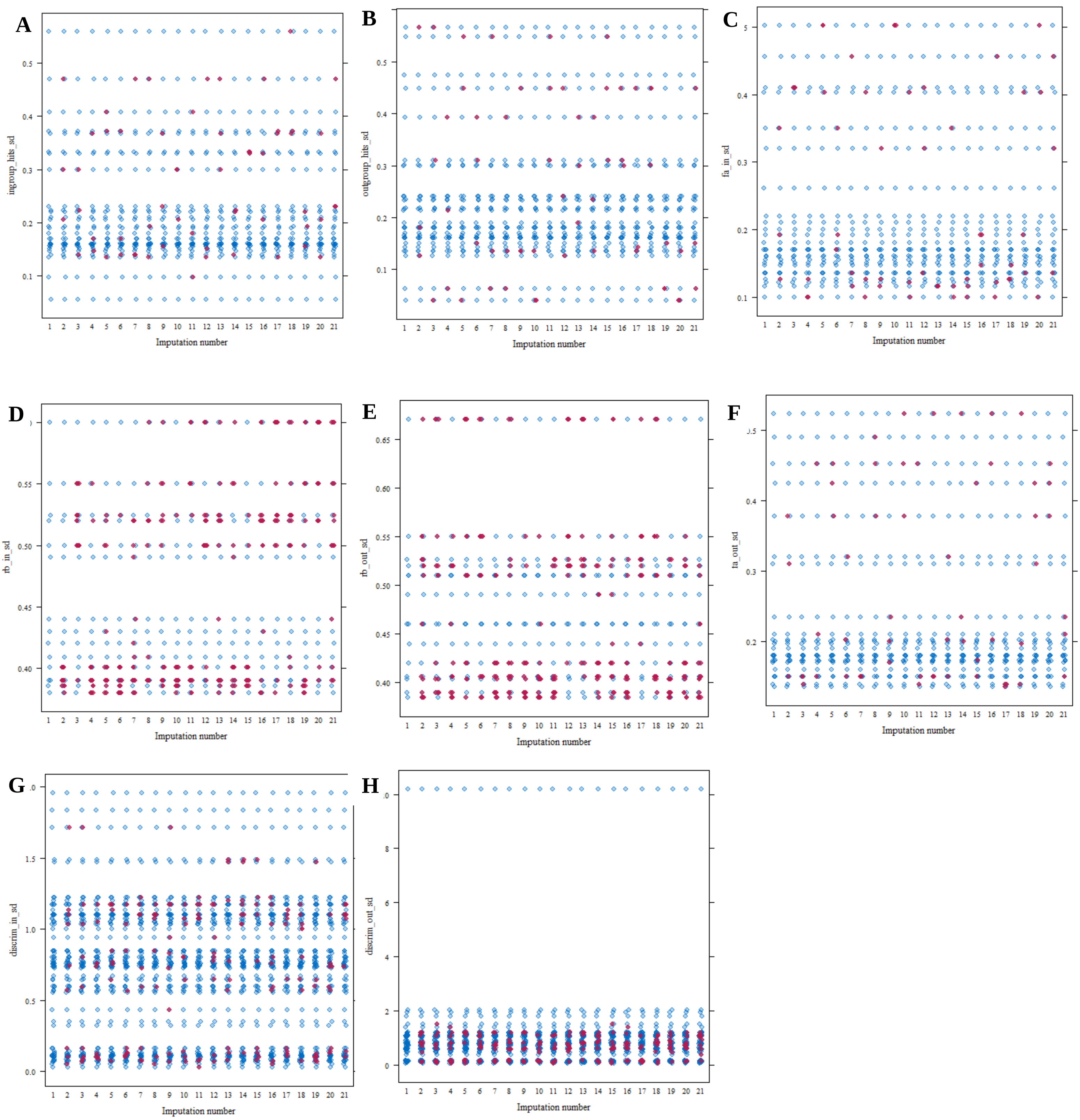
Imputation Check for Imputed Identification Data Standard Deviations (BW Plot)



Note. (A) Plot for Outgroup Hits – Standard deviations. (B) Plot for ingroup false alarms – standard deviations. (C) Plot for outgroup false alarms – standard deviations. (D) Plot for ingroup response bias – standard deviations. (E) Plot for outgroup response bias – standard deviations.

Figure J2

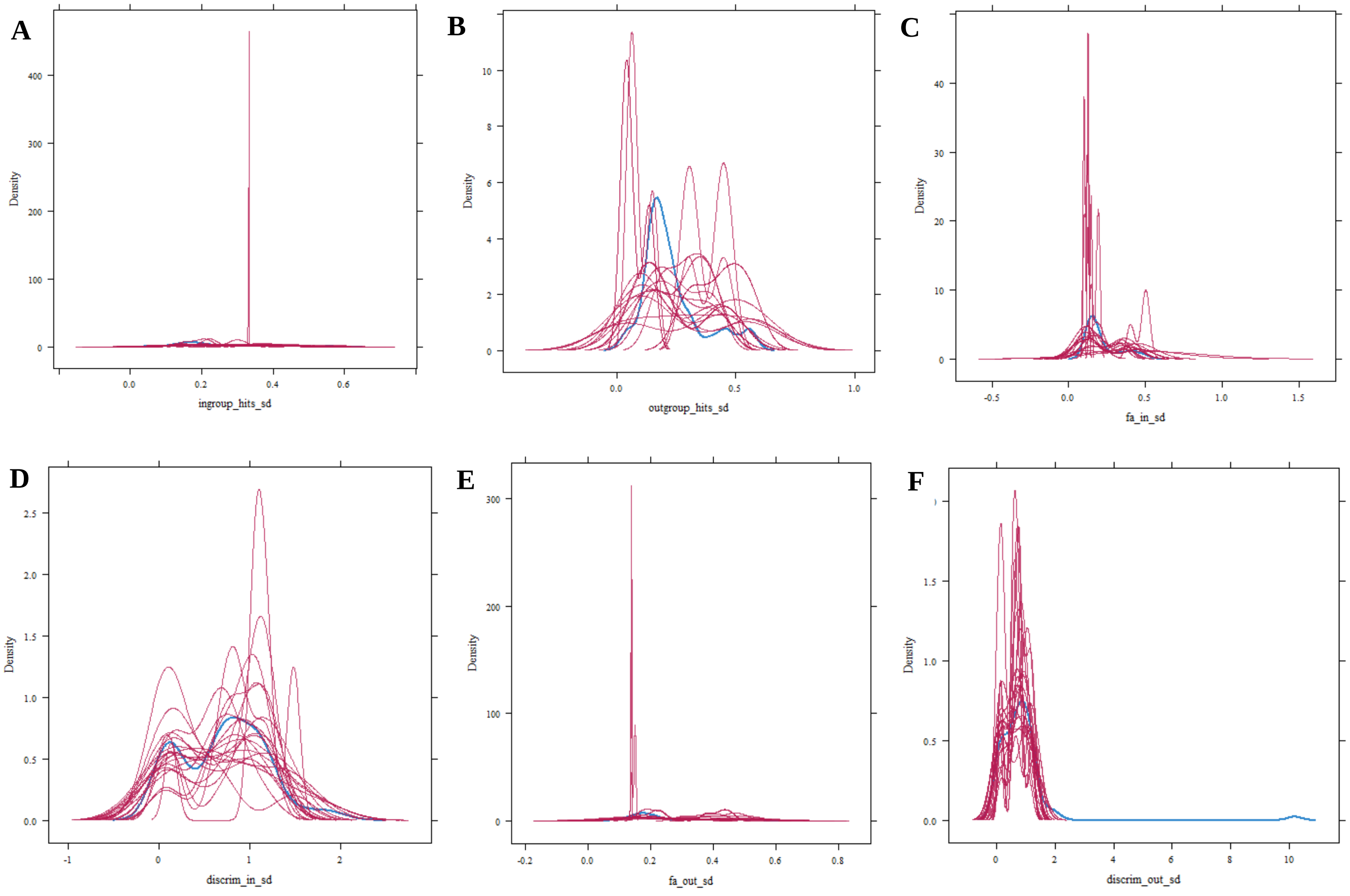
Imputation Check for Imputed Identification Data Standard Deviations (Strip Plots)

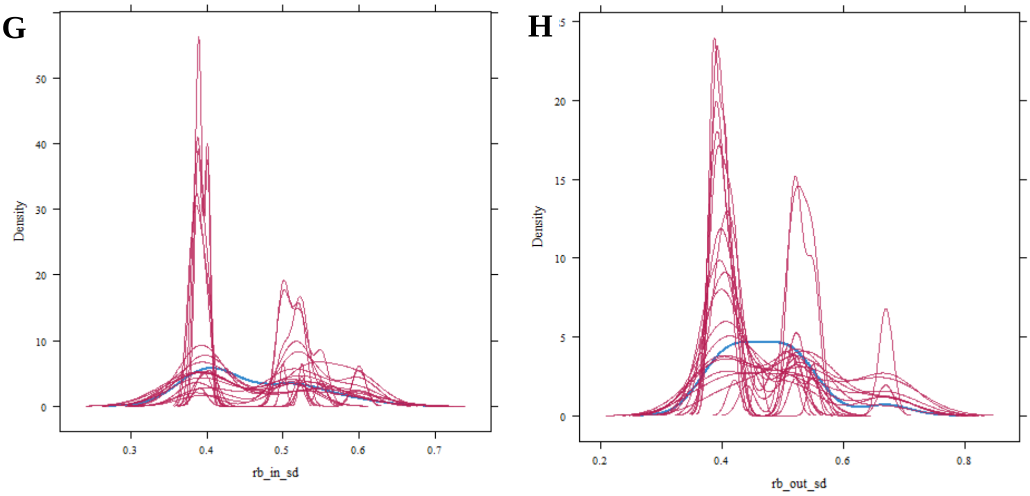


*Note***.** (A) Plot for ingroup hits – standard deviations. (B) Plot for outgroup hits – standard deviations. (C). Plot for ingroup false alarms – standard deviations. (D)Plot for ingroup response bias – standard deviations. (E) Plot for outgroup response bias – standard deviations. (F) Plot for outgroup false alarms - standard deviations. (G) Plot for ingroup discriminability - standard deviations. (H) Plot for outgroup discriminability - standard deviations.

Figure J3

Imputation Check for Imputed Identification Data Standard Deviations (Density Plots)

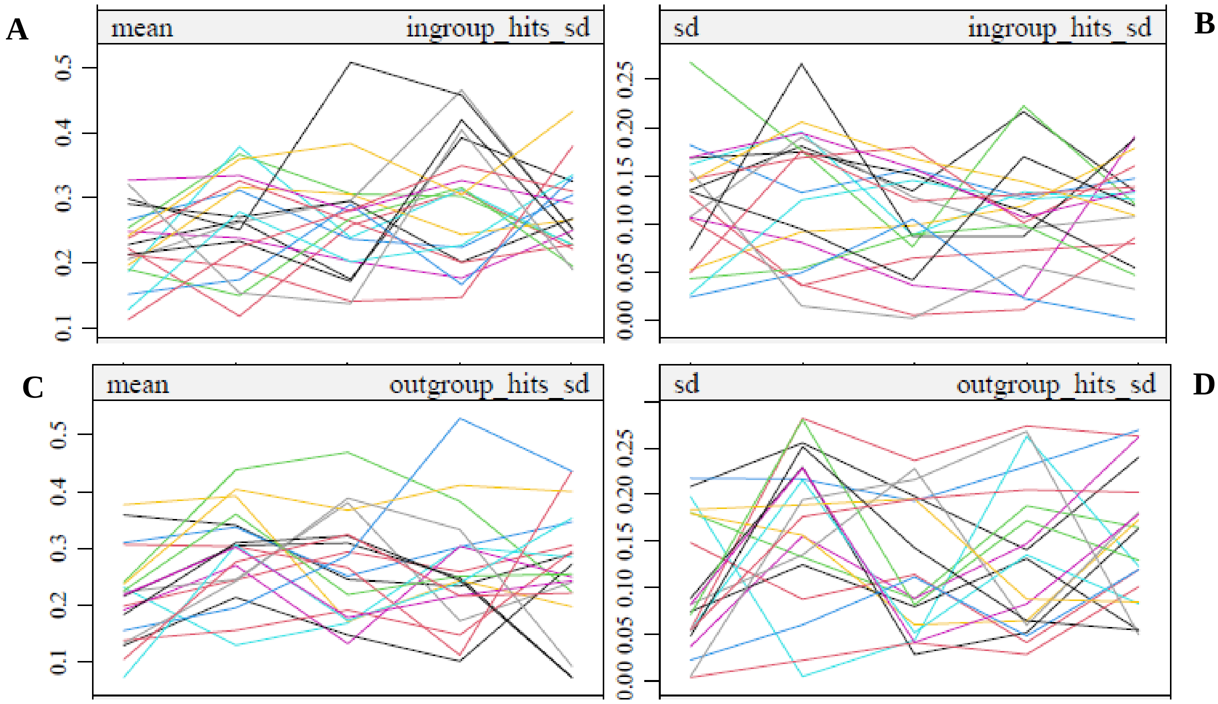


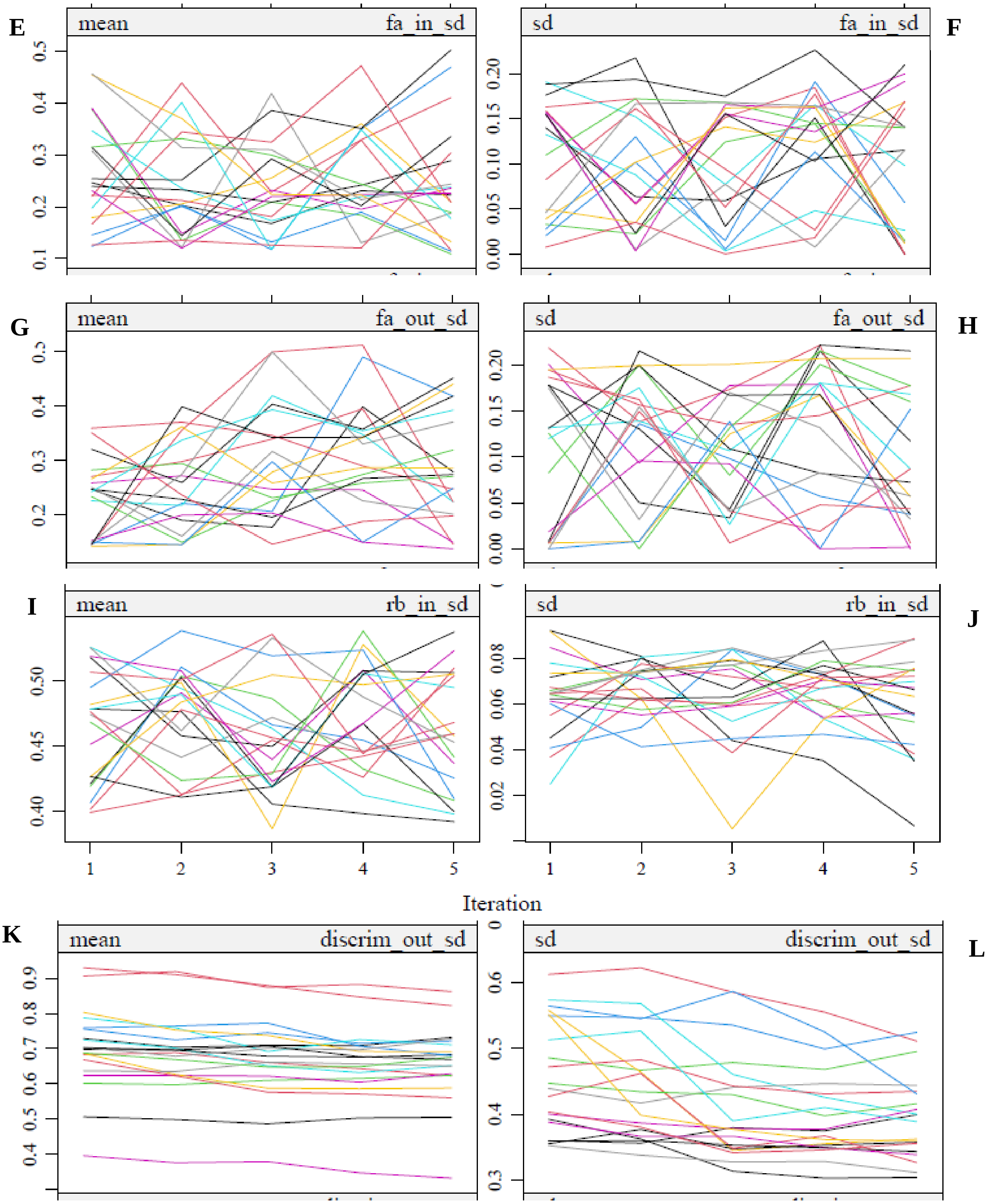


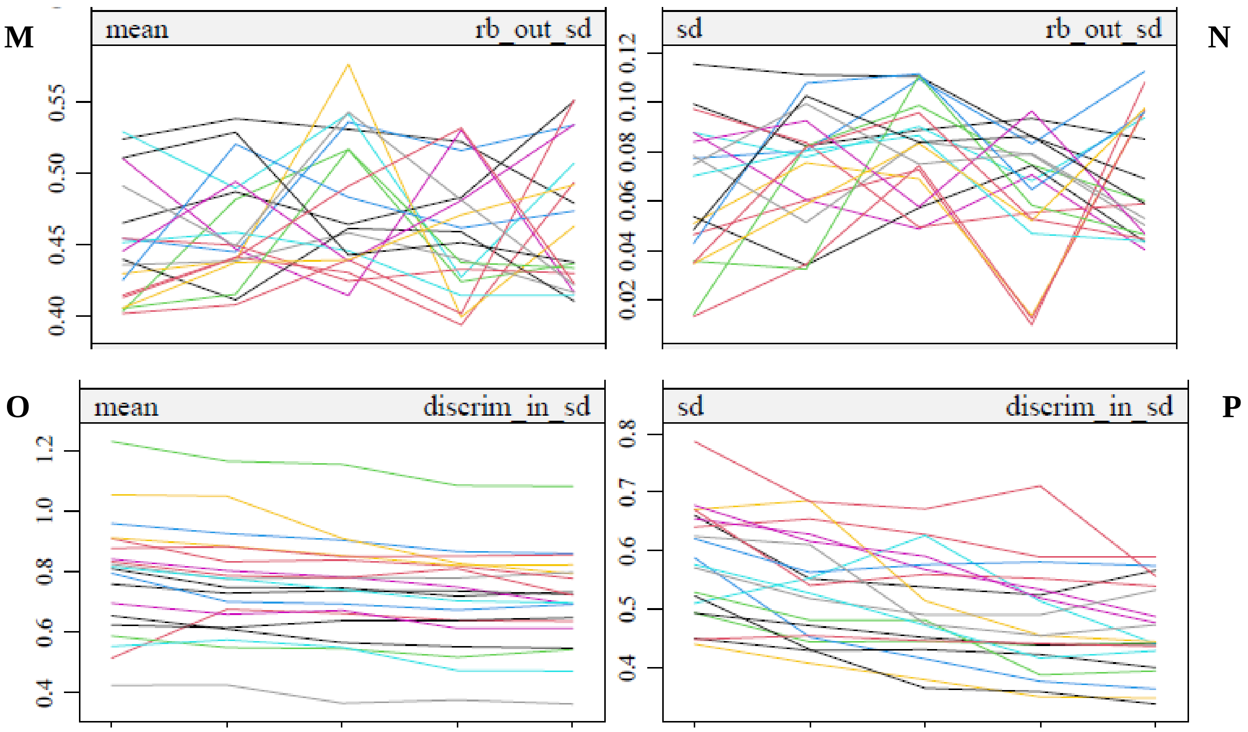
Note. (A) Plot for ingroup hits – standard deviations. (B) Plot for outgroup hits - standard deviations. (C) Plot for ingroup false alarms - standard deviations. (D) Plot for ingroup discriminability – standard deviations. (E) Plot for outgroup false alarms - standard deviations. (F) Plot for outgroup discriminability - standard deviations. (G) Plot for ingroup response bias - standard deviations. (H) Plot for outgroup response bias – standard deviations.

Figure J4

Imputation Check for Imputed Identification Data Standard Deviations (Convergence Plots)




E
mean
fa_in_sd
0.1
0.2
0.3
0.4
0.5
F
sd
fa_in_sd
0.00
0.05
0.10
0.15
0.20
G
mean
fa_out_sd
0.2
0.3
0.4
H
sd
fa_out_sd
0.00
0.05
0.10
0.15
0.20
I
mean
rb_in_sd
0.40
0.45
0.50
J
sd
rb_in_sd
0.02
0.04
0.06
0.08
1
2
3
4
5
1
2
3
4
5
Iteration
K
mean
discrim_out_sd
0.4
0.5
0.6
0.7
0.8
0.9
L
sd
discrim_out_sd
0.3
0.4
0.5
0.6



*Note***.** Plots for imputed standard deviations – imputed values only. (A) Plot for ingroup hits. (B) Plot of variability of imputed values for ingroup hits. (C) Plot for outgroup hits. (D) Plot of variability of imputed values for outgroup hits. (E) Plot for ingroup false alarms. (F) Plot of variability of imputed values for ingroup false alarms. (G) Plot for outgroup false alarms. (H) Plot of variability of imputed values for outgroup false alarms. (I) Plot for ingroup response bias. (J) Plot of variability of imputed values for ingroup response bias. (K) Plot for outgroup discriminability (d'). (L) Plot of variability of imputed values for outgroup discriminability (d'). (M) Plot for outgroup response bias. (N) Plot of variability of imputed values for outgroup response bias. (O) Plot for ingroup discriminability (d'). (P) Plot of variability of imputed values for ingroup discriminability (d').

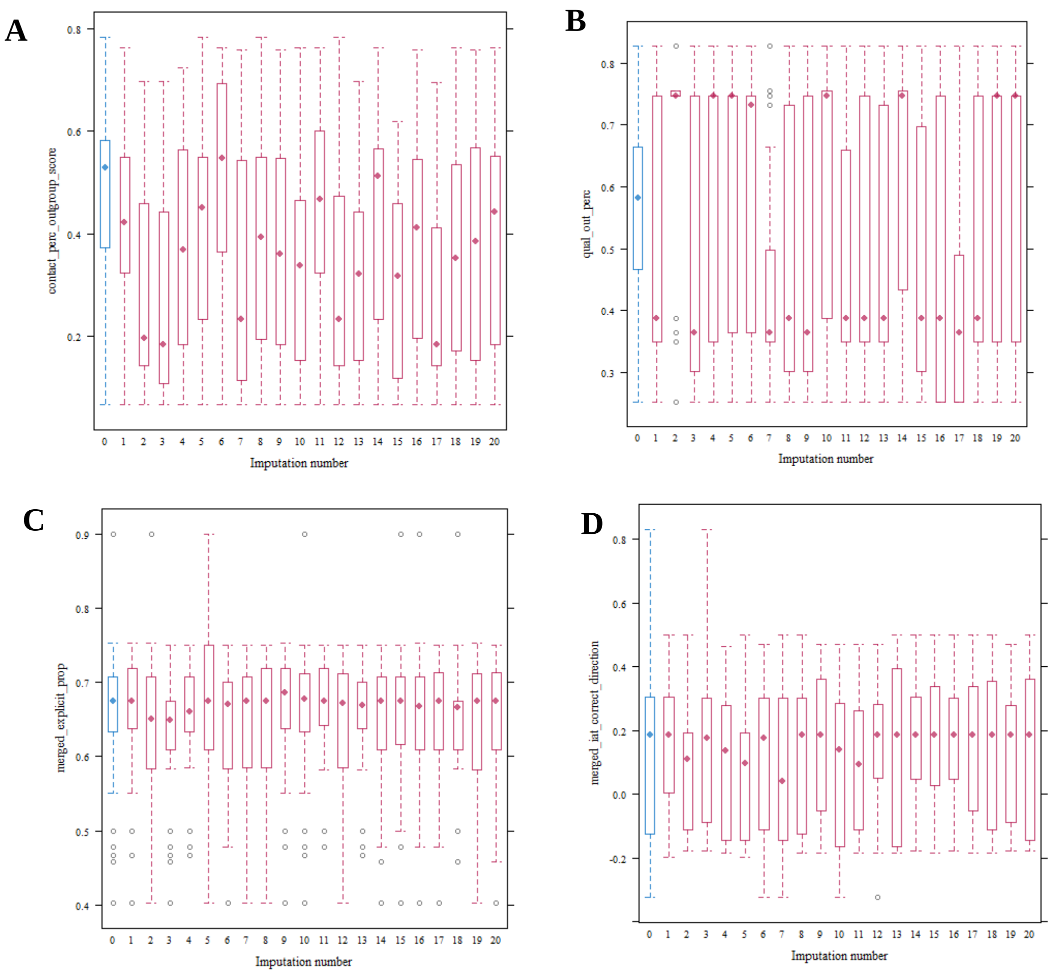
Appendix K

Imputation for Outgroup Contact and Prejudice

Incomplete outgroup contact and prejudice data was completed via imputation. All imputations were checked for reasonableness. All imputations were deemed reasonable approximations of the original data.

Figure K1

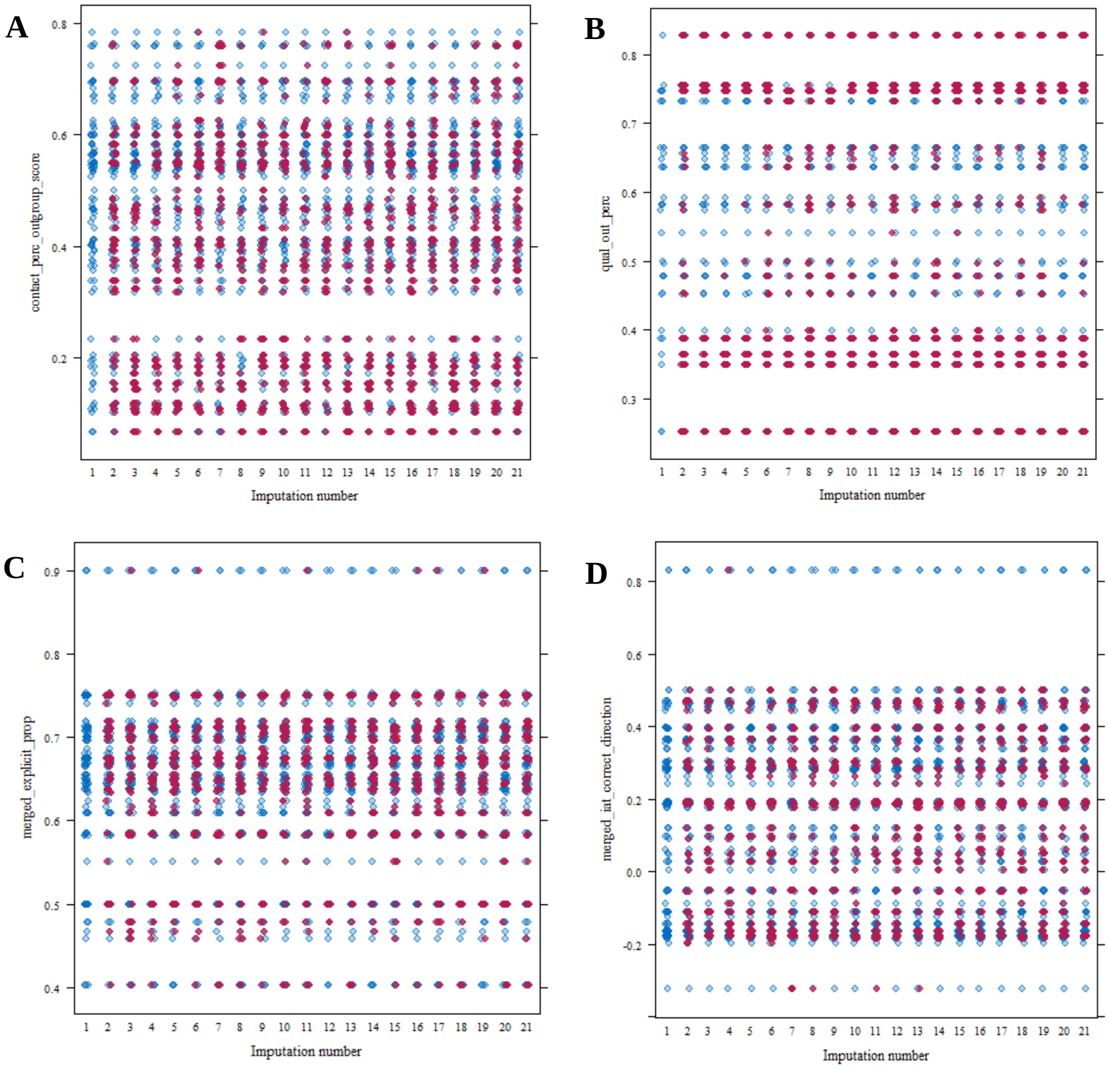
Imputation Check for Imputed Outgroup Contact and Prejudice Data (BW Plots)



Note. (A) Plot for quantity of outgroup contact (% Score). (B) Plot for quality of outgroup contact (% Score). (C) Plot for explicit outgroup prejudice (proportion). (D) Plot for implicit outgroup prejudice.

Figure K2

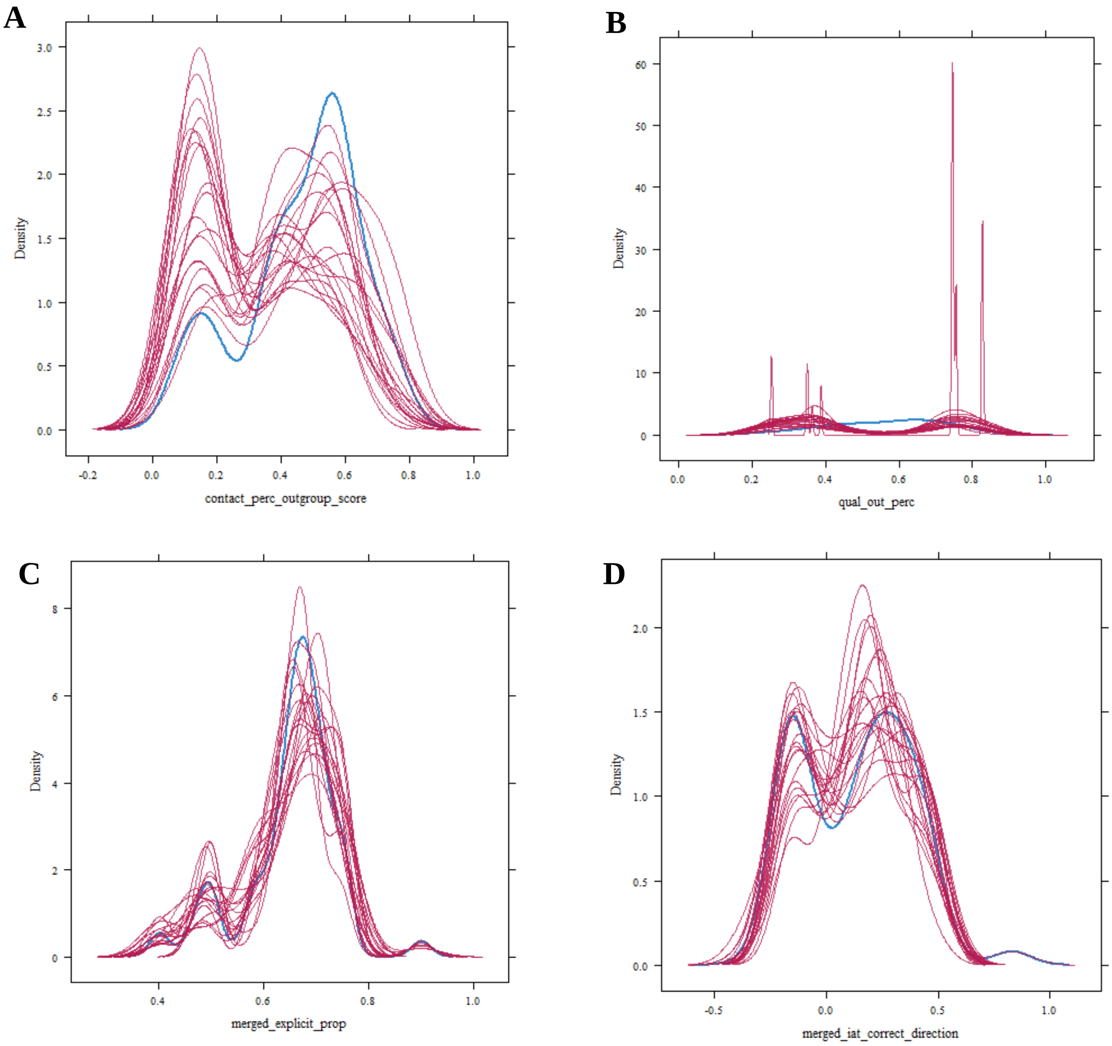
Imputation Check for Imputed Outgroup Contact and Prejudice Data (Strip Plots)



Note. (A) Plot for quantity of outgroup contact (% Score). (B) Plot for quality of outgroup contact (% Score). (C) Plot for explicit outgroup prejudice (proportion). (D) Plot for implicit outgroup prejudice.

Figure K3

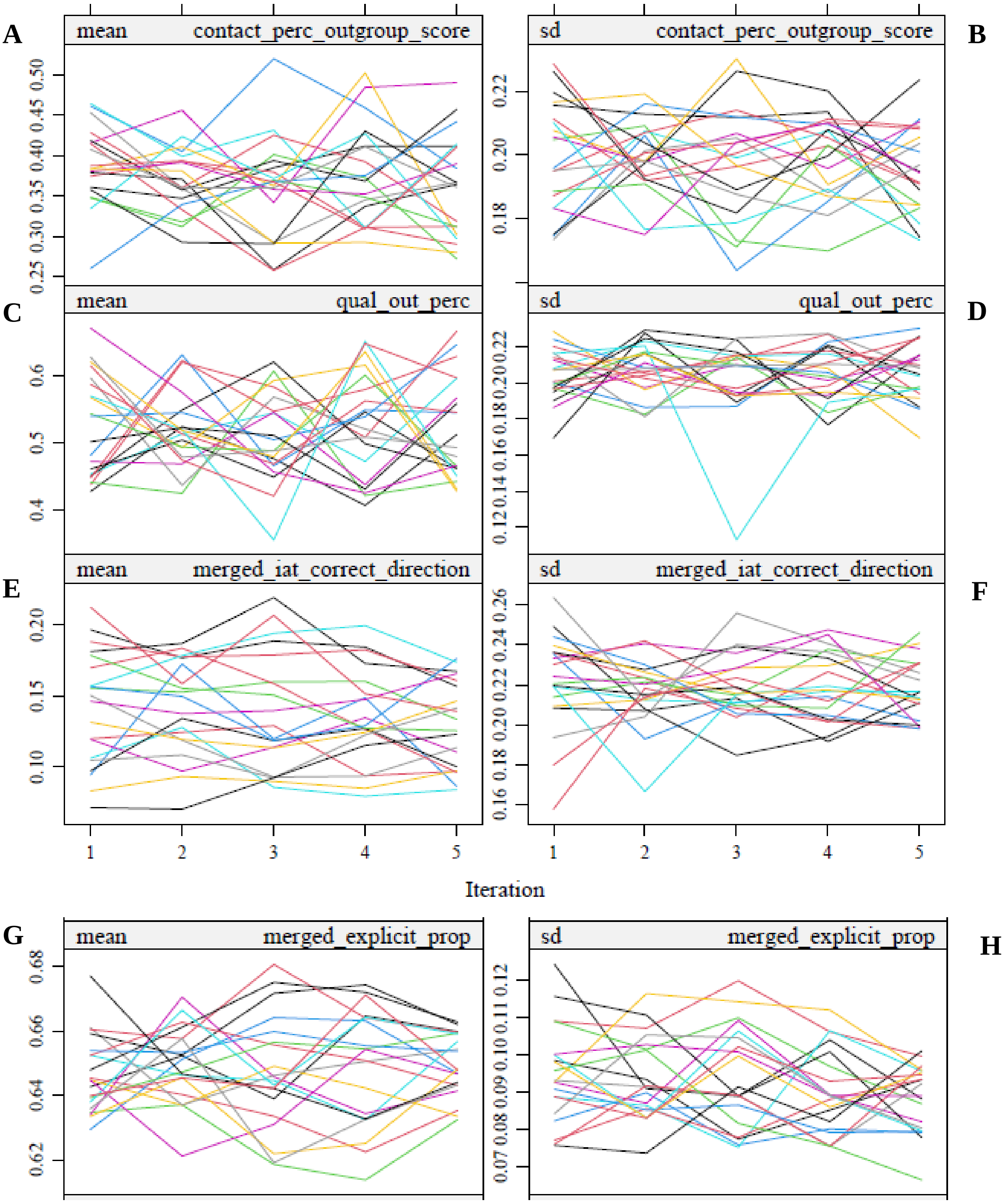
Imputation Check for Imputed Outgroup Contact and Prejudice Data (Density Plots)



Note. (A) Plot for quantity of outgroup contact (% Score). (B) Plot for quality of outgroup contact (% Score). (C) Plot for explicit outgroup prejudice (proportion). (D) Plot for implicit outgroup prejudice.

Figure K4

Imputation Check for Imputed Outgroup Contact and Prejudice Data (Convergence Plots)



*Note***.** Plots for imputed values only. (A) Plot for quantity of outgroup contact. (B) Plot of variability of imputed values for quantity of outgroup contact. (C) Plot for quality of outgroup contact. (D) Plot of variability of imputed values for quality of outgroup contact. (E) Plot for implicit outgroup prejudice. (F) Plot of variability of imputed values for implicit outgroup prejudice. (G) Plot for explicit outgroup prejudice. (H) Plot of variability of imputed values for explicit outgroup prejudice.

Appendix L

Validation Checks for Harvard's Project Implicit Prejudice Data

An independent sample t-test was used to test the validity, or rather the feasibility, of using Harvard's '*Project Implicit*' prejudice data. Articles which recorded and reported outgroup prejudice scores were used in this check. The actual outgroup prejudice scores were compared to matched to sample '*Project Implicit'* scores i.e. matched on location, year, participant age and participant to target ethnic-nationality pairings.

Implicit Prejudice

Using the Hancock and Rhodes (2008) or Blanton et al. (2015) convention, the average reported implicit outgroup prejudice value was classified as being neutral while the average '*Project Implicit*' implicit outgroup value was classified as having slight prejudice (M = 0.02, M =0.17 respectively, see Appendix G for convention). Implicit outgroup prejudice values were not significantly different from one another (t (13.04) =-2.09, p = 0.056)[55]. The merged data was therefore validated as an acceptable proxy for reported data in instances where no such data was available.

Explicit Prejudice

Average explicit outgroup prejudice values, with higher values indicating less outgroup prejudice or lower explicit outgroup prejudice, were high for both reported and '*Project Implicit*' values (M= 0.70, M = 0.57 respectively). Indicating that on average, the samples had more favourable outgroup views.

Similar to implicit outgroup prejudice, a trend was noted wherein average reported values demonstrated greater outgroup prejudice when compared with '*Project Implicit*' data (M= 0.70, M = 0.57 respectively, '*Project Implicit*' data was not exclusively conducted in an experimental setting)[56].

Average explicit outgroup prejudice values were significantly different from one another (t (34) =2.809, p=0.008). While not fully validated as an acceptable proxy for reported data, all merged explicit outgroup values were analyzed with caution.

---

[55] The implicit prejudice t-test accounted for heterogeneity of variance.

[56] Project Implicit data is captured online. It could be argued that without experimenter and co-participant observations i.e. lower social desirability bias, anonymity and a familiar setting yielded values which are a more realistic of innate prejudice.

Appendix M

Match Specificity for Harvard's 'Project Implicit' Explicit Outgroup Prejudice Data

Index	ID	Citation	Level of Geographic Specificity	Direct Year Match	Difference in Year Match	Use of a geographic proxy	Age filtering not exact?	Reported Outgroup Prejudice	Merged Outgroup Prejudice
1	106	Stelter et al., 2022	Country	Yes	-	-	-	0.56	0.71
2	107	Stelter et al., 2022	Country	Yes	-	-	-	0.56	0.71
3	108	Stelter et al., 2022	Country	No	6	-	-	0.45	0.58
4	109	Stelter et al., 2022	Country	No	6	-	-	0.45	0.58
5	110	Stelter et al., 2022	Country	No	6	-	-	0.49	0.59
6	111	Stelter et al., 2022	Country	No	6	-	-	0.49	0.59
7	152	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.83	0.65
8	153	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.83	0.65
9	154	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.83	0.65
10	155	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.87	0.50
11	156	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.87	0.50
12	157	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.87	0.50
13	158	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.71	0.64
14	159	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.71	0.64

Index	ID	Citation	Level of Geographic Specificity	Direct Year Match	Difference in Year Match	Use of a geographic proxy	Age filtering not exact?	Reported Outgroup Prejudice	Merged Outgroup Prejudice
15	160	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.71	0.64
16	164	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.79	0.40
17	165	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.79	0.40
18	166	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.79	0.40

Note. ‘Difference in Year Match’ would only be filled in if there was not an exact year match. For example, Index 3 to 6. ‘Use of geographic proxy; would only be filled in when the participant’s country was not present in the ‘*Project Implicit*’ dataset. As noted above, no geographic proxies were used. ‘Age filtering not exact?’ would only be filled in when outgroup prejudice data was not an exact match for the sample (i.e. within 5 year window above or below the average sample age).

Appendix N

Match Specificity for Harvard's 'Project Implicit' Implicit Outgroup Prejudice Data

Index	ID	Citation	Level of Geographic Specificity	Direct Year Match	Difference in Year Match	Use of a geographic proxy	Age filtering not exact?	Reported Outgroup Prejudice	Merged Outgroup Prejudice
1	26	Jerovich, 2017	Country	No	1	-	-	0.01	-0.20
2	27	Jerovich, 2017	Country	No	1	-	-	0.03	0.18
3	152	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.03	0.47
4	153	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.03	0.47
5	154	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.03	0.47
6	155	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.02	0.46
7	156	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.02	0.46
8	157	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.02	0.46
9	158	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.02	0.03
10	159	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.02	0.03
11	160	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.02	0.03
12	164	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.00	-0.14
13	165	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.00	-0.14
14	166	Wheat, 2010	Metropolitan Statistical Area	Yes	-	-	-	0.00	-0.14

Appendix O

Frequency of Task Types in the Sample

The analysis made use of identification tasks only i.e. tasks that had a delay between encoding and testing. Frequency of task types sampled is depicted in the table below

Task	Percentage (%)
Old-New*	68.00
CFMT	14.67
Delayed Matching	6.67
Lineup	6.67
AFC*	4.00

*Note***. ***Alternative force choice tasks are a specific type of Old-New task. Instead of viewing a single face and being asked to make a decision, participants could be shown two faces – side by side and are instructed to pick which face was previously shown (example of a 2 AFC task)

Appendix P

Frequency (%) of Type of Encoding Instructions Used

Encoding Instructions	Percentage (%)
Standard	84.75
Directed	6.78
Incidental	5.08
Mixed	3.39

Note. Standard instructions = participants were informed that they would be tested on the study faces at a later date. Directed = participants were instructed to study the target faces in a specific way i.e. pay close attention to features. Incidental = participants were not expecting an identification task and did not actively study the target faces. Mixed = collapsed condition. Data could be a mix of standard and directed encoding for example.

Appendix Q

Frequency of Outgroup Contact Assessment

Methods used to assess outgroup contact

	Percentage **(%)**
Self-report	85.33
PV as proxy	5.33
Outgroup population metrics	5.33
Manipulation	4.00

Note. ‘PV as proxy’ = participant variable as proxy. For example, country in which the participant resides being used to infer degree of outgroup contact or length of time living in a particular area i.e. labels such as high / low outgroup contact could be inferred. ‘Outgroup population metrics’ = using census data to infer contact i.e. using proportion of outgroup members in a city as an estimation of outgroup contact. Manipulation = manipulating the degree of outgroup contact for example outgroup training or exposure.

Appendix R

Descriptive Statistics for Outgroup Prejudice

	Implicit Prejudice			Explicit Prejudice	
	Mean		SD	Mean	SD
Merged	0.16	slight	0.26	0.64	0.09
Merged (Completed)	0.14	neutral	0.24	0.65	0.09
Combined Option One	0.15	slight	0.25	0.65	0.10
Combined Option Two*	0.14	neutral	0.24	0.66	0.09

Note. Merged = ‘Project Implicit’ data. Merged (completed) = ‘Project Implicit’ data that was completed via multiple imputation. Combined Option One = actual data was used when available, if not, ‘Project Implicit’ data was used as a substitute. Combined Option Two = missing values were completed via multiple imputation.

Appendix S

Descriptive Statistics for Outgroup Prejudice Across Time

Decade	Data	Average Prejudice	Global Average Prejudice	Largest Mean Difference
1970s	explicit merged	0.38	0.38	0.00
1970s	explicit merged completed	0.38		
1970s	explicit prejudice opt1	0.38		
1970s	explicit prejudice opt2	0.38		
1970s	implicit merged	0.17	0.17	0.00
1970s	implicit merged completed	0.17		
1970s	implicit prejudice opt1	0.17		
1970s	implicit prejudice opt2	0.17		
1980s	explicit merged	0.36	0.36	0.00
1980s	explicit merged completed	0.36		
1980s	explicit prejudice opt1	0.36		
1980s	explicit prejudice opt2	0.36		
1980s	implicit merged	-0.05	-0.05	0.00
1980s	implicit merged completed	-0.05		
1980s	implicit prejudice opt1	-0.05		
1980s	implicit prejudice opt2	-0.05		
1990s	explicit merged	0.26	0.26	0.00
1990s	explicit merged completed	0.26		
1990s	explicit prejudice opt1	0.26		
1990s	explicit prejudice opt2	0.26		
1990s	implicit merged	0.14	0.13	0.03
1990s	implicit merged completed	0.11		
1990s	implicit prejudice opt1	0.14		
1990s	implicit prejudice opt2	0.11		
2000s	explicit merged	0.36	0.35	0.01
2000s	explicit merged completed	0.35		

Decade	Data	Average Prejudice	Global Average Prejudice	Largest Mean Difference
2000s	explicit prejudice opt1	0.36		
2000s	explicit prejudice opt2	0.35		
2000s	implicit merged	0.00		
2000s	implicit merged completed	0.02		
2000s	implicit prejudice opt1	0.00	0.01	0.03
2000s	implicit prejudice opt2	0.02		
2010s	explicit merged	0.35		
2010s	explicit merged completed	0.35		
2010s	explicit prejudice opt1	0.31	0.33	0.05
2010s	explicit prejudice opt2	0.31		
2010s	implicit merged	0.15		
2010s	implicit merged completed	0.16		
2010s	implicit prejudice opt1	0.12	0.14	0.04
2010s	implicit prejudice opt2	0.12		
2020s	explicit merged	0.35		
2020s	explicit merged completed	0.34		
2020s	explicit prejudice opt1	0.37	0.35	0.02
2020s	explicit prejudice opt2	0.35		
2020s	implicit merged	0.17		
2020s	implicit merged completed	0.17		
2020s	implicit prejudice opt1	0.17	0.17	0.00
2020s	implicit prejudice opt2	0.17		

Appendix T

Sample Bias – Frequency (%) of Countries Sampled

	Country	n	Percentage (%)
1	United States	33	44.59
2	Australia	13	17.57
3	Germany	6	8.11
4	Canada	5	6.76
5	Malaysia	4	5.41
6	South Africa	4	5.41
7	China & Germany	2	2.7
8	Ukraine	2	2.7
9	China	1	1.35
10	Portugal	1	1.35
11	Scotland & England	1	1.35
12	Zimbabwe	1	1.35
13	Zimbabwe & England	1	1.35

Note. Participant data is not always reported in relation to one specific place. As a result combination lines such as ‘Scotland & England’ represent instances whereby the reported data was collapsed across sampling sites.

Appendix U

Frequency (%) of Most Used Target Ethnic Nationality Pairings

	Targets Tested	n	Percentage (%)
1	White - Black	32	35.16
2	White - Asian	22	24.18
3	Same - Other	10	10.99
4	Asian - Indian	4	4.4
5	White - Other	3	3.3
6	Asian - Asian	2	2.2
7	Hispanic - Black	2	2.2
8	Slavic - Asian	2	2.2
9	Slavic - Black	2	2.2
10	Slavic - Indian	2	2.2
11	Slavic - MENA	2	2.2
12	White - Hispanic	2	2.2
13	White - Indian	2	2.2
14	White - MENA	2	2.2
15	Asian - Black	1	1.1
16	White - IPOC	1	1.1

Note. Identification tasks typically use two ethnic nationalities when testing identification performance i.e. testing participants on ‘White and Black faces’ or ‘White and Asian faces’. When (i) more than two ethnic nationalities are used for target faces and/or (ii) more than one participant ethnic nationality is used, results are often reported in a collapsed manner under the label ‘same-other’. When only one participant ethnic nationality is used i.e. white and participants are tested on more than two target face ethnic nationalities, the label ‘white-other’ is used.

Appendix V

OENE Validation - Hits

Descriptive Statistics

Both datasets, i.e. original and imputed, had a higher incidence of hits, or correct identifications, for ingroup members exceeding that of hits for outgroup members (62.50% and 62.86% of cases respectively). Descriptively supporting the presence of an OENE for hits. This was further affirmed via an exploration of central tendencies (Table V1). The overall average hits, or rather accuracy of identifications, for in-group members is higher than that of outgroup members. However, the average mean difference was relatively small (0.03)

Table V1

Descriptives Table for Correct Identifications

		Ingroup (prop)	Ingroup SD	Outgroup (prop)	Outgroup SD	Mean difference
Original All	Ingroup H > Outgroup H	0.73	0.10	0.64	0.14	0.09
	Ingroup H < Outgroup H	0.63	0.13	0.68	0.09	-0.06
	Overall	0.69	0.12	0.66	0.12	0.04
Imputed All	Ingroup H > Outgroup H	0.73	0.10	0.64	0.13	0.09
	Ingroup H < Outgroup H	0.63	0.12	0.69	0.09	-0.06
	Overall	0.69	0.12	0.66	0.12	0.03

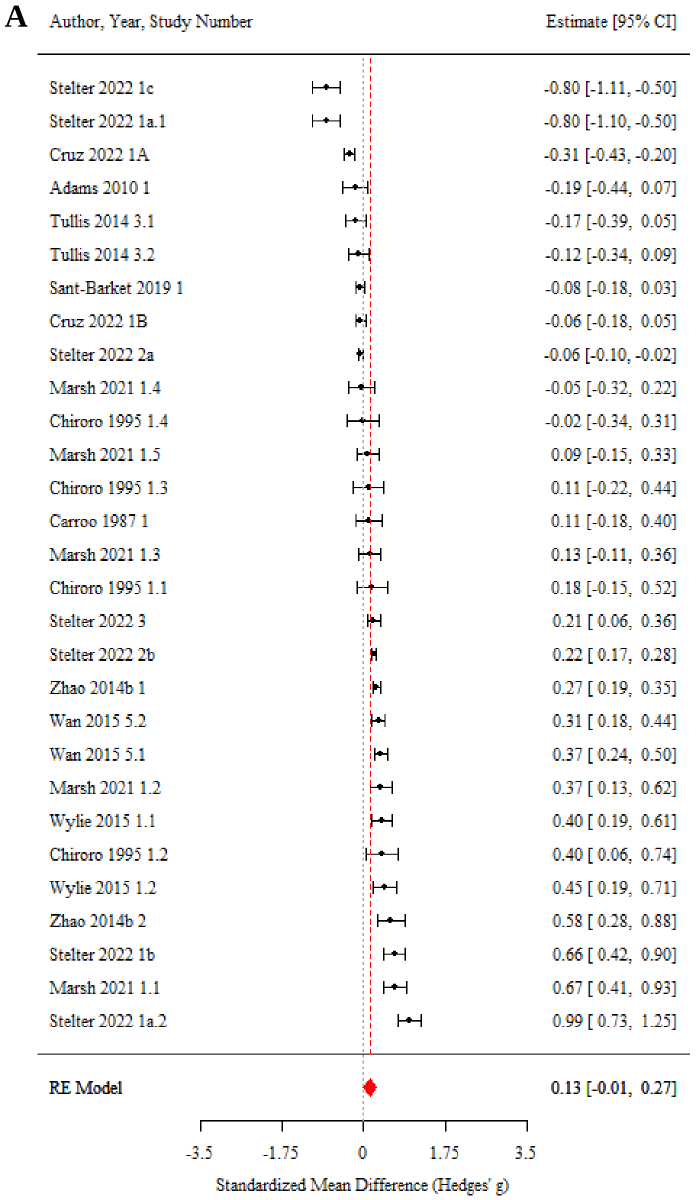
Note. ‘Original-all’ = Actual data that had both the necessary means and standard deviations to calculate an effect. ‘Imputed-all’ = Data that was completed via imputation i.e. missing standard deviations were imputed and used to calculate an effect.

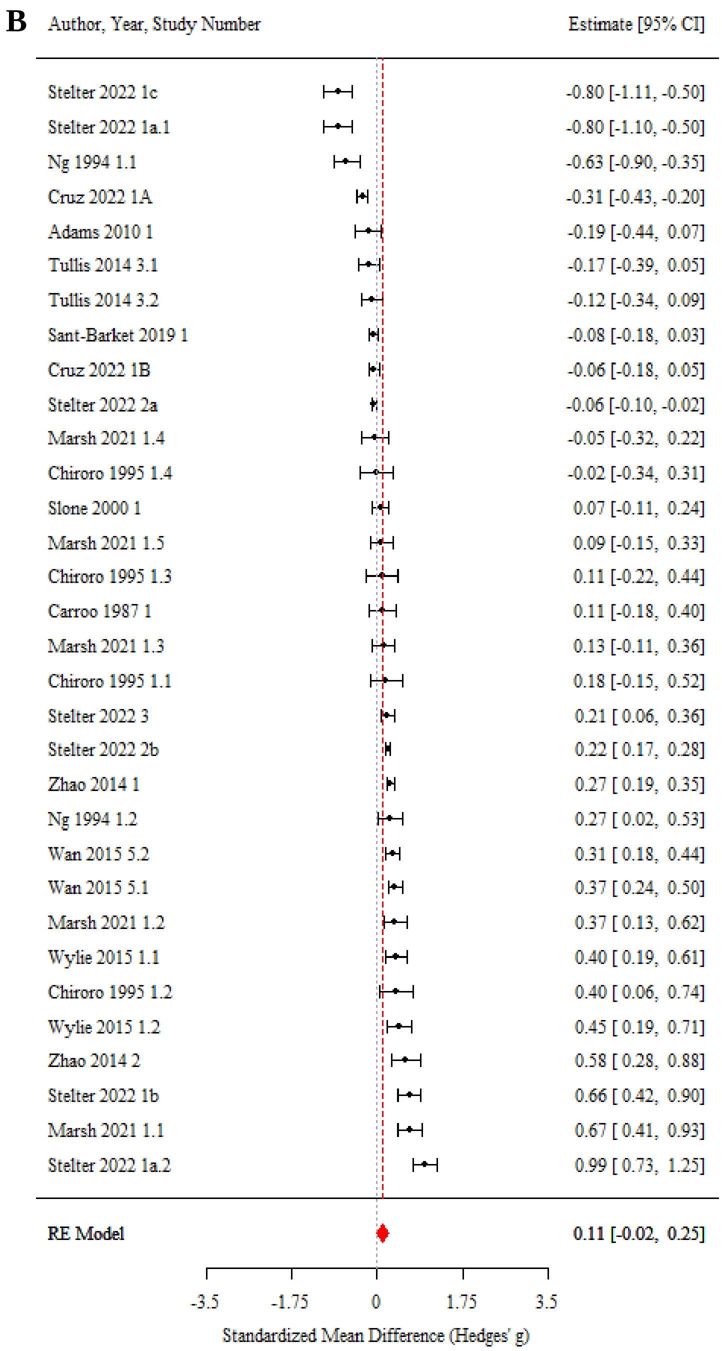
Meta-Analysis

The aggregate effect (*SMD*), for all datasets, confirm the presence of an OENE for hits. Thus supporting hypothesis 1a. The strength of the effect sizes were however, negligible and none were statistically significant (p >.05). No publication bias was evident in the sample.

Figure V1

Forest Plots for Correct Identifications (Hits) after Influential Cases and Outliers were Removed

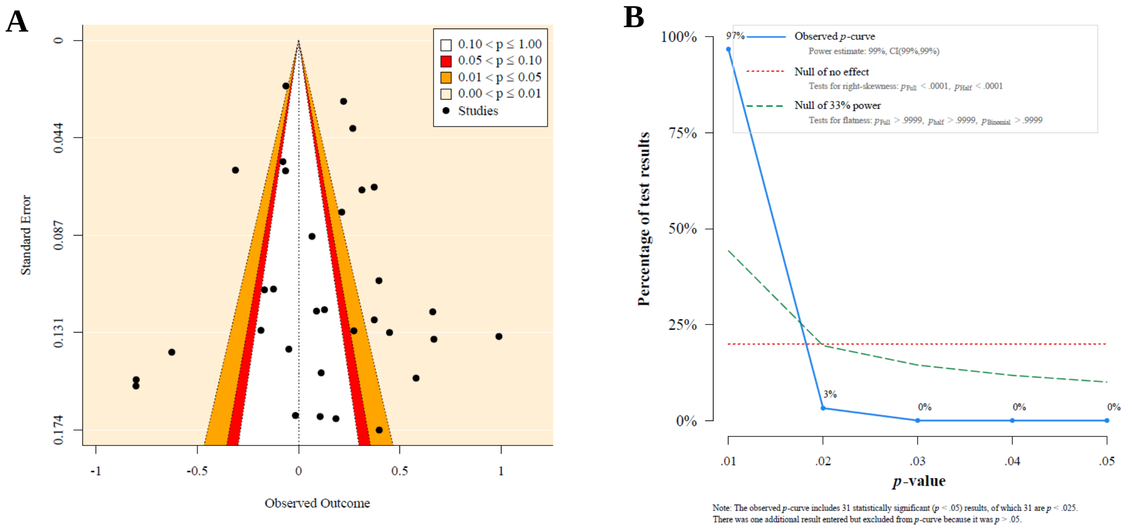




Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (V1-A) Forest plot for original or actual data after the removal of outliers and influential cases i.e. 'original-redux' model. (V1-B) Forest plot using the complete dataset after outliers and influential cases were removed – standard deviations were completed via multiple imputation i.e. 'imputed-redux' model.

Figure V2

Diagnostic Plots for Correct Identifications (Hits) in the Completed Dataset after Outliers and Influential Cases were Removed



Note. (V2-A) Funnel Plot testing for publication bias in the 'imputed-redux' model. (V2-B) P-curve analysis testing for publication bias in the 'imputed-redux' model.

Moderators

Quantity of outgroup contact in or out of the critical period

Whilst the interaction term coding (a) the level of outgroup contact ('high' or 'low') and (b) if measured contact took place within or outside of the theorized critical period for contact is a significant moderator of the aggregate effect for hits, the direction of the estimate does not follow the expected pattern. Namely, higher levels of outgroup contact post critical period ('later-life') increase the size of the observed OENE for hits (for both Original- and Imputed- All datasets: 2.12, z= -5.51, p <.0001). Conversely, lower levels of outgroup contact post critical period reduce the size of the observed OENE for hits (for both Original- and Imputed- All datasets: -1.98, z= 6.20, p <.0001).

Whilst it is theorized that outgroup contact within the critical period will lead to greater reductions in the size of the OENE than equal amounts of outgroup contact post critical period, greater outgroup contact post critical period should reduce the size of the OENE for hits more than lower levels of outgroup contact.

Quantity of outgroup contact time bands

Quantity of outgroup contact during adulthood significantly increased the size of the observed OENE for hits (Original-All: 0.58, z =2.06, p<.05; Imputed-All: 0.69, z=2.60, p<.01) whilst, quantity of outgroup contact across the entire lifespan significantly reduced the OENE for hits (Imputed-All: -.61, z=-2.18, p<.05). Both estimates follow the expected pattern namely, outgroup contact across the lifespan being more beneficial than outgroup contact during adulthood only.

Length of delay

Delay is a significant moderator which follows the expected pattern of results. Namely, a longer delay, in minutes, between encoding, or study of faces, and testing, significantly reduces the observed OENE for hits (Original-All: -.09, z=-2.70, p<.001; Imputed-All: -.08, z=-2.79, p<.01). This is noteworthy as delay is often non-existent or relatively short in the sampled studies (*Mdn* = 0) and delays in identification within real-life settings would typically not be immediate. Given the significant reduction in the size of the OENE for hits, it is recommended that studies employ longer delays which have the additional benefit of greater ecological validity.

Task/Cognitive demands

A difficult task is expected to increase the size of the observed OENE. By comparison a task with low task and/or cognitive demands while easier may not necessarily equate to a reduction in the OENE. Easier tasks could limit the intrinsic motivation needed to individuate outgroup members. Tasks requiring lower cognitive demands were found to significantly increase the observed OENE for hits (Original-All: 0.81, z=2.75, p<.01; Original-Redux: 0.33, z=2.00, p<.05; Imputed-All: 0.76, z=3.04, p<.01; Imputed-Redux: 0.35, z=2.11, p<.05).

Task

Cambridge Face Memory Tasks (CFMT) have a significantly higher OENE for hits (Original-All: 0.85, z=2.81, p<.01; Original-Redux: 0.38, z=2.09, p<.05; Imputed-All: 0.82, z=3.07, p<.01; Imputed-Redux: 0.38, z=2.07, p<.05). When compared to CFMT's, Old-New Identification tasks significantly reduce the observed OENE for hits (Original All; -.94, z=-2.70, p<.01; Imputed-All: -.89, z=-2.96, p<.01)

Table V2

Significant Moderators for Correct Identifications (Hits) Across all Models

			Model 4 (Original-all)										Model 3 (Original-redux)										Model 2 (Imputed-all)										Model 1 (Imputed-redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Quantity of outgroup contact, in or out of critical period (Interaction term)	Core	Later life-High	5	3	2.12	No	6.20	<.0001	1.45	2.79	***	92.11	-	-	-	-	-	-	-	-	-	-	5	3	2.12	No	6.20	<.0001	1.45	2.79	***	90.81	-	-	-	-	-	-	-	-	-	-
		Later life-Low vs Later life-High			-1.98	No	-5.51	<.0001	-2.68	-1.28	***		-	-	-	-	-	-	-	-	-	-			-1.98	No	-5.51	<.0001	-2.68	-1.28	***		-	-	-	-	-	-	-	-	-	-
Quantity of outgroup contact time bands (multiple)	Core	Adult (18+)	32	13	0.58	Yes	2.06	0.040	0.03	1.14	*	23.39	-	-	-	-	-	-	-	-	-	-	35	15	0.69	Yes	2.60	0.009	0.17	1.21	**	27.19	-	-	-	-	-	-	-	-	-	-
		Life span vs Adult (18+)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			-0.61	Yes	-2.18	0.030	-1.16	-0.06	*		-	-	-	-	-	-	-	-	-	-
Length of delay	Methodological / Task	-	26	10	-0.09	Yes	-2.70	0.007	-0.15	-0.02	***	64.68	-	-	-	-	-	-	-	-	-	-	29	12	-0.08	Yes	-2.79	0.005	-0.14	-0.03	**	59.85	-	-	-	-	-	-	-	-	-	-
Task / Cognitive Demand	Methodological / Task	Low vs High	32	13	0.81	-	2.75	0.006	0.23	1.38	**	32.93	29	11	0.33	-	2.00	0.046	0.01	0.65	*	11.19	35	15	0.76	-	3.04	0.002	0.27	1.26	**	33.97	32	13	0.35	-	2.11	0.035	0.02	0.67	*	11.54
Task	Methodological / Task	CFMT	32	13	0.85	-	2.81	0.005	0.26	1.45	**	26.74	29	11	0.38	-	2.09	0.037	0.02	0.73	*	7.14	35	15	0.82	-	3.07	0.002	0.30	1.35	**	28.46	32	13	0.38	-	2.07	0.038	0.02	0.73	*	7.93
		Old-New vs CFMT			-0.94	-	-2.70	0.007	-1.62	-0.26	**		-	-	-	-	-	-	-	-	-	-			-0.89	-	-2.96	0.003	-1.48	-0.30	**		-	-	-	-	-	-	-	-	-	-

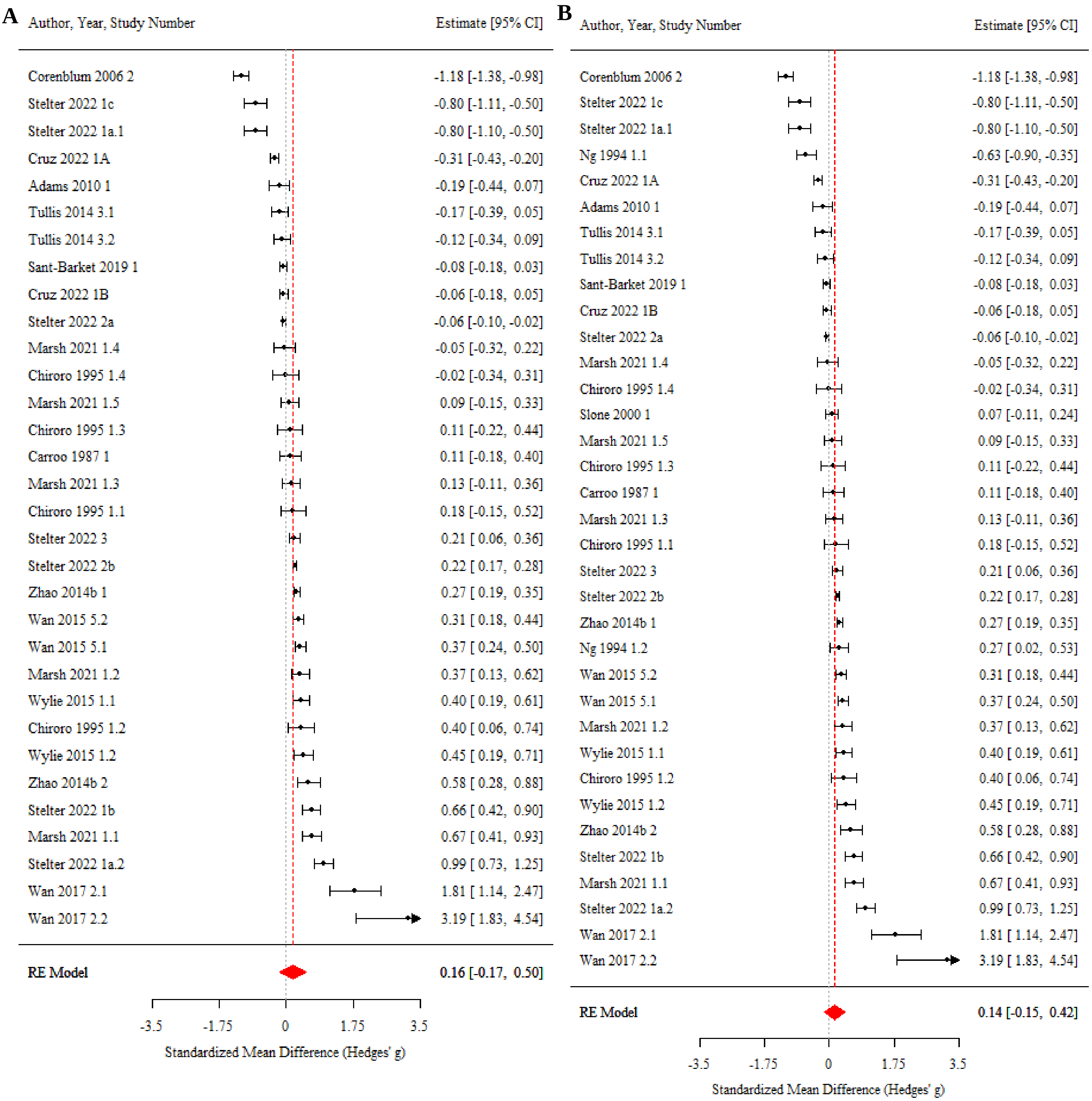
Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

Appendix W

OENE Validation – Hits: Supplemental Forest Plots

Figure W1

Forest Plots for Correct Identifications (Hits)



Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (W1-A) Forest plot for original or actual data prior to removing influential cases or outliers – ‘original-all’ model. (W1-B) Forest plot for data completed via imputation prior to removing influential cases or outliers – ‘imputed-all’ model.

Appendix X

OENE-Hits: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

The regression test for funnel plot asymmetry was significant (4.27, *z*=3.97, p<.001). This suggests the presence of publication bias.

Original-redux

The regression test was not significant (0.17, *z*=.11, p>.05). This suggests publication bias was minimal, or alternatively, that publication bias was not significantly influential on the aggregate estimate.

Imputed-all

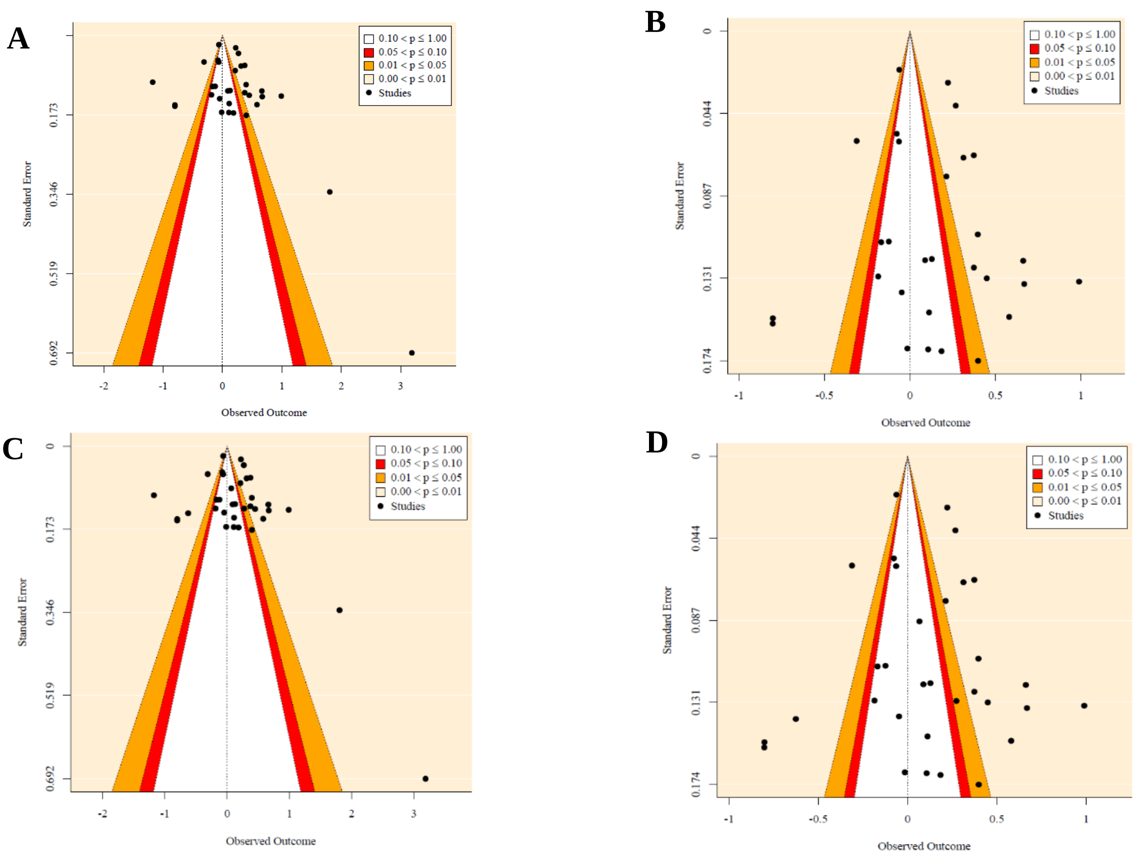
The regression test was significant (4.19, z=3.91, p<.001)

Imputed-redux

The regression test was non-significant (-.12, z=-.07, p>.05)

Figure X1

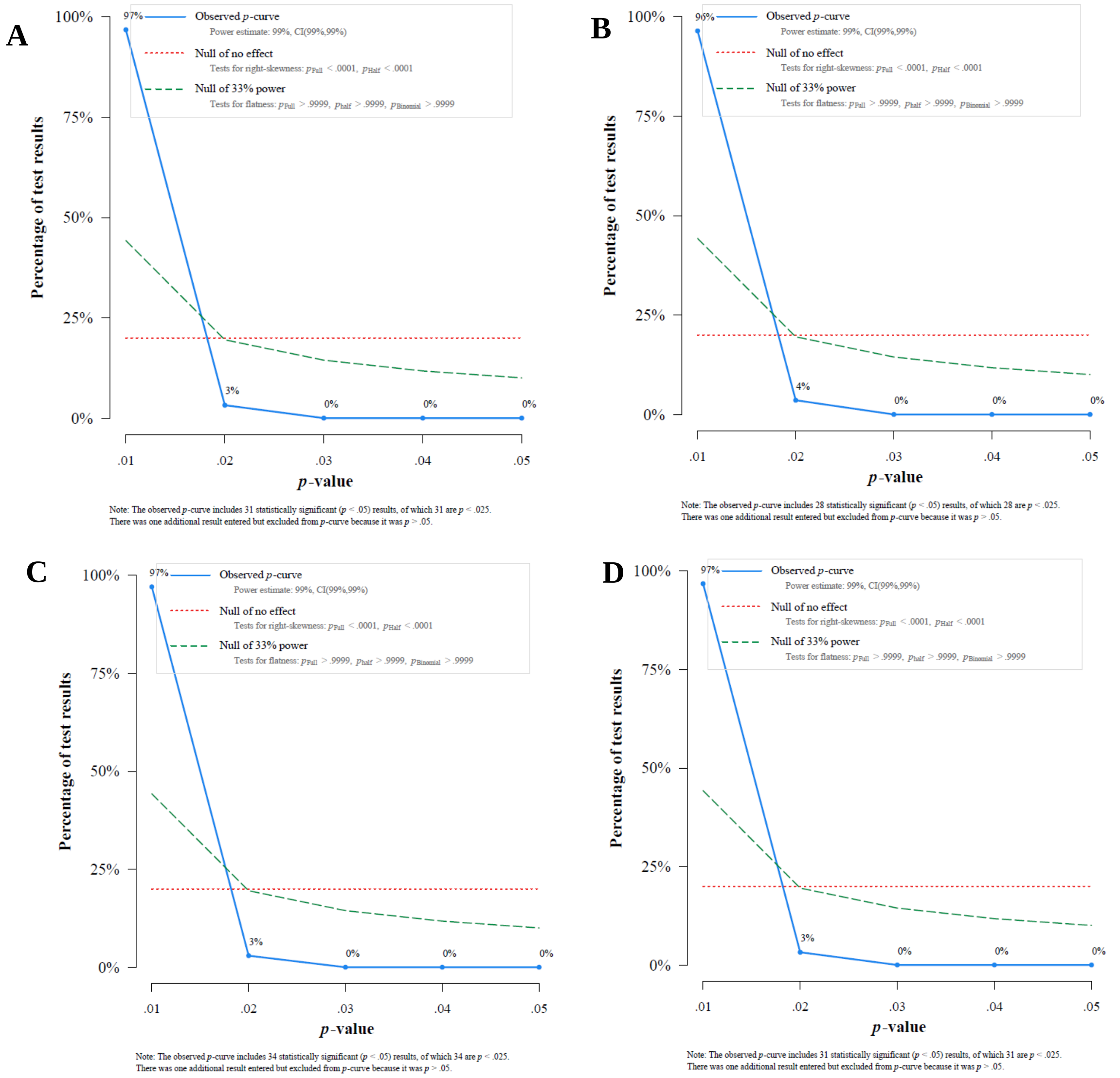
Funnel Plots for Correct Identifications (Hits) Before and After Influential Cases and Outliers were Removed



Note. (X1-A) Funnel plot testing for publication bias using reported data –'original-all' model. (X1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model. (X1-C) Funnel plot testing for publication bias using completed data – 'imputed-all' model. (X1-D) Funnel plot testing for publication bias using completed data after influential cases and outliers were removed – 'imputed-redux' model

Figure X2

P-curve Analysis for Correct Identifications (Hits) Before and After Influential Cases and Outliers were Removed



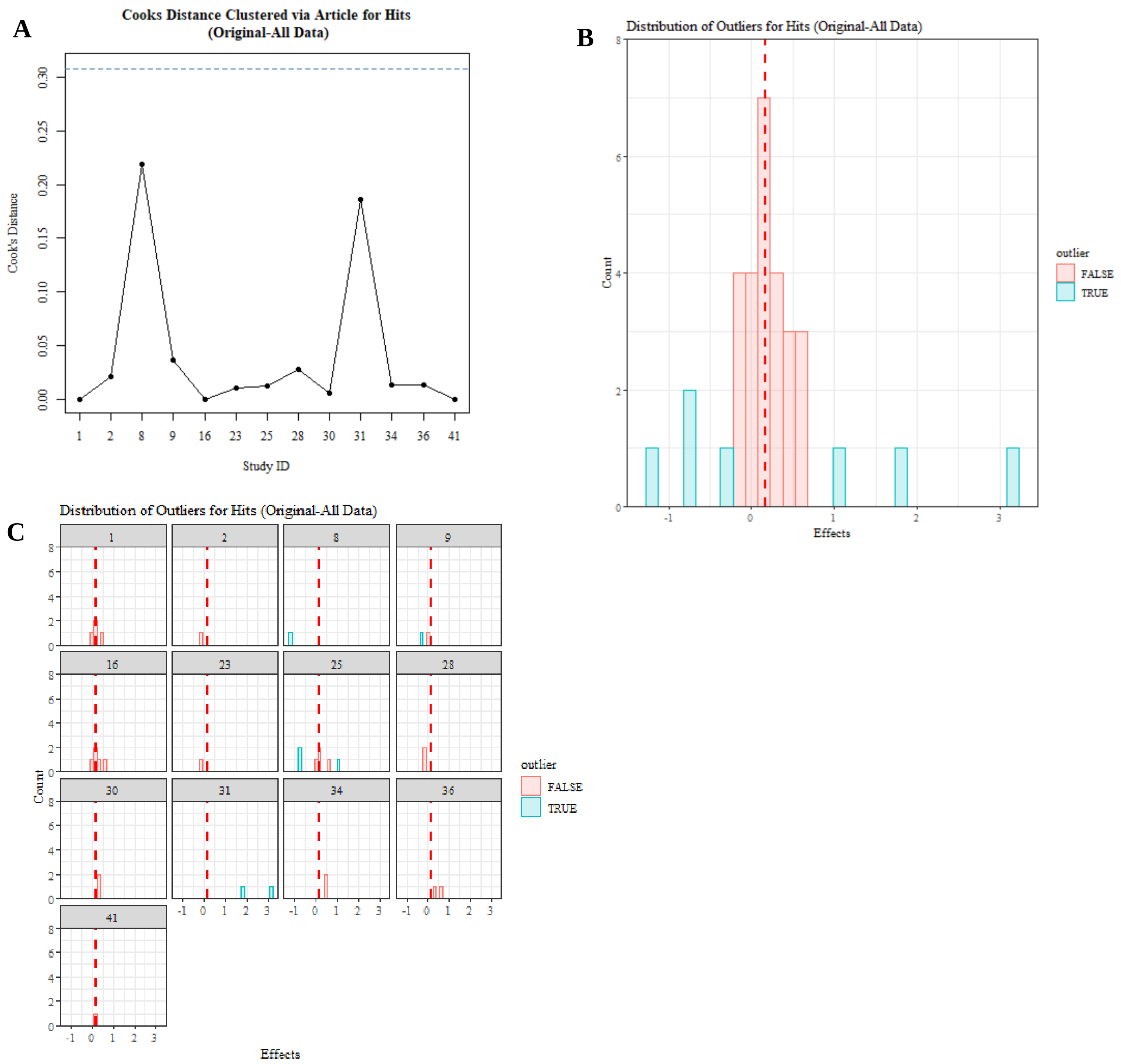
Note. (X2-A) P-curve plot testing for publication bias using reported data –'original-all' model. (X2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model. (X2-C) P-curve plot testing for publication bias using completed data – ‘imputed-all’ model. (X2-D) P-curve plot testing for publication bias using completed data after influential cases and outliers were removed – ‘imputed-redux’ model

Appendix Y

OENE-Hits: Diagnostic Plots

Figure Y1

Diagnostic Plots for Reported Data – 'original-all' Model



Note. (Y1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (Y1-B) Distribution of effect size outliers. (Y1-C) Distribution of effect size outliers faceted by article ID.

Figure Y2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model

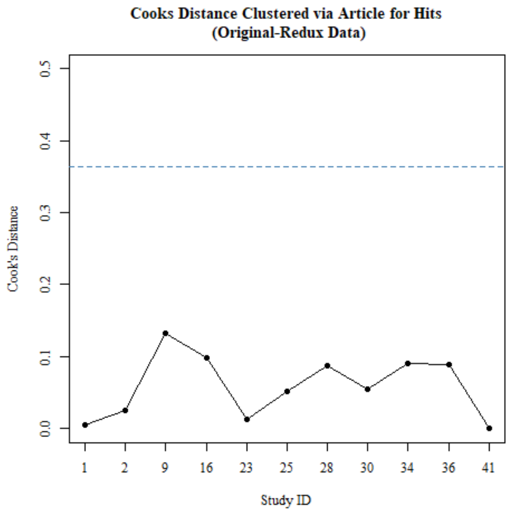
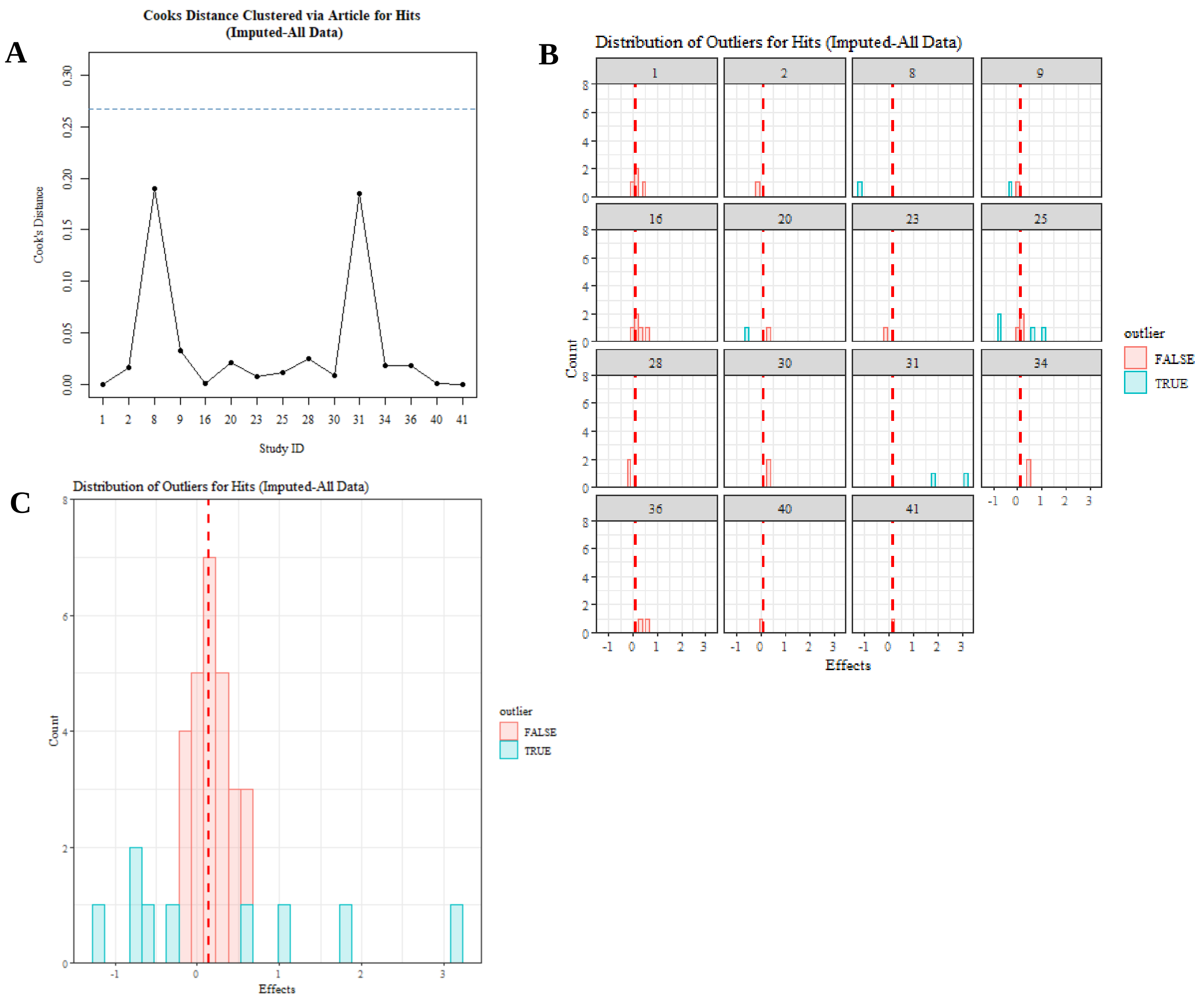


Figure Y3

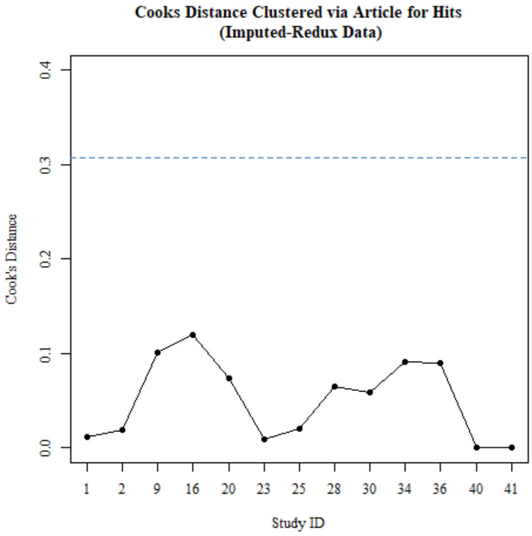
Diagnostic Plots for Completed Data – 'imputed-all' Model



Note. (Y3-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (Y3-B) Distribution of effect size outliers faceted by article ID. (Y3-C) Distribution of effect size outliers.

Figure Y4

Cooks Distance for Completed Data after Influential Cases and Outliers were Removed – 'imputed-redux' Model



Appendix Z

OENE Validation – False Alarms

Descriptive Statistics

All datasets had a higher incidence of false-alarms, or incorrect identifications, for outgroup members exceeding false-alarms for ingroup members (Original: 74.07%, Imputed: 73.33%). Descriptively thus affirming the presence of an OENE for false alarms. The percentage mean difference is noticeably larger than the mean difference observed for hits (7% or 8% respectively, see Table Z1).

Table Z1

Descriptive Table for Incorrect Identifications (False Alarms)

		Ingroup (prop)	Ingroup SD	Outgroup (prop)	Outgroup SD	Mean difference
Original All	Outgroup FA > Ingroup FA	0.20	0.13	0.33	0.15	-0.13
	Outgroup FA < Ingroup FA	0.33	0.13	0.26	0.08	0.07
	Overall	0.23	0.14	0.31	0.14	-0.08
Imputed All	Outgroup FA > Ingroup FA	0.20	0.12	0.32	0.14	-0.12
	Outgroup FA < Ingroup FA	0.31	0.13	0.25	0.07	0.06
	Overall	0.23	0.13	0.30	0.13	-0.07

Note. ‘Original-all’ = Actual data that had both the necessary means and standard deviations to calculate an effect. ‘Imputed-all’ = Data that was completed via imputation i.e. missing standard deviations were imputed and used to calculate an effect.

Meta-Analysis

The aggregate effect (*SMD*), for all datasets, confirm the presence of a statistically significant OENE for false alarms ($p<.05$). Hypothesis 1b is therefore supported.

Original

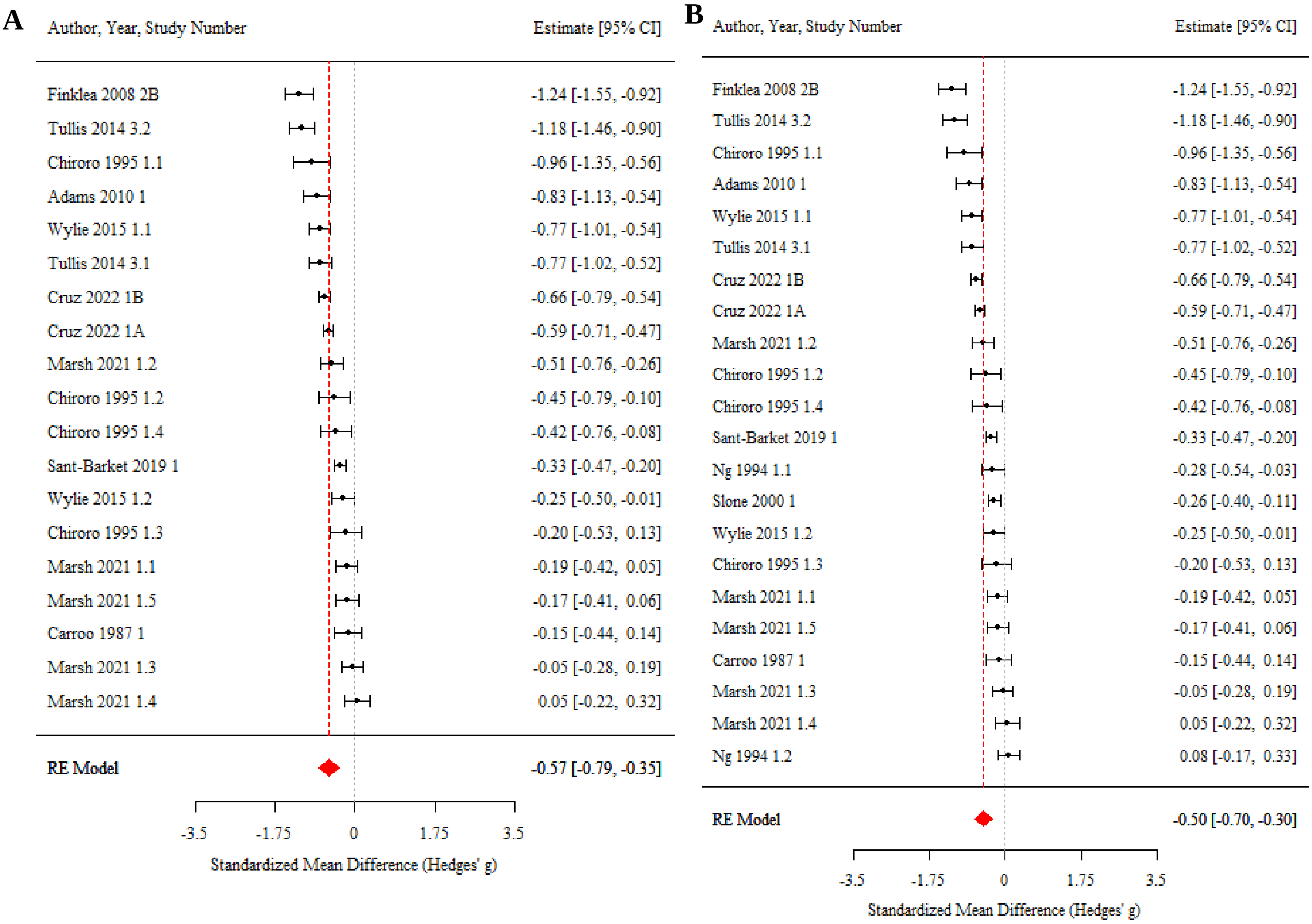
The strength of the aggregate effect size was classified as medium-to-large for both ‘-all’ and ‘-redux’ models (g=-.62 & g=-.57 respectively).

Imputed

The aggregate effect size for ‘-all’ was classified as medium-to-large whilst ‘-redux’ was classified as medium (g=-0.55, g=-50 respectively)

Figure Z1

Forest Plots for Incorrect Identifications (False Alarms) after Influential Cases and Outliers were Removed

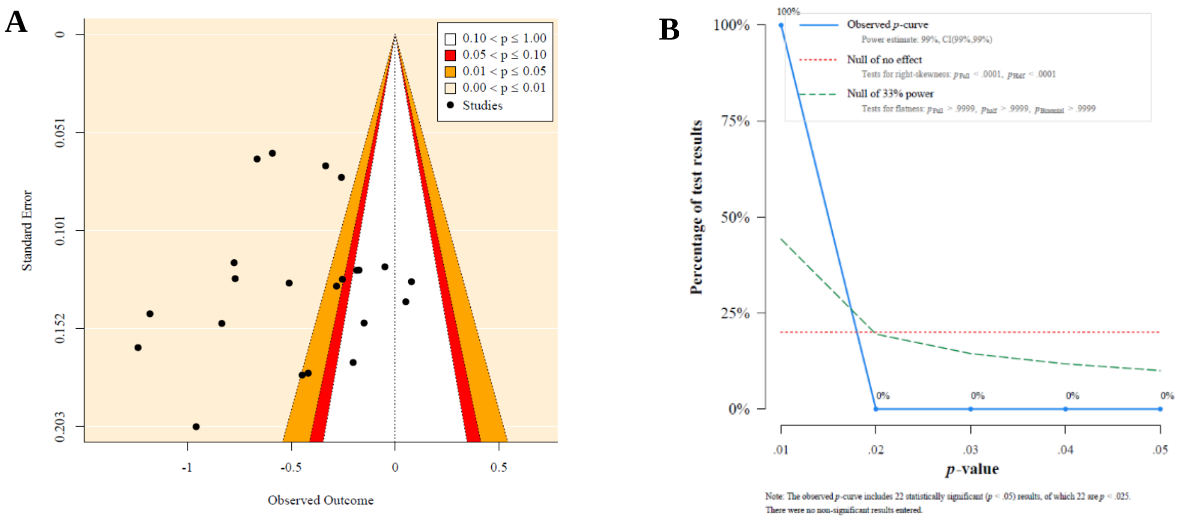


Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (Z1-A) Forest plot for original or actual data after the removal of outliers and influential cases i.e. 'original-redux' model. (Z1-B) Forest plot using the complete dataset after outliers and influential cases were removed – standard deviations were completed via multiple imputation i.e. 'imputed-redux' model.

Some publication bias was present within the sample

Figure Z2

Diagnostic Plots for Incorrect Identifications (False Alarms) in the Completed Dataset after Outliers and Influential Cases were Removed



Note. (Z2-A) Funnel Plot testing for publication bias in the ‘imputed-redux’ model. (Z2-B) P-curve analysis testing for publication bias in the ‘imputed-redux’ model

Moderators

Quality of outgroup contact

Reported. Quality of outgroup contact follows the expected direction in that greater outgroup contact reduces the size of the observed OENE for false-alarms (Both Original-all & Imputed-all: 2.32, $x=3.73$, $p<.001$). While only significant within the ‘-all’ models, this moderator accounts for a significant reduction in total variance (Original-all: 33.23% and Imputed-all: 24.79%). This underscores the importance of accounting for degree of outgroup contact

Completed. The same pattern is noted for completed data (Original-all: 2.06, $z=3.82$, $p<.001$; Original-redux: 0.95, $z=2.19$, $p<.05$; Imputed-all: 2.06, $z=4.04$, $p<.001$; Imputed-redux: 1.07, $z=2.58$, $p<.01$)

Quantity of outgroup contact time bands

Outgroup contact during adulthood follows the expected pattern of results. Namely, that contact during adulthood only increases the size of the OENE for false-alarms (Original-all: -1.12, $z=-2.54$, $p<.05$; Imputed-all: -1.04, $z=-2.48$, $p<.05$).

Outgroup contact across the lifespan does not follow the expected direction in that it would be expected that such contact would reduce the size of the OENE. The opposite is noted (Original-redux: -.55, x=-4.49, p<.001; Imputed-redux: -.47, z=-4.39, p<.001).

Explicit prejudice

Combined Option 1. More favourable outgroup attitudes, or lower explicit outgroup prejudice, follows the expected direction in that lower explicit prejudice reduces the observed OENE for false-alarms (Original-redux: 1.89, z=3.51, p<.001; Imputed-redux: 2.03, z=3.03, p<.01). The importance of accounting for explicit prejudice is underscored by the large proportion variance explained via its inclusion within the model (Original-redux: 67.58%; Imputed-redux: 48.02%)

Combined Option 2. Combined option two follows the same pattern of results (Imputed-redux: 1.35; z=2.03, p<.05; 21.96%)

Merged. Merged data follows the same pattern (Original-redux: 1.80, z=3.51, p<.001; Imputed-redux: 2.03, z=3.03, p<.01). The total variance explained is similarly large (Original-redux: 67.58%; Imputed-redux: 48.02%)

Merged Completed. Merged-completed data follows the same pattern (Imputed-redux: 1.35, z=2.03, p<.05; 21.96%)

Length of encoding (categorical)

A longer encoding, or rather greater time available to study target faces, is only beneficial if participants are motivated to individuate outgroup members.

A longer encoding (between 3-8 seconds) significantly increases the OENE for false-alarms when compared to a shorter encoding time (0-2 seconds; Original-all -.71; z=-1.97; p<.05; Imputed-all:-71, z=-2.01, p<.05)

Length of delay

Numeric. The greater the delay between encoding and test, the larger the observed OENE for false-alarms. This follows the expected direction (Original-all:-.18, z=-3.78, p<.001; Imputed-all: -.15, z=-3.15, p<.01). The proportion variance explained when delay is accounted for is large (46.08% and 32.85% respectively)

Categorical. The same pattern of results holds for the categorical delay predictor. A delay between 2 and 15 minutes increases the size of the OENE for false-alarms (Original-all:-.95, z=-3.70,p<.001; Original-redux: -.68, z=-4.31, p<.001; Imputed-all: -.81, z=-3.48, p<.001; Imputed-redux: -.58, z=-3.90, p<.001)

Type of encoding

Type of encoding refers to the degree to which outgroup faces are processed during the identification task for example basic versus deep. Basic outgroup face processing would be expected from a standard identification task, wherein participants are expecting to be tested on the study faces. A basic level of outgroup processing followed the expected direction of results in that such processing increased the OENE for false-alarms (Original-all: -.70, z=-3.10, p<.01; Original-redux: -.65; z=-5.45, p<.001; Imputed-all: -.65, z=-3.22, p<.01; Imputed-redux: -.59, z=-5.11, p<.001)

Positionality of participants relative to outgroup members

When majority members are tested on minority outgroup members, the size of the observed OENE for false-alarms increases (Original-all:-.86, z=-3.85, p<.001; Original-redux: -.60, z=-4.55, p<.001; Imputed-all: -.76, z=-4.02, p<.001; Imputed-redux: -.55, z=-4.68, p<.001). When minority members that are tested on majority outgroup members are compared to majority members tested on minority outgroup members, the size of the OENE for false-alarms decreases (Original-all: .87, z=4.10, p<.001; Imputed-all: 0.78, z=4.00, p<.001).

This follows the expected pattern of results in that the OENE is most evident within a ‘majority- minority outgroup’ context.

Task/Cognitive demands

High task or cognitive demands does follow the expected direction in that a harder task increases the size of the OENE for false-alarms (Original-all: -.65, z=-2.91, p<.01; Original-redux: -.58, z=-4.46, p<.001; Imputed-all: -.56, z=-2.99, p<.01; Imputed-redux: -.50, z=-4.32, p<.001)

Sample country’s positionality

Countries classified as belonging to the ‘Global North’ follow the expected direction in that such countries exhibit a larger OENE for false-alarms (Original-all:-.65, z=-2.87, p<.01; Original-redux: -.59, z=-4.48, p<.001; Imputed-all: -.56, z=-2.95, p<.01; Imputed-redux: -.50, z=-4.06, p<.001).

This follows the expected direction as such countries are often less integrated and less diverse than countries belonging to the ‘Global South’ and as such opportunities for varied outgroup contact could be lower.

Motivation

No motivation manipulation or instructions significantly increases the observed OENE for false-alarms (Original-all: -.64, z=-2.80, p<.01; Original-redux: -.58, z=-4.11, p<.001; Imputed-all: -.56, z=-2.88, p<.01; Imputed-redux: -.50, z=-4.06, p<.001)

This follows the expected direction expected as motivation manipulations or instructions should decrease the size of the OENE. As the majority of the sample has no such motivation manipulation, it is recommended that future studies compensate for this via the inclusion of such manipulations which have the capacity to decrease the OENE.

Novel faces at test

When novel, or new versions of study faces, are used during testing and thus memory for the face and not pictorial memory is being tapped, the OENE is expected to be larger. In this sample the opposite was observed. Namely, when novel faces were not used, which the observed OENE for false-alarms increased (Original-redux: -.64, z=-4.32, p<.001; Imputed-all: -.41, z=-2.06, p<.05; Imputed-redux; -.52, z=-4.00, p<.001).

Number of faces at encoding

As the number of to-be-studied faces present during encoding increases, thus increasing the difficulty of the task, the OENE for false-alarms increases (Original-redux: -.02, z=-3.66, p<.001; Imputed-redux: -.01, z=-2.02, p<.05). This follows the expected direction and accounting for number of faces viewed accounts for 56.61% and 11.49% of the variance explained, respectively.

Number of faces at testing

The number of faces used during the encoding does not necessarily equate to the number of faces used during testing. As such the number of to-be-studied faces at testing was additionally assessed. The same pattern of results is observed wherein the greater the number of tested faces, the larger the OENE for false-alarms (Original-redux: -.01, z=-2.24, p<.05, 25.45%)

Lures

The greater the number of lures, or never before seen faces used during testing, the larger the observed OENE for false-alarms (Original-redux: -.02, z=-3.46, p<.001; Imputed-redux: -.02, z=-2.18, p<.05).

This follows the expected direction as the difficulty of the memory task increases.

Total faces at testing

The greater the number of old, previously studied, and new, novel, faces at testing, the greater the observed OENE for false-alarms (Original-redux: -.01, z=-3.07, p<.01). This follows the expected pattern as a larger number of faces observed during testing would increase task complexity/difficulty.

Publication

Studies that were not published had a larger OENE for false-alarms (Original-redux: -.75, $z=-2.79$, $p<.01$; Imputed-redux: -.75; $z=-2.83$, $p<.01$). This is not the expected direction in that published studies are expected to have a larger OENE.

Fixed or self-paced encoding

A self-paced encoding would allow for more time to study the target faces however, greater time does not necessarily equate to the motivation to individuate outgroup faces.

A mixed encoding (combination of self-paced and fixed) increased the OENE for false-alarms (Original-redux: -.97, $z=-3.19$, $p<.01$; Imputed-redux: -.97; $z=-3.26$, $p<.01$).

Table Z2

Significant Moderators for Incorrect Identifications (False Alarms) Across all Models

			Model 4 (Original-all)										Model 3 (Original-redux)										Model 2 (Imputed-all)										Model 1 (Imputed-redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Quantity of outgroup contact (reported)	Core	-	16	4	2.32	Yes	3.73	0.000	1.10	3.54	***	33.23	-	-	-	-	-	-	-	-	-	-	16	4	2.32	Yes	3.73	0.000	1.10	3.54	***	24.79	-	-	-	-	-	-	-	-	-	-
Quantity of outgroup contact (completed)	Core	-	27	11	2.06	Yes	3.82	0.000	1.00	3.11	***	EMV	19	9	0.95	Yes	2.19	0.029	0.10	1.81	*	28.12	30	13	2.06	Yes	4.04	<.0001	1.06	3.05	***	EMV	22	11	1.07	Yes	2.58	0.010	0.26	1.88	**	30.77
Quantity of outgroup contact time bands (multiple)	Core	Adult (18+)	27	11	-1.12	Yes	-2.54	0.011	-1.99	-0.26	*	EMV	-	-	-	-	-	-	-	-	-	-	30	13	-1.04	Yes	-2.48	0.013	-1.86	-0.22	*	EMV	-	-	-	-	-	-	-	-	-	-
		Life span	-	-	-	-	-	-	-	-	-	-	19	9	-0.55	No	-4.49	<.0001	-0.78	-0.31	***	EMV	-	-	-	-	-	-	-	-	-	-	22	11	-0.47	No	-4.39	<.0001	-0.69	-0.26	***	EMV
Explicit prejudice for outgroup members (combined option one)	Core	-	-	-	-	Yes	-	-	-	-	-	-	13	9	1.80	Yes	3.51	0.001	0.79	2.80	***	67.58	-	-	-	Yes	-	-	-	-	-	-	16	11	2.03	-	3.03	0.002	0.72	3.34	**	48.02
Explicit prejudice for outgroup members (combined option two, using merged data completed with imputation)	Core	-	-	-	-	Yes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Yes	-	-	-	-	-	-	22	11	1.35	Yes	2.03	0.042	0.05	2.65	*	21.96
Harvard's 'Project Implicit' values for outgroup explicit prejudice	Core	-	-	-	-	Yes	-	-	-	-	-	-	13	9	1.80	Yes	3.51	0.001	0.79	2.80	***	67.58	-	-	-	Yes	-	-	-	-	-	-	16	11	2.03	-	3.03	0.002	0.72	3.34	**	48.02
Harvard's 'Project Implicit' values for outgroup explicit prejudice (completed)	Core	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	22	11	1.35	No	2.03	0.042	0.05	2.65	*	21.96
Length of encoding (categories)	Methodological / Task	Short vs Brief	24	9	-0.71	-	-1.97	0.049	-1.41	0.00	*	5.70	-	-	-	-	-	-	-	-	-	-	27	11	-0.71	-	-2.01	0.045	-1.39	-0.02	*	8.73	-	-	-	Yes	-	-	-	-	-	-
Length of delay	Methodological / Task	-	27	11	-0.18	-	-3.78	0.000	-0.27	-0.08	***	46.08	-	-	-	-	-	-	-	-	-	-	30	13	-0.15	-	-3.15	0.002	-0.24	-0.06	**	32.85	-	-	-	Yes	-	-	-	-	-	-
Length of delay (categories)	Methodological / Task	Brief	27	11	-0.95	-	-3.70	0.000	-1.45	-0.45	***	17.25	19	9	-0.68	Yes	-4.31	<.0001	-0.99	-0.37	***	EMV	30	13	-0.81	-	-3.48	0.001	-1.26	-0.35	***	8.87	22	11	-0.58	Yes	-3.90	<.0001	-0.87	-0.29	***	EMV

			Model 4 (Original-all)										Model 3 (Original-redux)										Model 2 (Imputed-all)										Model 1 (Imputed-redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Type of encoding	Methodological / Task	Basic	27	11	-0.70	-	-3.10	0.002	-1.15	-0.26	**	EMV	19	9	-0.65	Yes	-5.45	<.0001	-0.88	-0.41	***	8.19	30	13	-0.65	-	-3.22	0.001	-1.05	-0.25	**	EMV	22	11	-0.59	Yes	-5.11	<.0001	-0.82	-0.37	***	2.39
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Majority-Minority	25	11	-0.86	-	-3.85	0.000	-1.29	-0.42	***	EMV	17	9	-0.60	Yes	-4.55	<.0001	-0.86	-0.34	***	EMV	28	13	-0.76	-	-4.02	<.0001	-1.13	-0.39	***	EMV	20	11	-0.55	-	-4.68	<.0001	-0.77	-0.32	***	EMV
		Minority-Majority vs Majority-Minority			0.87	Yes	4.10	<.0001	0.45	1.28	***		-	-	-	-	-	-	-	-	-	-			0.78	Yes	4.00	<.0001	0.40	1.16	***		-	-	-	-	-	-	-	-	-	-
Task / Cognitive Demand	Methodological / Task	High	27	11	-0.65	Yes	-2.91	0.004	-1.08	-0.21	**	EMV	19	9	-0.58	Yes	-4.46	<.0001	-0.84	-0.33	***	EMV	30	13	-0.56	Yes	-2.99	0.003	-0.93	-0.19	**	EMV	22	11	-0.50	Yes	-4.32	<.0001	-0.73	-0.27	***	EMV
Sample Country's global positionality	Methodological / Task	Global North	25	11	-0.65	Yes	-2.87	0.004	-1.09	-0.21	**	EMV	17	9	-0.59	Yes	-4.48	<.0001	-0.85	-0.33	***	EMV	28	13	-0.56	Yes	-2.95	0.003	-0.94	-0.19	**	EMV	20	11	-0.50	Yes	-4.31	<.0001	-0.73	-0.28	***	EMV
Presence of motivation manipulation or instructions	Methodological / Task	None	27	11	-0.64	Yes	-2.80	0.005	-1.10	-0.19	**	EMV	19	9	-0.58	Yes	-4.11	<.0001	-0.86	-0.31	***	EMV	30	13	-0.56	Yes	-2.88	0.004	-0.94	-0.18	**	EMV	22	11	-0.50	Yes	-4.06	<.0001	-0.74	-0.26	***	EMV
Use of novel faces at testing	Methodological / Task	No	-	-	-	Yes	-	-	-	-	-	-	19	9	-0.64	No	-4.32	<.0001	-0.93	-0.35	***	EMV	30	13	-0.41	Yes	-2.06	0.040	-0.80	-0.02	*	2.86	22	11	-0.52	-	-4.00	<.0001	-0.78	-0.27	***	EMV
Number of to-be-studied faces present during encoding	Methodological / Task	-	-	-	-	Yes	-	-	-	-	-	-	19	9	-0.02	Yes	-3.66	0.00	-0.03	-0.01	***	56.61	-	-	-	Yes	-	-	-	-	-	-	22	11	-0.01	Yes	-2.02	0.044	-0.02	0.00	*	11.49
Number of tested faces (present at both encoding and test)	Methodological / Task	-	-	-	-	Yes	-	-	-	-	-	-	19	9	-0.01	Yes	-2.24	0.025	-0.02	0.00	*	25.45	-	-	-	Yes	-	-	-	-	-	-	-	-	-	Yes	-	-	-	-	-	-
Number of lures (new faces) present during testing	Methodological / Task	-	-	-	-	Yes	-	-	-	-	-	-	17	8	-0.02	Yes	-3.46	0.001	-0.03	-0.01	***	50.67	-	-	-	Yes	-	-	-	-	-	-	20	10	-0.02	Yes	-2.18	0.029	-0.03	0.00	*	1.89
Total number of studied (old) and new faces at testing	Methodological / Task	-	-	-	-	-	-	-	-	-	-	-	17	8	-0.01	Yes	-3.07	0.002	-0.02	0.00	**	39.22	-	-	-	No	-	-	-	-	-	-	-	-	-	No	-	-	-	-	-	-
Published	Methodological / Task	No	-	-	-	-	-	-	-	-	-	-	19	9	-0.75	No	-2.79	0.005	-1.28	-0.22	**	EMV	-	-	-	-	-	-	-	-	-	-	22	11	-0.75	Yes	-2.83	0.005	-1.27	-0.23	**	EMV

			Model 4 (Original-all)										Model 3 (Original-redux)										Model 2 (Imputed-all)										Model 1 (Imputed-redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Fixed or self-paced encoding	Methodologica l / Task	Mixed	-	-	-	-	-	-	-	-	-	-	19	9	-0.97	-	-3.19	0.001	-1.56	-0.37	**	0.13	-	-	-	-	-	-	-	-	-	-	22	11	-0.97	-	-3.26	0.001	-1.55	-0.39	**	17.35

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

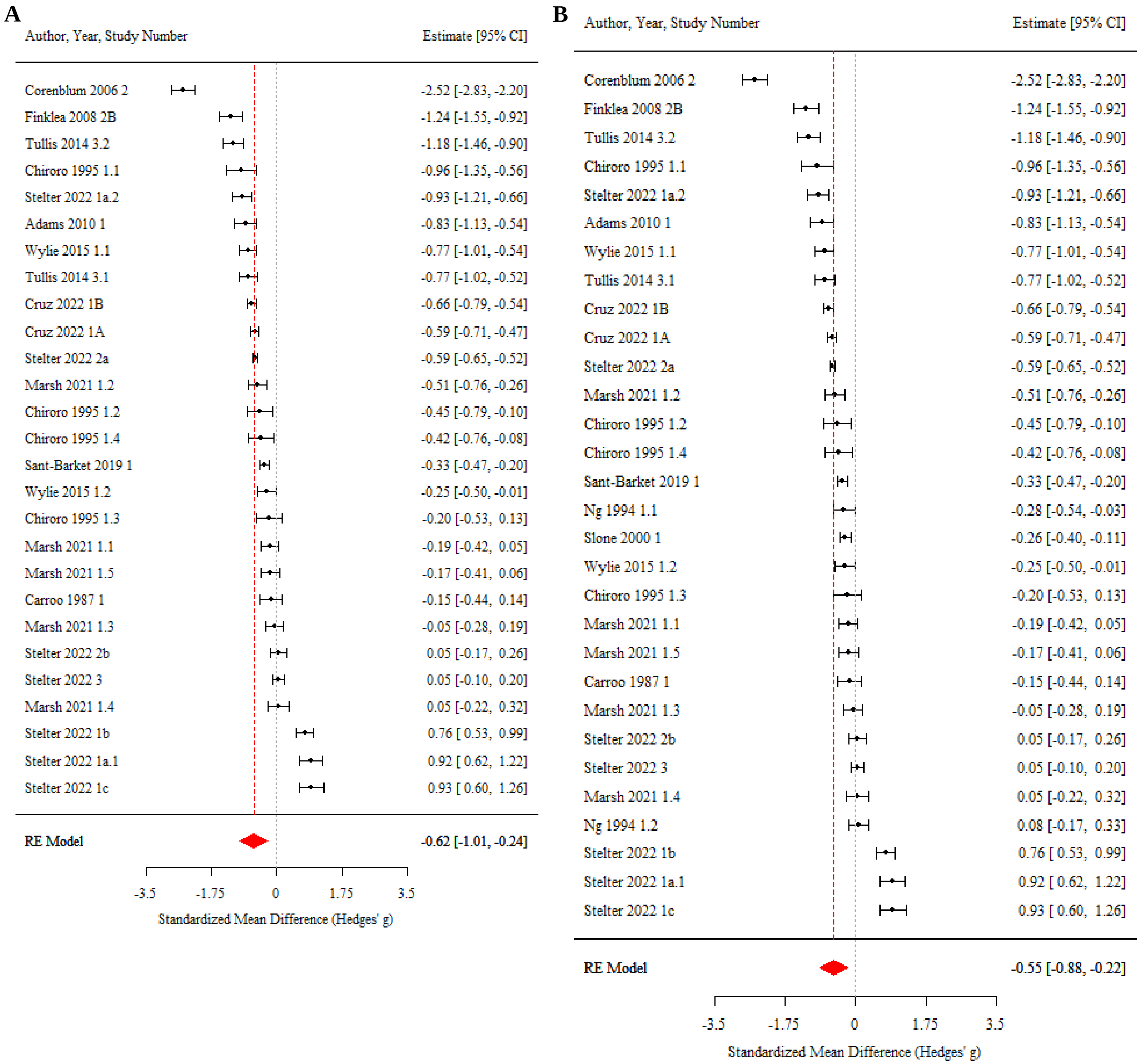
The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance

Appendix AA

OENE Validation - False Alarms: Supplemental Forest Plots

Figure AA1

Forest Plots for Incorrect Identifications (False Alarms)



Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (AA1-A) Forest plot for original or actual data prior to removing influential cases or outliers – ‘original-all’ model. (AA1-B) Forest plot for data completed via imputation prior to removing influential cases or outliers – ‘imputed-all’ model.

Appendix AB

OENE-False Alarms: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

The regression test was not significant (2.69, z=.80, p>.05).

Original-redux

. The regression test was not significant (-3.33, z=-1.12, p>.05).

Imputed-all

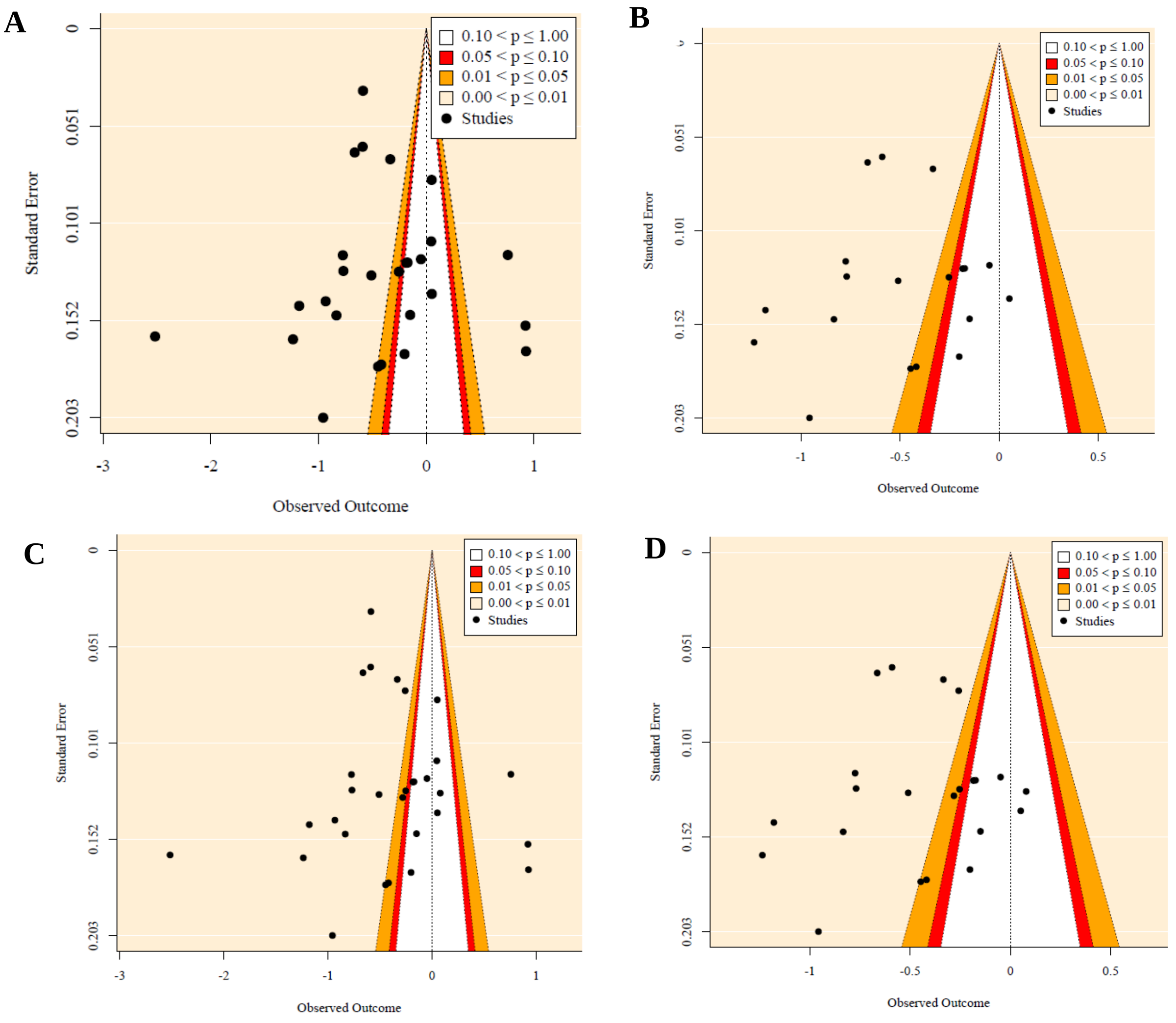
The regression test was not significant (1.74, z=.55, p>.05).

Imputed-redux

The regression test was not significant (-3.52, z=-1.29, p>.05).

Figure AB1

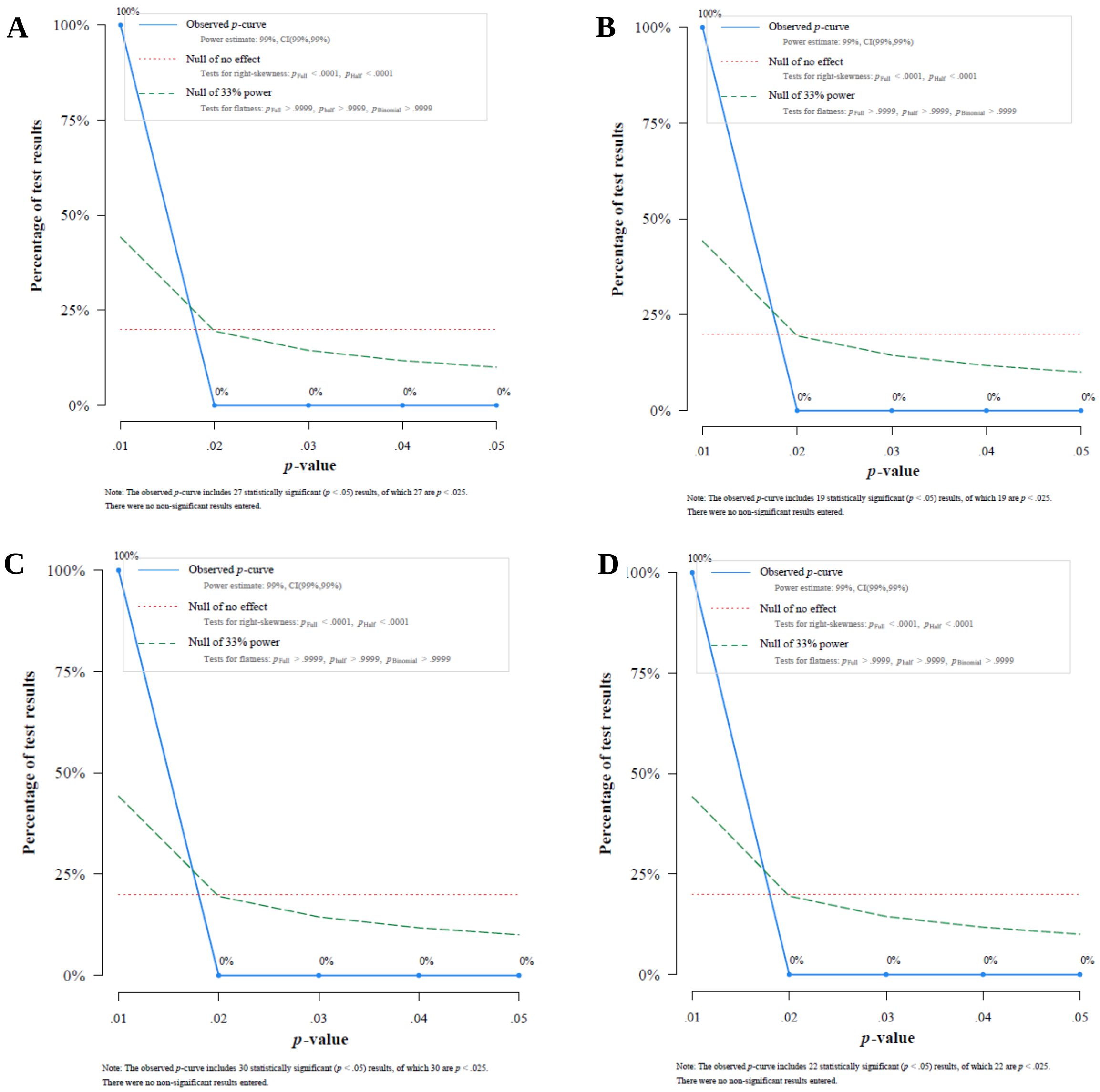
Funnel Plots for Incorrect Identifications (False Alarms) Before and After Influential Cases and Outliers were Removed



Note. (AB1-A) Funnel plot testing for publication bias using reported data –'original-all' model. (AB1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model. (AB1-C) Funnel plot testing for publication bias using completed data – 'imputed-all' model. (AB1-D) Funnel plot testing for publication bias using completed data after influential cases and outliers were removed – 'imputed-redux' model

Figure AB2

P-curve Analysis for Incorrect Identifications (False Alarms) Before and After Influential Cases and Outliers were Removed



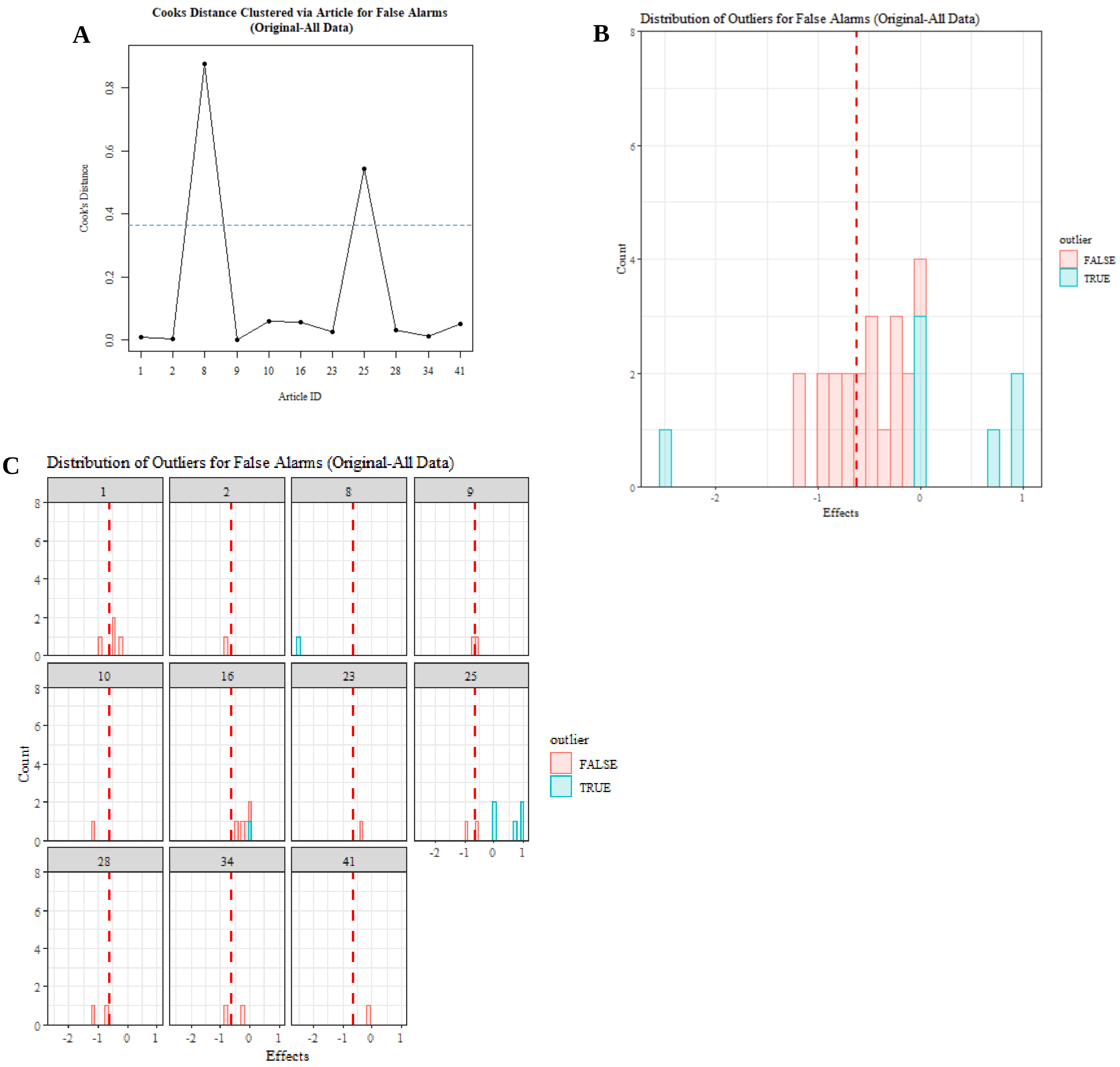
Note. (AB2-A) P-curve plot testing for publication bias using reported data – 'original-all' model. (AB2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model. (AB2-C) P-curve plot testing for publication bias using completed data – ‘imputed-all’ model. (AB2-D) P-curve plot testing for publication bias using completed data after influential cases and outliers were removed – ‘imputed-redux’ model

Appendix AC

OENE-False Alarms: Diagnostic Plots

Figure AC1

Diagnostic Plots for Reported Data – 'original-all' Model



Note. (AC1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AC1-B) Distribution of effect size outliers. (AC1-C) Distribution of effect size outliers faceted by article ID.

Figure AC2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model

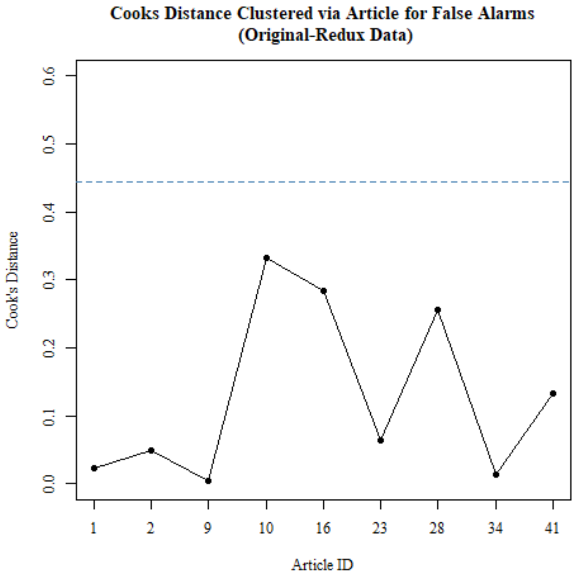
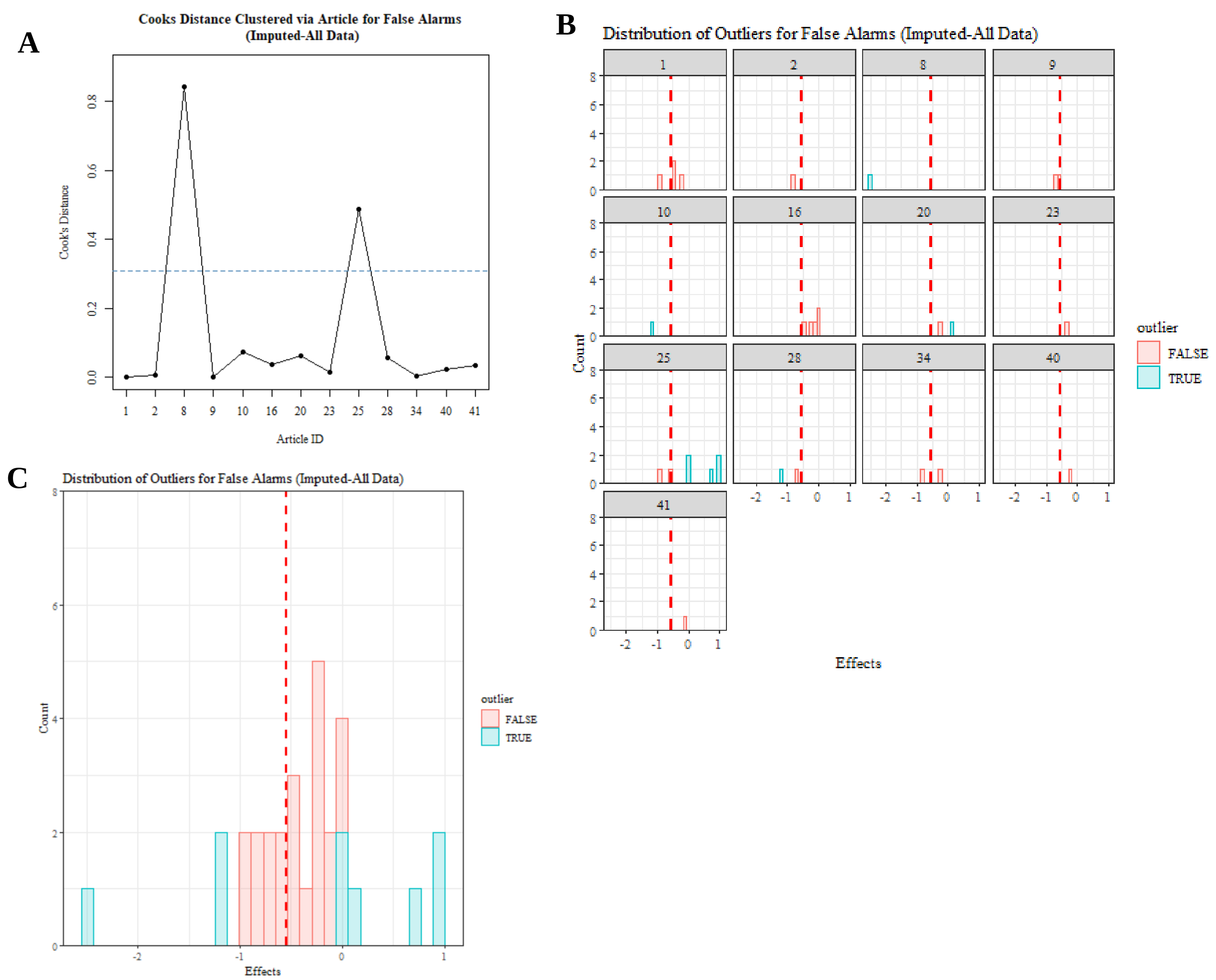


Figure AC3

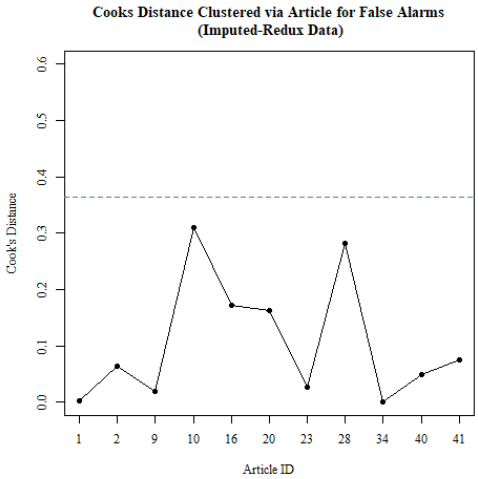
Diagnostic Plots for Completed Data – 'imputed-all' Model



Note. (AC3-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AC3-B) Distribution of effect size outliers faceted by article ID. (AC3-C) Distribution of effect size outliers

Figure AC4

Cooks Distance for Completed Data after Influential Cases and Outliers were Removed – 'imputed-redux' Model



Appendix AD

OENE Validation – Response Bias

Descriptive Statistics

A positive value indicates a more conservative criterion for identification decisions whilst a negative value indicates a more liberal criterion for such decisions. The ideal situation would therefore entail a neutral value (0, Stanislaw & Todorov, 1999). Of the two methods to calculate response bias namely, C and beta, the majority of the sample used C (Original: 73.33%, Imputed: 80.00%).

Furthermore, the majority of the sample had a lower response bias value for outgroup targets when compared to ingroup targets (66.67%). Average response bias values overall affirm this trend, with the mean difference between ingroup and outgroup values falling between 12 and 13%. This descriptively affirms the presence of an OENE for response bias.

Table AD1

Descriptives Table for Response Bias

		Ingroup	Ingroup SD	Outgroup	Outgroup SD	Mean difference
Original All	Ingroup RB > Outgroup RB	0.39	0.17	0.12	0.21	0.27
	Ingroup RB < Outgroup RB	0.05	0.12	0.19	0.17	-0.14
	Overall	0.28	0.22	0.14	0.19	0.13
Imputed All	Ingroup RB > Outgroup RB	0.29	0.20	0.04	0.18	0.24
	Ingroup RB < Outgroup RB	0.04	0.10	0.16	0.16	-0.12
	Overall	0.21	0.21	0.08	0.18	0.12

Note. ‘Original-all’ = Actual data that had both the necessary means and standard deviations to calculate an effect. ‘Imputed-all’ = Data that was completed via imputation i.e. missing standard deviations were imputed and used to calculate an effect.

Meta-Analysis

The aggregate effect (*SMD*), for all datasets, confirm the presence of a statistically significant OENE for response bias ($p<.05$). Hypothesis 1d is therefore supported.

Original

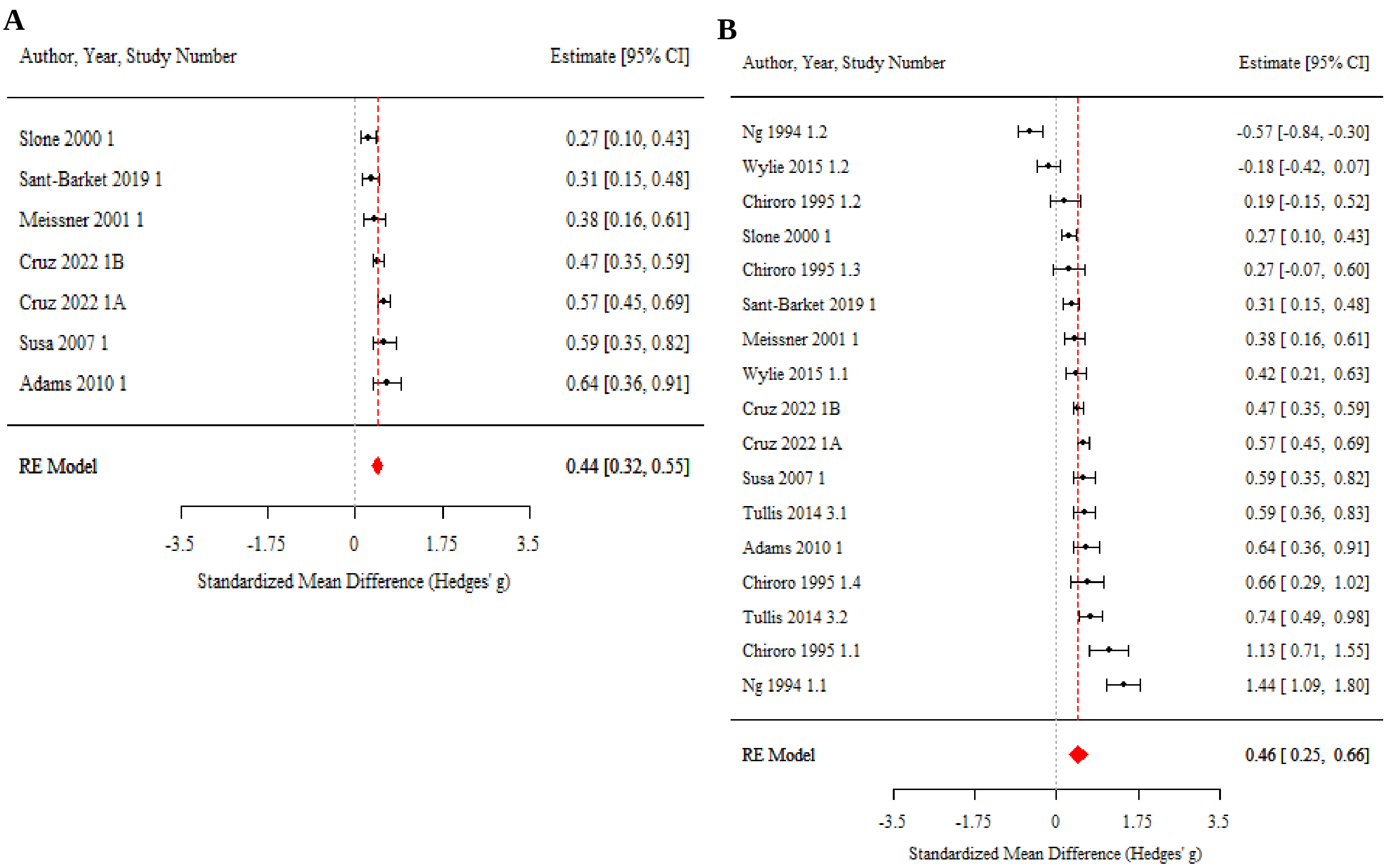
The strength of the aggregate effect size for '-all' was classified as medium-to-large, whilst (-redux' was classified as small-to-medium (*g*=.63 & *g*= .44 respectively).

Imputed

The aggregate effect size for both '-all' and 'redux' models was classified as small-to-medium (*g*=.47, *g*=.46 respectively).

Figure AD1

Forest Plots for Response Bias after Influential Cases and Outliers were Removed



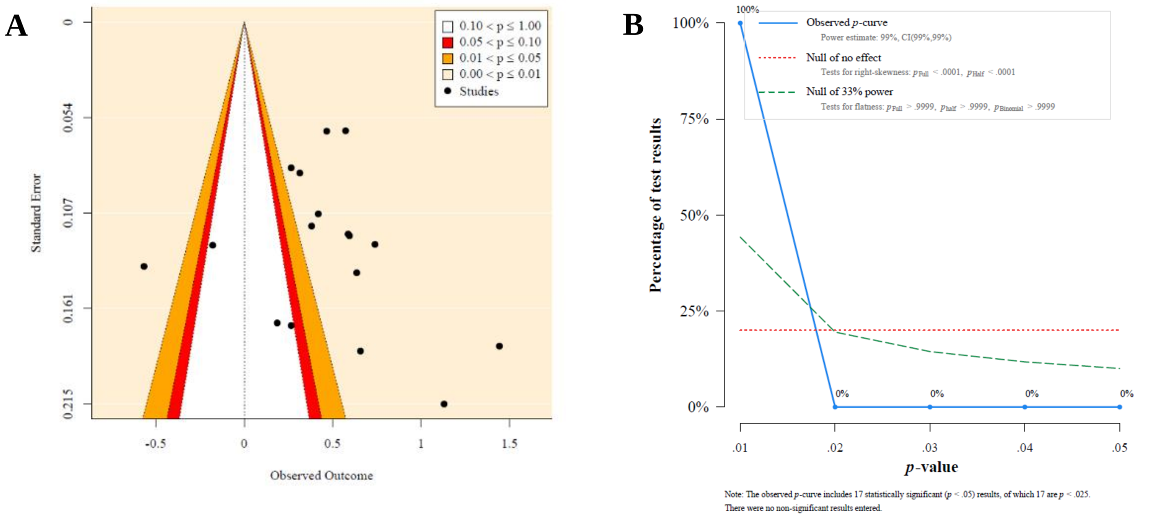
Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (AD1-A) Forest plot for original or actual data after the removal of outliers and influential cases i.e. 'original-redux' model. (AD1-B) Forest plot using the complete dataset after outliers and influential cases were removed – standard deviations were completed via multiple imputation i.e. 'imputed-redux' model.

Publication Bias

Publication bias was present in the sample.

Figure AD2

Diagnostic Plots for Response Bias in the Completed Dataset after Outliers and Influential Cases were Removed



Note. (AD2-A) Funnel Plot testing for publication bias in the 'imputed-redux' model. (AD2-B) P-curve analysis testing for publication bias in the 'imputed-redux' model.

Moderators

Implicit Outgroup Prejudice

Combined option 1. Higher implicit outgroup prejudice increases the size of the observed OENE for response bias (Original-all: 2.45, z=2.76, p<.01). This follows the expected direction in that prejudice exacerbates the size of the OENE.

Combined option 2. The same pattern is observed (Original-all: 2.31, z=2.69, p<.01)

Merged. The same pattern is observed (Original-all: 2.45, z=2.76, p<.01)

Merged completed. The same pattern is observed (Original-all: 2.31, z=2.69, p<.01).

Quantity of outgroup contact time bands

Quantity of outgroup contact during adulthood follows the expected pattern in that contact during adulthood only increases the observed OENE for response bias (Original-all: 0.95, z=2.28, p<.05)

Contact with outgroup members across the life span does not follow the expected direction. Life span outgroup contact increases the OENE for response bias (Original-redux: 0.42, z=6.72, p<.001; Imputed-redux: .45, z=4.00, p<.001).

Quality of outgroup contact

Higher quality of outgroup contact reduces the observed OENE for response bias (Imputed-redux: -1.14, z=02.14, p<.05). Accounting for quality of contact explains 20.19% of the explained variance. Quality of contact therefore follows the expected pattern of results.

Length of delay

Numeric. The greater the delay between encoding and test, the greater the observed OENE for response bias (Original-all: .21, z=4.68, p<.001; Original-redux: 0.05, z=2.81, p<.01; Imputed-all: 0.15, z=3.37, p<.001). This follows the expected direction. The inclusion of delay explains a considerable amount of variance within the models (74.51%, 93.30% and 37.60% respectively)

Categorical. A brief delay, between 2-15minutes, increases the OENE for response bias (Original-all: .83, z=2.60, p<.01; Original-redux: .48, z=8.99, p<.001; Imputed-all: .65, z=3.05, p<.01; Imputed-redux: .40, z=2.90, p<.01).

Type of encoding

A basic level of encoding increases the observed OENE for response bias (Original-all: .67, z=2.05, p<.05; Original-redux: .45, z=6.45, p<.001; Imputed-all: .56, z=2.83, p<.01; Imputed-redux: .46, z=3.70, p<.001). This follows the expected direction.

Positionality of participants relative to outgroup members

Consistent with the expected pattern of results, majority members tested on minority outgroup members exhibited a larger OENE for response bias (Original-redux: .42, z=5.36, p<.001; Imputed-all: .56, z=3.12, p<.01; Imputed-redux: .52, z=4.30, p<.001). Minority members tested on majority outgroup members exhibited a lower OENE for response bias when compared to majority members tested on minority outgroup members (Original-all: -.47, z=-2.20, p<.05; Imputed-all: -.61, z=-2.61, p<.01; Imputed-redux: -1.48, z=-4.27, p<.001)

Motivation

No motivation manipulation or instructions increases the observed OENE for response bias (Original-all: .65, z=1.96, p<.05; Original-redux: .41, z=5.81, p<.001; Imputed-all: .47, z=2.55, p<.05; Imputed-redux: .44, z=3.51, p<.001). This is consistent with the expectant pattern of results.

Novel faces at test

Not using novel faces at test increases the size of the OENE for response bias (Original-redux: 0.46, z=5.16, p<.001; Imputed-redux: .51, z=3.28, p<.01). This is not the expected pattern of results.

Fixed or self-paced encoding

Self-paced length of encoding time increased the size of the observed OENE for response bias (Original-redux: 0.38, $z=2.22$, $p<.05$). The same pattern of results was observed for a mixed encoding, a combination of self-paced and fixed (Imputed-redux: .67, $z=2.08$, $p<.05$).

Publication

Non- published studies exhibited a larger OENE for response bias (Original-redux: .42, $z=4.31$, $p<.001$). This is counter to expectations.

Task/Cognitive demands

High task and or cognitive demands increases the OENE for response bias (Imputed-all: 0.51, $z=2.84$, $p<.01$; Imputed-redux: .50, $z=5.53$, $p<.001$). This follows the expected pattern of results.

Sample country's positionality

Consistent with expectations, countries belonging to the 'Global North' exhibit a larger OENE for response bias (Imputed-all: .47, $z=2.56$, $p<.05$; Imputed-redux: .43, $z=3.56$, $p<.001$).

Table AD2

Significant Moderators for Response Bias Across Models

			Model 4 (Original-all)										Model 3 (Original-redux)										Model 2 (Imputed-all)										Model 1 (Imputed-redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Implicit prejudice for outgroup members (combined option one)	Core	-	13	6	2.45	Yes	2.76	0.006	0.71	4.19	**	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Implicit prejudice for outgroup members (combined option two, using merged data completed with imputation)	Core	-	15	8	2.31	Yes	2.69	0.007	0.62	3.99	**	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Harvard's 'Project Implicit' values for outgroup implicit prejudice	Core	-	13	6	2.45	Yes	2.76	0.006	0.71	4.19	**	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Harvard's 'Project Implicit' values for outgroup implicit prejudice (completed)	Core	-	15	8	2.31	Yes	2.69	0.007	0.62	3.99	**	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Quantity of outgroup contact time bands (multiple)	Core	Adult (18+)	15	8	0.95	Yes	2.28	0.023	0.13	1.76	*	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Life span	-	-	-	-	-	-	-	-	-	-	7	6	0.42	No	6.75	<.0001	0.29	0.54	***	4.89	-	-	-	-	-	-	-	-	-	-	17	10	0.45	No	4.00	<.0001	0.23	0.66	***	EMV
Quality of outgroup contact (completed)	Core	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	10	-1.14	Yes	-2.14	0.033	-2.19	-0.09	*	20.19
Length of delay	Methodological / Task	-	15	8	0.21	Yes	4.69	<.0001	0.12	0.30	***	74.51	7	6	0.05	Yes	2.81	0.005	0.02	0.09	**	93.30	30	13	0.15	Yes	3.37	0.001	0.06	0.23	***	37.60	-	-	-	-	-	-	-	-	-	-
Length of delay (categories)	Methodological / Task	Brief	15	8	0.83	Yes	2.60	0.009	0.20	1.46	**	7.37	7	6	0.48	Yes	8.99	<.0001	0.37	0.58	***	45.75	30	13	0.65	Yes	3.05	0.002	0.23	1.07	**	5.45	17	10	0.40	Yes	2.90	0.004	0.13	0.67	**	EMV
Type of encoding	Methodological / Task	Basic	15	8	0.67	Yes	2.05	0.040	0.03	1.31	*	EMV	7	6	0.45	Yes	6.45	<.0001	0.31	0.59	***	EMV	30	13	0.56	Yes	2.83	0.005	0.17	0.94	**	EMV	17	10	0.46	Yes	3.70	0.000	0.22	0.70	***	EMV
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Majority-Minority	-	-	-	-	-	-	-	-	-	-	6	5	0.42	Yes	5.36	<.0001	0.27	0.58	***	EMV	27	12	0.56	Yes	3.12	0.002	0.21	0.91	**	1.23	14	9	0.52	Yes	4.30	<.0001	0.28	0.76	***	27.13

			Model 4 (Original-all)										Model 3 (Original-redux)										Model 2 (Imputed-all)										Model 1 (Imputed-redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
		Minority-Majority vs Majority-Minority	14	7	-0.47	Yes	-2.20	0.028	-0.88	-0.05	*	EMV	-	-	-	-	-	-	-	-	-	-			-0.61	Yes	-2.61	0.009	-1.07	-0.15	**				-1.48	Yes	-4.27	<.0001	-2.15	-0.80	***	
Presence of motivation manipulation or instructions	Methodological / Task	None	15	8	0.65	Yes	1.96	0.050	0.00	1.30	*	EMV	7	6	0.41	Yes	5.81	<.0001	0.27	0.55	***	EMV	30	13	0.47	Yes	2.55	0.011	0.11	0.83	*	EMV	17	10	0.44	Yes	3.51	0.000	0.19	0.68	***	EMV
Use of novel faces at testing	Methodological / Task	No	-	-	-	-	-	-	-	-	-	-	7	6	0.46	No	5.16	<.0001	0.28	0.63	***	EMV	-	-	-	-	-	-	-	-	-	-	17	10	0.51	No	3.28	0.001	0.20	0.81	**	EMV
Fixed or self-paced encoding	Methodological / Task	Self-paced	-	-	-	-	-	-	-	-	-	-	7	6	0.38	-	2.22	0.026	0.04	0.72	*	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Mixed	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	10	0.67	-	2.08	0.037	0.04	1.29	*	EMV
Published	Methodological / Task	No	-	-	-	-	-	-	-	-	-	-	7	6	0.42	No	4.31	<.0001	0.23	0.61	***	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Task / Cognitive Demand	Methodological / Task	High	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	30	13	0.50	Yes	2.84	0.005	0.16	0.85	**	EMV	17	10	0.50	Yes	4.53	<.0001	0.28	0.72	***	2.31
Sample Country's global positionality	Methodological / Task	Global North	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	28	13	0.47	Yes	2.56	0.010	0.11	0.83	*	EMV	15	10	0.43	Yes	3.56	0.000	0.19	0.67	***	EMV

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

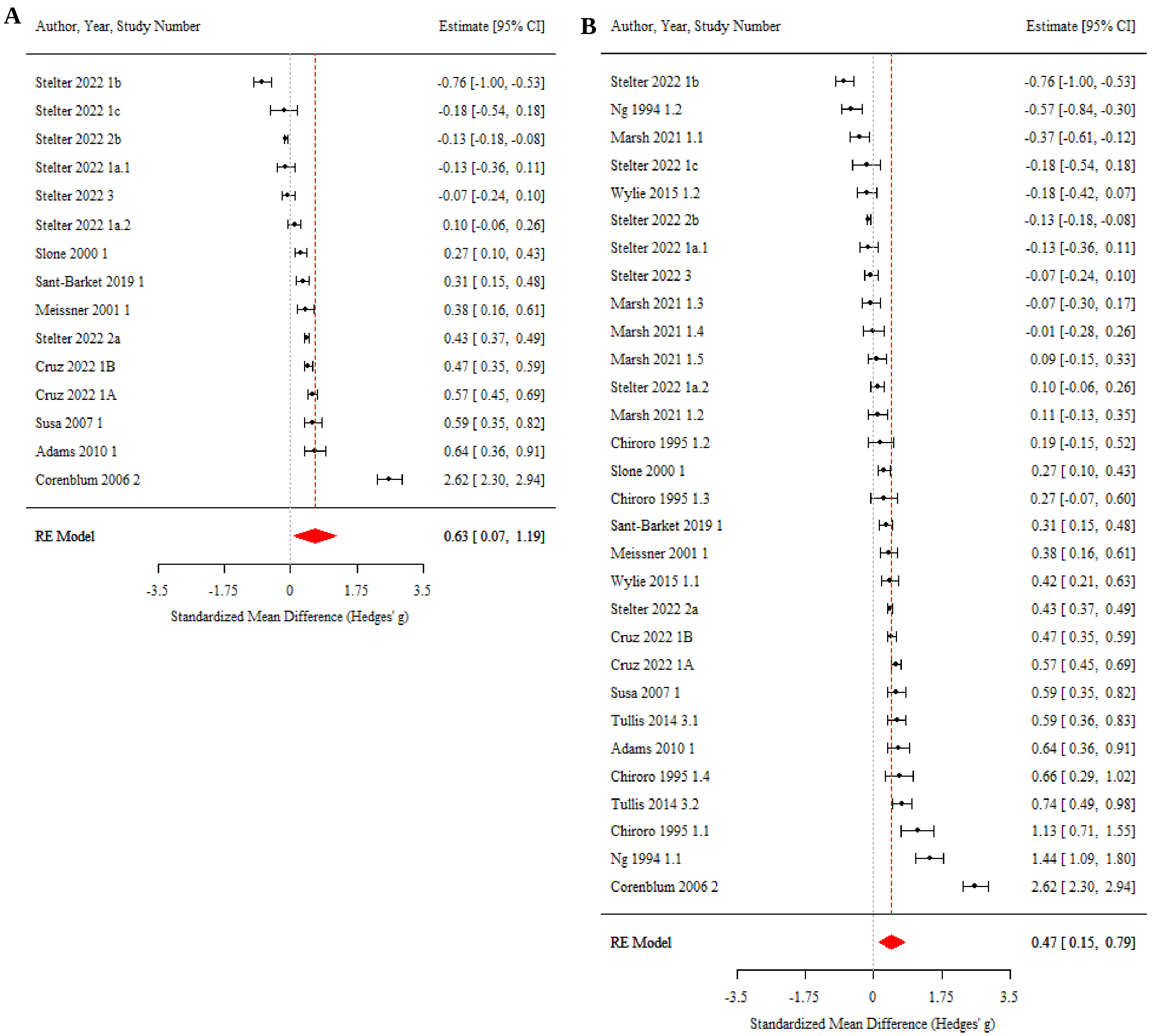
The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the model meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as 'EMV' or exceeds model variance

Appendix AE

OENE Validation-Response Bias: Supplemental Forest Plots

Figure AE1

Forest Plots for Response Bias



Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (AE1-A) Forest plot for original or actual data prior to removing influential cases or outliers – ‘original-all’ model. (AE1-B) Forest plot for data completed via imputation prior to removing influential cases or outliers – ‘imputed-all’ model.

Appendix AF

OENE-Response Bias: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

The regression test was not significant (-2.16, z=-.93, p>.05).

Original-redux

. The regression test was not significant (1.96, z=.79, p>.05).

Imputed-all

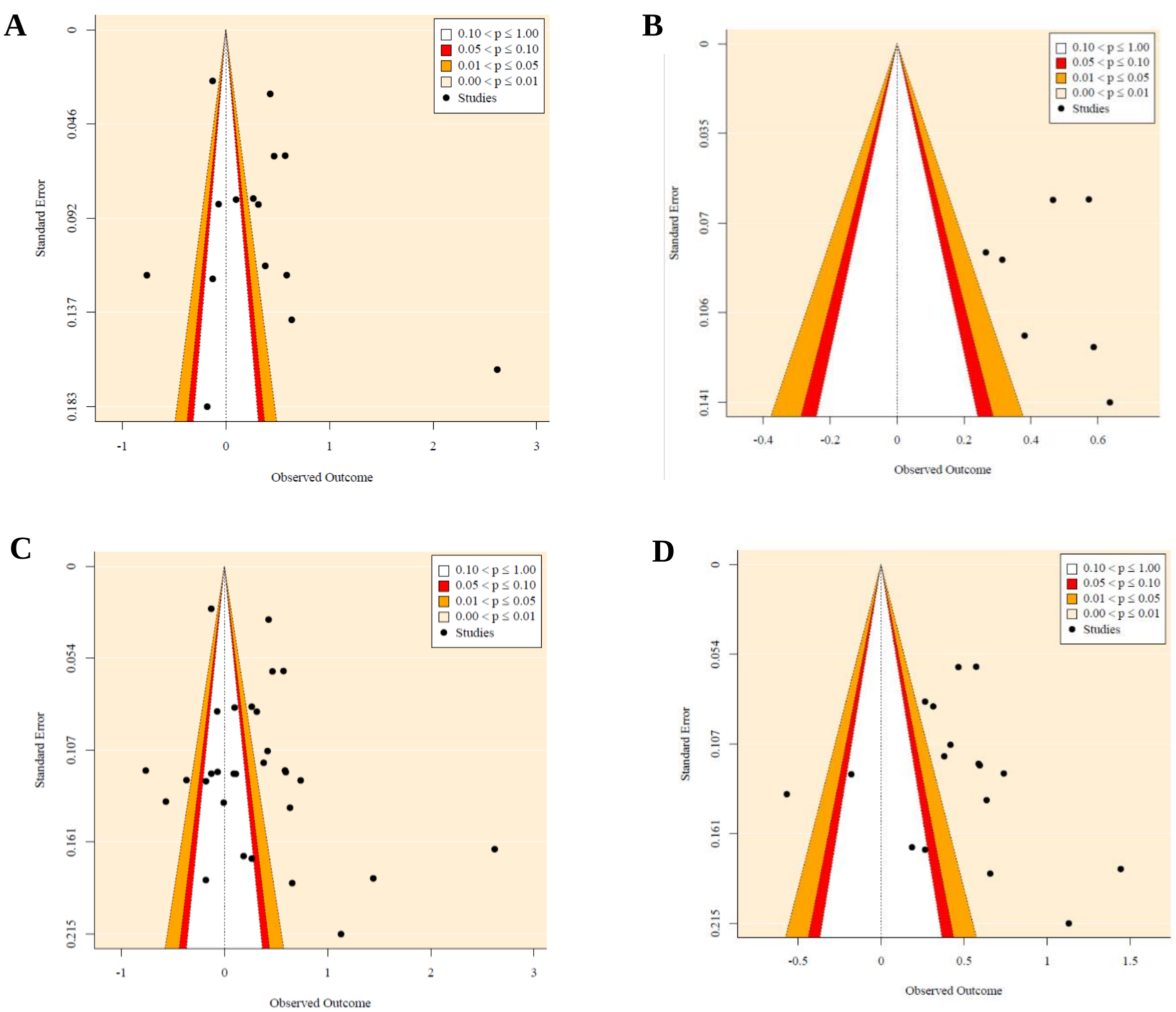
The regression test was not significant (2.62, z=.97, p>.05).

Imputed-redux

The regression test was not significant (3.03, z=1.23, p>.05).

Figure AF1

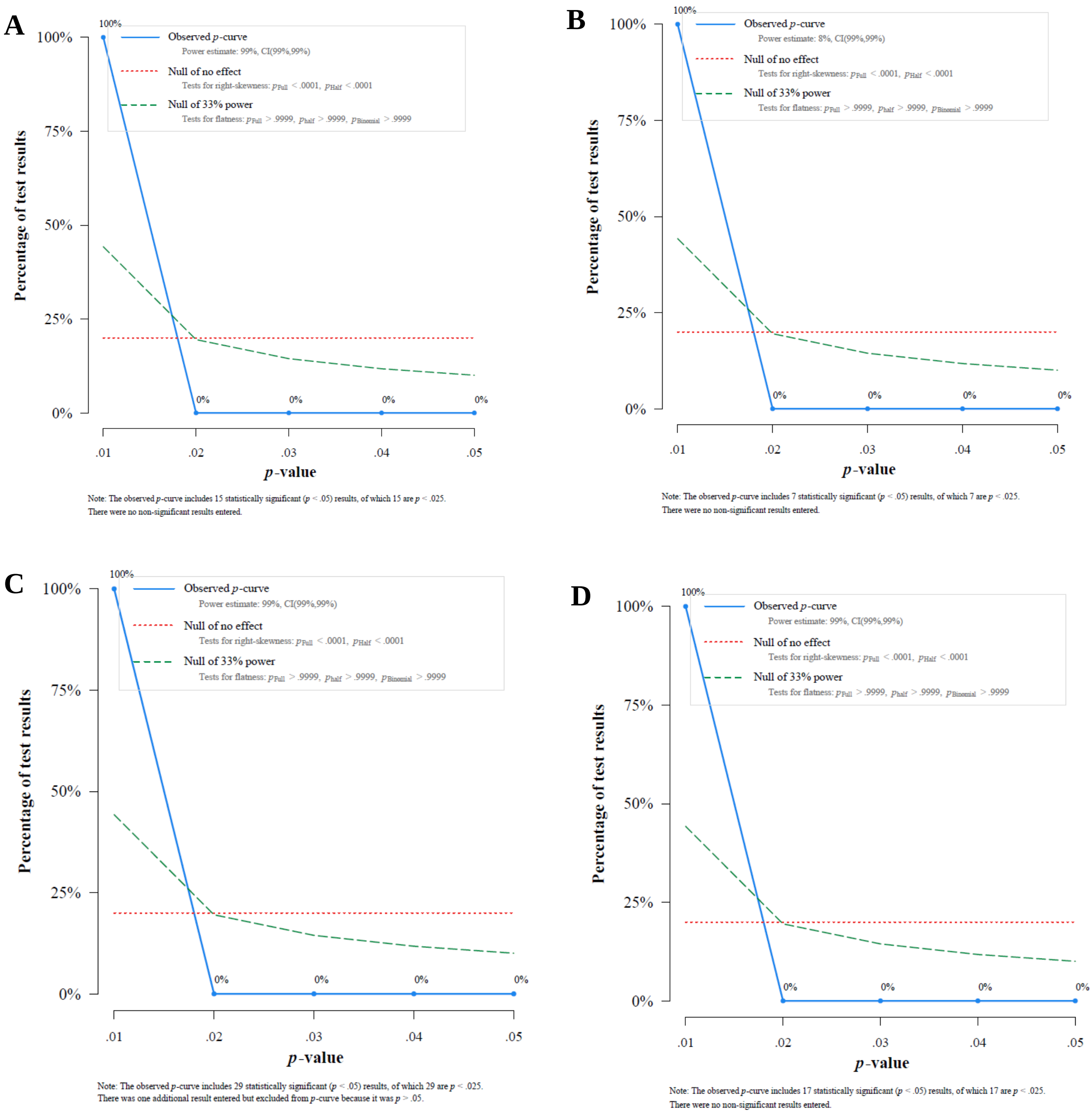
Funnel Plots for Response Bias Before and After Influential Cases and Outliers were Removed



Note. (AF1-A) Funnel plot testing for publication bias using reported data – 'original-all' model. (AF1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model. (AF1-C) Funnel plot testing for publication bias using completed data – 'imputed-all' model. (AF1-D) Funnel plot testing for publication bias using completed data after influential cases and outliers were removed – 'imputed-redux' model

Figure AF2

P-curve Analysis for Response Bias Before and After Influential Cases and Outliers were Removed



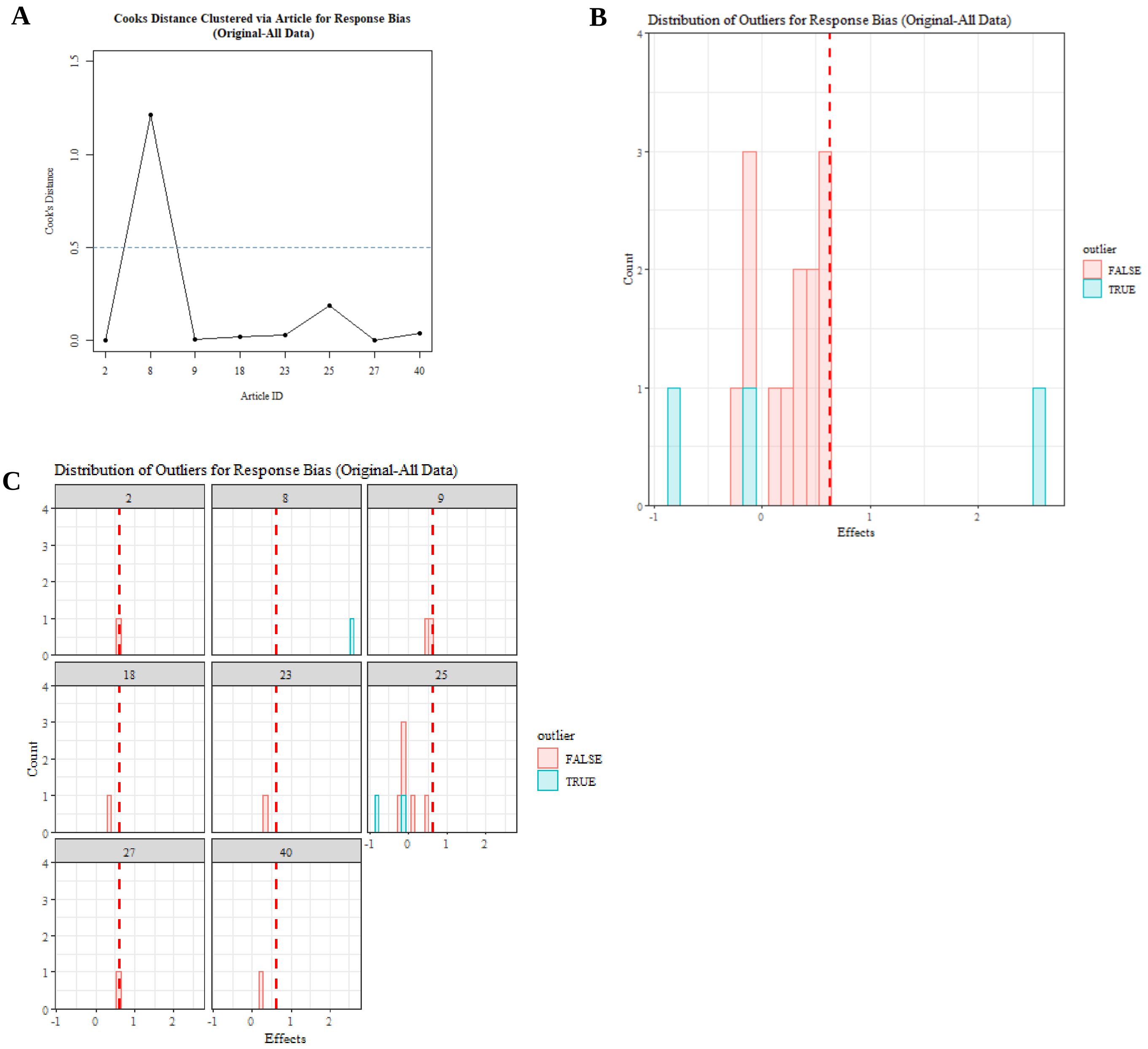
Note. (AF2-A) P-curve plot testing for publication bias using reported data – ‘original-all’ model. (AF2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model. (AF2-C) P-curve plot testing for publication bias using completed data – ‘imputed-all’ model. (AF2-D) P-curve plot testing for publication bias using completed data after influential cases and outliers were removed – ‘imputed-redux’ model

Appendix AG

OENE-Response Bias: Diagnostic Plots

Figure AG1

Diagnostic Plots for Reported Data – 'original-all' Model



Note. (AG1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AG1-B) Distribution of effect size outliers. (AG1-C) Distribution of effect size outliers faceted by article ID.

Figure AG2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model

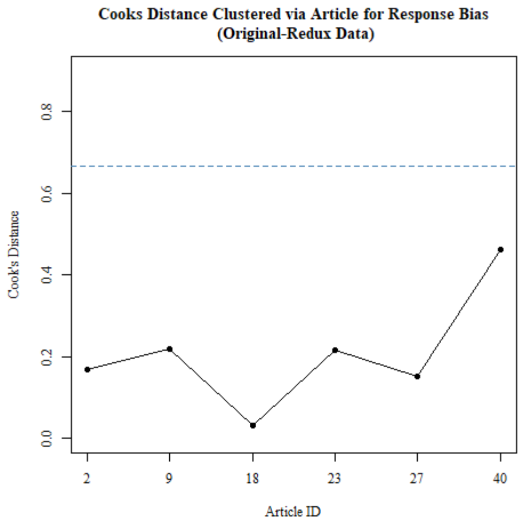
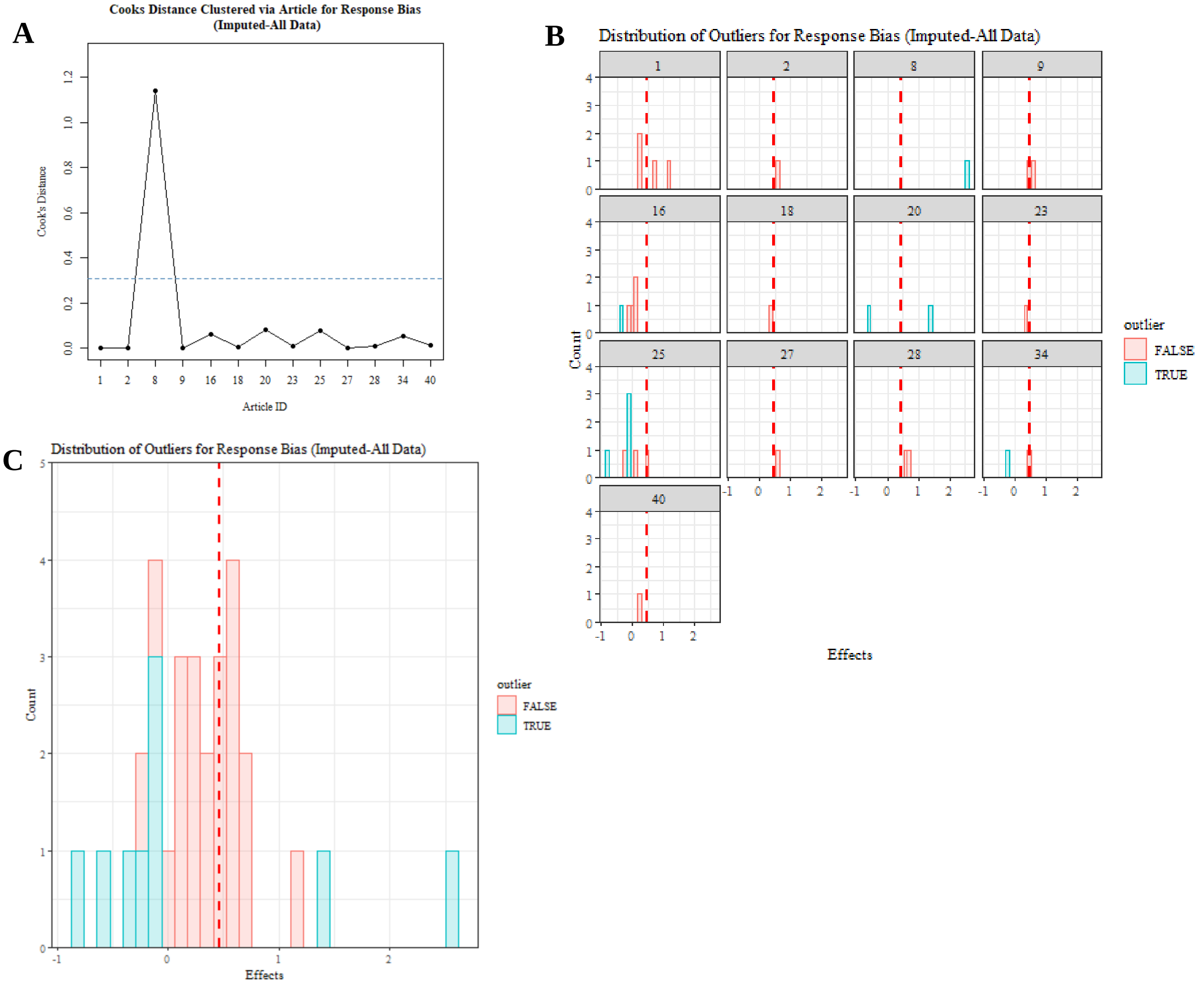


Figure AG3

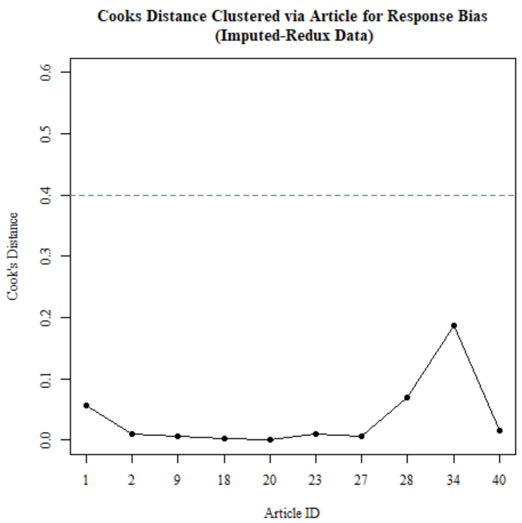
Diagnostic Plots for Completed Data – 'imputed-all' Model



Note. (AG3-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AG3-B) Distribution of effect size outliers faceted by article ID. (AG3-C) Distribution of effect size outliers.

Figure AG4

Cooks Distance for Completed Data after Influential Cases and Outliers were Removed – 'imputed-redux' Model



Appendix AH

OENE Validation – Discriminability (d-prime)

Descriptive Statistics

The majority of the sample used d prime as opposed to A prime to measure discrimination (Original-all: 78.21%, Imputed-all: 80.68%). There is descriptive support for the presence of an OENE for discrimination as the majority of the sample exhibited higher ingroup discrimination values when compared to outgroup values (Original-all: 88.46%, Imputed-all: 89.77%). The average mean difference for discrimination was sizable

Table AH1

Descriptives Table for Overall Identification Accuracy (Discriminability)

		Ingroup	Ingroup SD	Outgroup	Outgroup SD	Mean difference
Original All	Ingroup D > Outgroup D	1.34	0.58	1.03	0.56	0.31
	Ingroup D < Outgroup D	1.03	0.81	1.31	0.62	-0.27
	Overall	1.31	0.62	1.06	0.57	0.25
Imputed All	Ingroup D > Outgroup D	1.35	0.61	1.01	0.62	0.34
	Ingroup D < Outgroup D	1.03	0.81	1.31	0.62	-0.27
	Overall	1.32	0.64	1.04	0.62	0.28

Meta-Analysis

The aggregate effect (*SMD*), for all datasets, confirm the presence of a statistically significant OENE for discrimination ($p<.05$). Hypothesis 1c is therefore supported.

Original

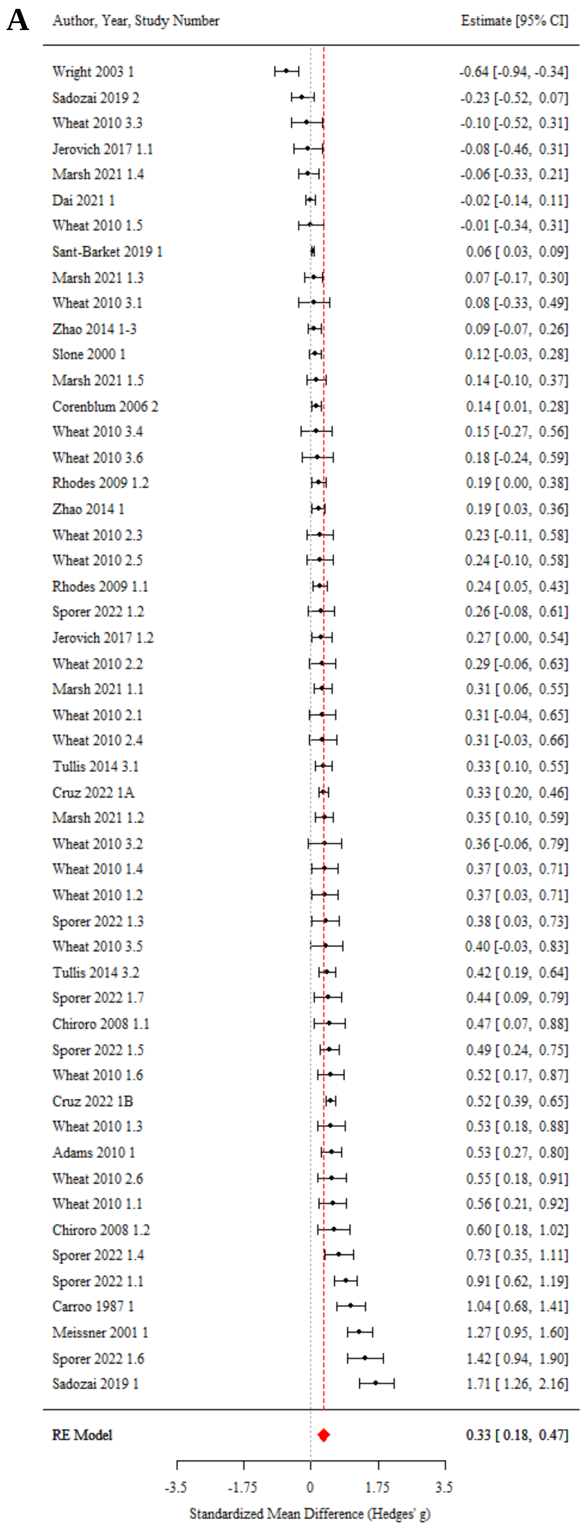
The strength of the aggregate effect size for '-all' was classified as medium-to-large, whilst (-redux' was classified as small-to-medium (g=.52 & g= .33 respectively).

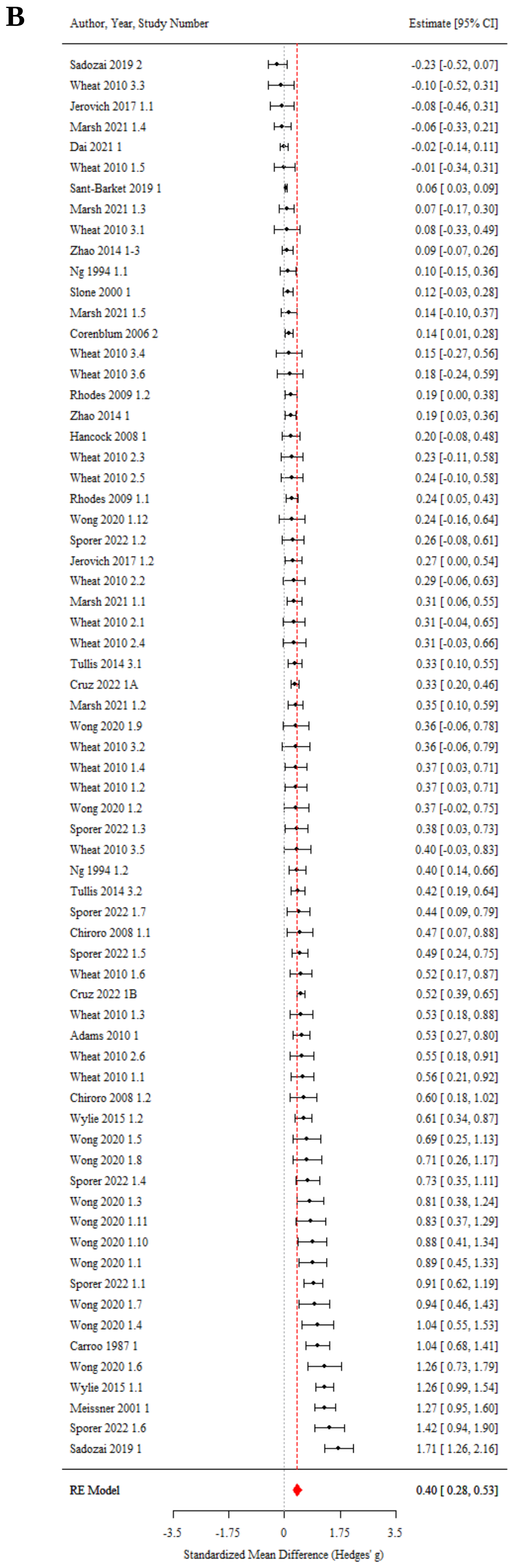
Imputed

The aggregate effect size for both '-all' was classified as medium-to-large and the aggregate effect for 'redux' was classified as small-to-medium (g=.69, g=.40 respectively)

Figure AH1

Forest Plots for Overall Identification Accuracy (Discriminability) after Influential Cases and Outliers were Removed





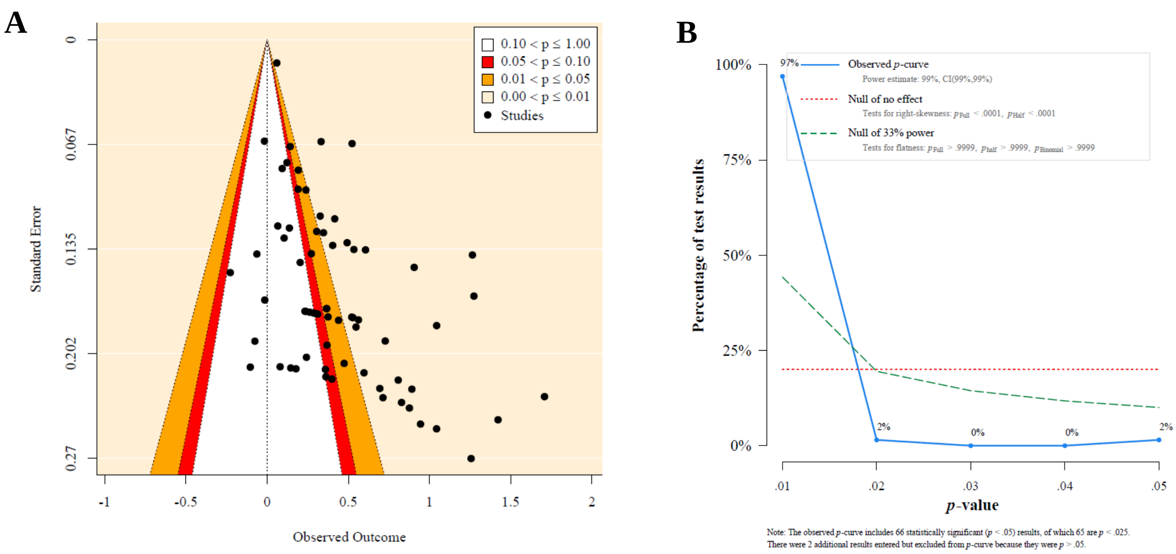
Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (AH1-A) Forest plot for original or actual data after the removal of outliers and influential cases i.e. 'original-redux' model. (AH1-B) Forest plot using the complete dataset after outliers and influential cases were removed – standard deviations were completed via multiple imputation i.e. 'imputed-redux' model.

Publication bias

Publication bias was present in the sample

Figure AH2

Diagnostic Plots for Overall Accuracy (Discriminability) in the Completed Dataset after Outliers and Influential Cases were Removed



Note. (AH2-A) Funnel Plot testing for publication bias in the 'imputed-redux' model. (AH2-B) P-curve analysis testing for publication bias in the 'imputed-redux' model.

Moderators

Implicit Outgroup Prejudice

Combined Option 1. Higher implicit outgroup prejudice increases the observed OENE for discrimination (Original-all: 1.30, z=2.61, p<.01). This is the expected direction and the inclusion thereof accounted for 28.13% of the total variance explained.

Merged. The same pattern was observed (Original-all: 1.03, z=2.39, p<.05, 17.05%; Original-redux: .53, z=2.17, p<.05, 24.64%).

Merged completed. The same pattern was observed (Original-all: 1.52, z-2.01, p<.05)

Quantity of outgroup contact

Reported. Higher outgroup contact decreases the observed OENE for discrimination (Original-all: -2.46, z=-4.35, p<.001; Original-redux: -1.22, z=-3.37, p<.001; Imputed-all: -2.44, z=-4.48, p<.001; Imputed-redux: -1.31, z=-3.25, p<.01). This is consistent with the expected pattern of results and the importance of accounting for outgroup contact is underscored by the large proportional reduction in total variance (30.43%, 70.30%, 68.27%; and 34.85% respectively)

Completed. The same pattern was observed (Original-all: -1.80, z=-3.64, p<.001; Imputed-all: -2.83, z=-4.91, p<.001; Imputed-redux: -.76, z=-2.23, p<.05).

Quantity of outgroup contact time bands

Outgroup contact during adulthood only increases the observed OENE for discrimination (Original-all: .73, z=2.41, p<.05). This is consistent with the expected direction.

Explicit Outgroup Prejudice

Merged. More favourable explicit outgroup attitudes, or lower explicit outgroup prejudice, increases the observed OENE for discrimination (Original-redux: .99, z=2.01, p<.05). This is not the expected pattern of results wherein, lower explicit prejudice should be associated with a reduction in the OENE.

Quality of outgroup contact time bands

Quality of outgroup contact across the lifespan should be associated with a reduction the size of the OENE. Within the sample, quality of outgroup contact across the lifespan was associated with an increase in the observed OENE for discrimination (Original-redux: .44, z=3.80, p<.001; Imputed-redux: .46, z=4.62, p<.001). This is counter to the expected direction.

Length of delay (categorical)

Consistent with expectations, a delay between 2 and 15 minutes increases the size of the OENE for discrimination (Original-all: .52, z=2.80, p<.01; Original-redux: .29, z=2.45, p<.05; Imputed-all: .88, z=4.18, p<.001; Imputed-redux: .48, z=5.41, p<.001).

Type of encoding

A basic level of face-processing during the identification task is associated with an increase in the observed OENE for discrimination

(Original-all: .55, z=3.46, p<.001; Original-redux: .30, z=2.94, p<.01; Imputed-all: .68, z=2.54, p<.05; Imputed-redux: .41, z=4.71, p<.001). This is consistent with the expected pattern of results.

Positionality of participants relative to outgroup members

Majority members tested on minority outgroup members exhibited a higher OENE for discrimination, consistent with expected pattern of results (Original-all: .51, z=4.20, p<.001; Original-redux: .33, z=4.78, p<.001; Imputed-all: .68, z=3.14, p<.01; Imputed-redux: .44, z=5.97, p<.001). By comparison, minority members that were tested on majority outgroup members exhibited a lower OENE for discrimination (Original-all: -.40, z=-2.34, p<.05).

Task/Cognitive demands

High task and/or cognitive demand followed the expected direction in that harder tasks increased the observed OENE for discrimination (Original-all: .51, z=3.88, p<.001; Original-redux: .29, z=3.66, p<.001; Imputed-all: .70, z=4.18, p<.001; Imputed-redux: .37, z=5.19, p<.001)

Sample country's positionality

Countries belonging to the 'Global North' exhibited a higher OENE for discrimination (Original-all: .57, z=3.91, p<.001; Original-redux: .51, z=4.05, p<.001; Imputed-all: .77, z=4.26, p<.001; Imputed-redux: .48, z=5.04, p<.001), while those belonging to the 'Global South' exhibited a lower OENE for discrimination (Original-redux: -1.18, z=-4.85, p<.001). This is consistent with expectations.

Motivation

No motivation manipulation or instructions increased the observed OENE for discrimination (Original-all: .55, z=4.06, p<.001; Original-redux: .33, z=3.78, p<.001; Imputed-all: .72, z=4.26, p<.001; Imputed-redux: .42, z=5.78, p<.001). This follows the expected pattern of results.

Task

Alternate forced choice (AFC) tasks increased the OENE for discrimination (Original-all: 2.71, z=3.87, p<.001). When compared to AFC tasks, both delayed matching (Original-all: -2.18, z=-2.92, p<.01) and Old-New tasks (Original-all: -2.27, z=-3.19, p<.01) exhibited a lower OENE for discrimination.

In samples without AFC tasks, delayed matching increased the OENE for discrimination (Original-redux: .46, z=3.07, p<.01)

Publication

Counter to expectations, non-published studies exhibited a larger OENE for discrimination (Original-all: .58, z=2.34, p<.05; Original-redux: .31, z=2.04, p<.05; Imputed-all: .69, z=2.11, p<.05; Imputed-redux: .30, z=2.22, p<.05).

Length of encoding (categorical)

A brief encoding, of between 0 to 2 seconds, increased the OENE for discrimination (Original-redux: .71, z=3.89, p<.001; Imputed-all: 1.27, z=3.02, p<.01; Imputed-redux: .63, z=3.26, p<.01). By comparison, a short encoding, of between 3 to 8 seconds, decreased the observed OENE for discrimination (Original-redux: -.47, z=-2.33, p<.05).

Encoding instructions

When the encoding instructions given to participants are either standard or incidental, i.e. they are unaware of a future memory test, as opposed to being directed encoding instructions, the observed OENE for discrimination increases (Imputed-all: 6.04, z=3.31, p<.001).

This is the expected direction as directed instructions which lead to a deeper level of face-processing should decrease the OENE more than either standard or incidental instructions.

Year

Year of study is used as a proxy for year the data was collected. In other words when contact, prejudice and the identification task occurred. Although global connectivity has increased and with it outgroup contact, it was argued that changes in time by itself should not result in an observed decrement in the observed OENE. Changes in time reducing the OENE for discrimination was therefore counter to expectations (Imputed-all: -.04, z=-3.13, p<.01).

Table AH2

Significant Moderators for Overall Accuracy (Discriminability) Across all Model

			Model 4 (Original-all)												Model 3 (Original-redux)												Model 2 (Imputed-all)												Model 1 (Imputed-redux)											
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Implicit prejudice for outgroup members (combined option one)	Core	-	51	18	1.30	Yes	2.61	0.009	0.32	2.27	**	28.13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Havard's 'Project Implicit' values for outgroup implicit prejudice	Core	-	45	18	1.03	Yes	2.39	0.017	0.19	1.88	*	17.05	32	15	0.53	Yes	2.17	0.030	0.05	1.00	*	24.64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Havard's 'Project Implicit' values for outgroup explicit prejudice (completed)	Core	-	78	25	1.52	Yes	2.01	0.044	0.04	3.01	*	8.04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Quantity of outgroup contact (reported)	Core	-	50	8	-2.46	Yes	-4.35	<.0001	-3.57	-1.35	***	30.43	31	6	-1.22	Yes	-3.37	0.001	-1.93	-0.51	***	70.30	53	10	-2.44	Yes	-4.48	<.0001	-3.50	-1.37	***	68.27	46	9	-1.31	Yes	-3.25	0.001	-2.10	-0.52	**	34.85
Quantity of outgroup contact (completed)	Core	-	78	25	-1.80	Yes	-3.64	0.000	-2.78	-0.83	***	7.82	-	-	-	-	-	-	-	-	-	-	88	31	-2.83	Yes	-4.91	<.0001	-3.96	-1.70	***	30.18	68	22	-0.76	Yes	-2.23	0.026	-1.42	-0.09	*	15.64
Quantity of outgroup contact time bands (multiple)	Core	Adult (18+)	78	25	0.73	Yes	2.41	0.016	0.14	1.33	*	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Havard's 'Project Implicit' values for outgroup explicit prejudice	Core	-	-	-	-	-	-	-	-	-	-	-	52	19	0.99	No	2.01	0.045	0.02	1.96	*	EMV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Quality of outgroup contact time bands (multiple)	Core	Life span	-	-	-	-	-	-	-	-	-	-	34	8	0.44	No	3.80	0.000	0.21	0.66	***	EMV	-	-	-	-	-	-	-	-	-	-	48	10	0.46	No	4.62	<.0001	0.27	0.66	***	EMV
Length of delay (categories)	Methodological / Task	Brief	78	25	0.52	Yes	2.80	0.005	0.16	0.88	**	EMV	52	19	0.29	Yes	2.45	0.014	0.06	0.52	*	EMV	88	31	0.88	Yes	4.18	<.0001	0.47	1.29	***	1.14	68	22	0.48	Yes	5.41	<.0001	0.31	0.65	***	6.62
Type of encoding	Methodological / Task	Basic	55	23	0.55	Yes	3.46	0.001	0.24	0.87	***	EMV	29	17	0.30	Yes	2.94	0.003	0.10	0.50	**	EMV	65	29	0.68	Yes	2.54	0.011	0.16	1.21	*	EMV	45	20	0.41	Yes	4.71	<.0001	0.24	0.58	***	EMV
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Majority-Minority	63	20	0.51	Yes	4.20	<.0001	0.27	0.75	***	27.51	42	15	0.33	Yes	4.78	<.0001	0.19	0.46	***	30.44	72	25	0.68	Yes	3.14	0.002	0.26	1.11	**	EMV	55	17	0.44	Yes	5.97	<.0001	0.29	0.58	***	16.37
		Minority-Majority vs Majority-Minority			-0.40	Yes	-2.34	0.019	-0.73	-0.07	*		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Moderator	Category	if categorical, levels	Model 4 (Original-all) No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	Model 3 (Original-redux) No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Task / Cognitive Demand	Methodological / Task	High	78	25	0.51	Yes	3.88	0.000	0.25	0.77	***	EMV	52	19	0.29	Yes	3.66	0.000	0.14	0.45	***	EMV
Sample Country's global positionality	Methodological / Task	Global North	74	23	0.57	Yes	3.91	<.0001	0.28	0.85	***	EMV	50	17	0.51	Yes	4.05	<.0001	0.26	0.76	***	EMV
		Global South vs Global North	-	-	-	-	-	-	-	-	-	-			-1.18	Yes	-4.85	<.0001	-1.66	-0.71	***	
Presence of motivation manipulation or instructions	Methodological / Task	None	78	25	0.55	Yes	4.06	<.0001	0.28	0.81	***	EMV	52	19	0.33	Yes	3.78	0.000	0.16	0.50	***	EMV
Task	Methodological / Task	AFC	78	25	2.71	-	3.87	0.000	1.34	4.09	***	15.59	-	-	-	-	-	-	-	-	-	-
		Delayed Matching vs AFC			-2.18	-	-2.92	0.004	-3.64	-0.72	**		-	-	-	-	-	-	-	-	-	-
		Old-New vs AFC			-2.27	-	-3.19	0.001	-3.67	-0.88	**		-	-	-	-	-	-	-	-	-	-
		Delayed Matching	-	-	-	-	-	-	-	-	-	-	52	19	0.46	-	3.07	0.002	0.17	0.76	**	EMV
Published	Methodological / Task	No	78	25	0.58	No	2.34	0.019	0.09	1.07	*	EMV	52	19	0.31	No	2.04	0.041	0.01	0.01	*	EMV
Length of encoding (categories)	Methodological / Task	Brief	-	-	-	-	-	-	-	-	-	-	30	15	0.71	-	3.89	0.000	0.35	1.07	***	EMV
		Short vs Brief	-	-	-	-	-	-	-	-	-	-			-0.47	-	-2.33	0.020	-0.86	-0.07	*	
Encoding instructions	Methodological / Task	Standard / Incidental vs Directed	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Year	Methodological / Task	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Moderator	Category	if categorical, levels	Model 2 (Imputed-all) No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)	Model 1 (Imputed-redux) No. effects	No. studies	Est.	ED	z	p	LB	UB	Sig	Var (%)
Task / Cognitive Demand	Methodological / Task	High	88	31	0.70	Yes	4.18	<.0001	0.37	1.02	***	EMV	68	22	0.37	Yes	5.19	<.0001	0.23	0.50	***	1.05
Sample Country's global positionality	Methodological / Task	Global North	84	29	0.77	Yes	4.26	<.0001	0.42	1.12	***	EMV	66	20	0.48	Yes	5.04	<.0001	0.29	0.67	***	EMV
		Global South vs Global North	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Presence of motivation manipulation or instructions	Methodological / Task	None	88	31	0.72	Yes	4.26	<.0001	0.39	1.05	***	EMV	68	22	0.42	Yes	5.78	<.0001	0.28	0.56	***	EMV
Task	Methodological / Task	AFC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Delayed Matching vs AFC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Old-New vs AFC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Delayed Matching	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Published	Methodological / Task	No	88	31	0.69	No	2.11	0.035	0.05	1.32	*	EMV	68	22	0.30	No	2.22	0.026	0.04	0.57	*	0.13
Length of encoding (categories)	Methodological / Task	Brief	66	27	1.27	-	3.02	0.003	0.45	2.10	**	EMV	46	18	0.63	-	3.26	0.001	0.25	1.00	**	EMV
		Short vs Brief	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Encoding instructions	Methodological / Task	Standard / Incidental vs Directed	65	29	6.04	Yes	4.33	<.0001	3.31	8.77	***	EMV	-	-	-	-	-	-	-	-	-	-
Year	Methodological / Task	-	88	31	-0.04	No	-3.13	0.002	-0.07	-0.02	**	5.23	-	-	-	-	-	-	-	-	-	-

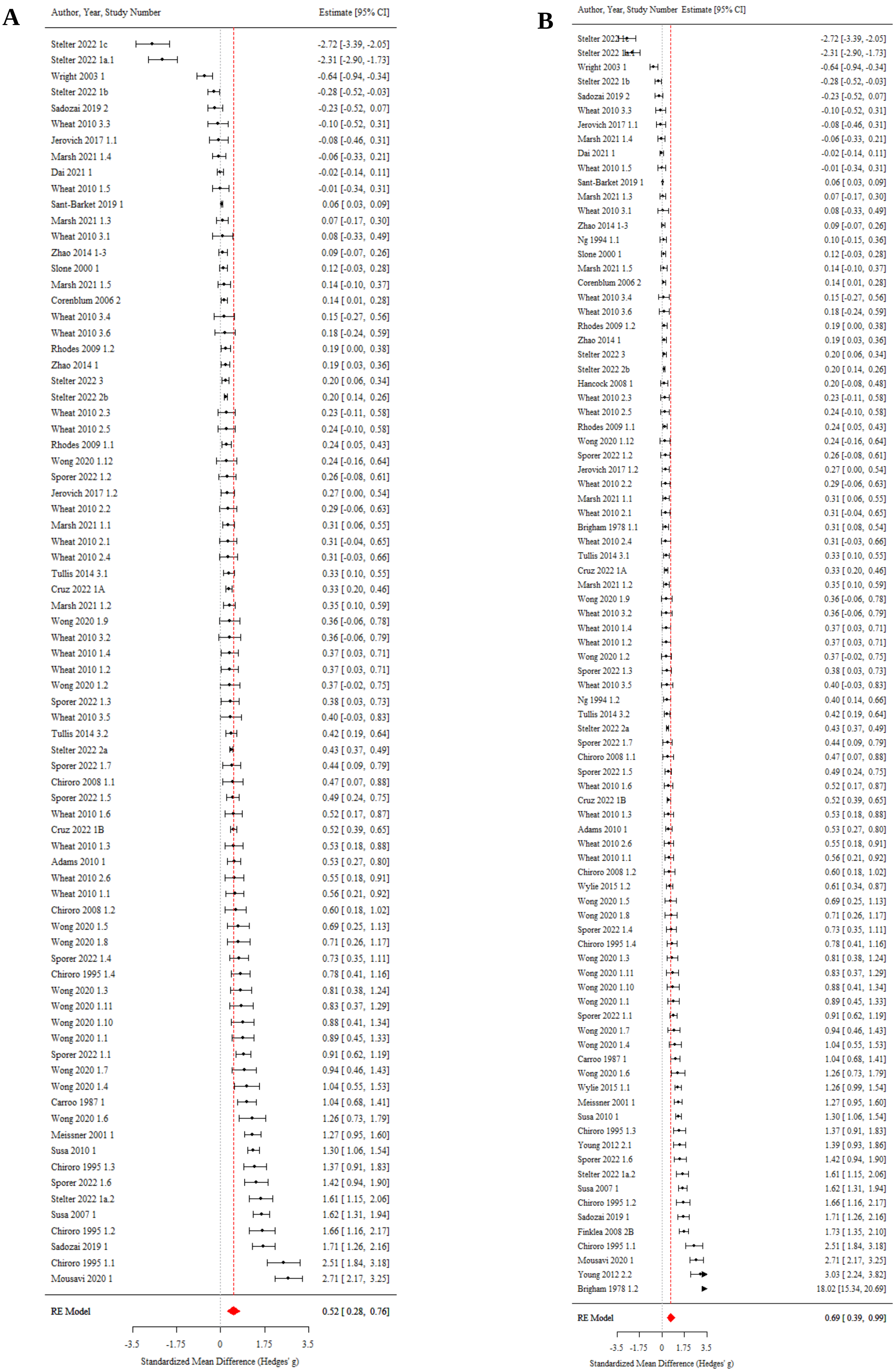
Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance

Appendix AI
OENE Validation – Discriminability (d-prime): Supplemental Forest Plots

Figure AI1

Forest Plots for Overall Accuracy (Discriminability)



Note. Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (AI1-A) Forest plot for original or actual data prior to removing influential cases or outliers – ‘original-all’ model. (AI1-B) Forest plot for data completed via imputation prior to removing influential cases or outliers – ‘imputed-all’ model.

Appendix AJ

OENE-Discriminability: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

The regression test was not significant (.47, z=.38, p>.05)

Original-redux

. The regression test was significant (3.37, z=2.92, p<.001)

Imputed-all

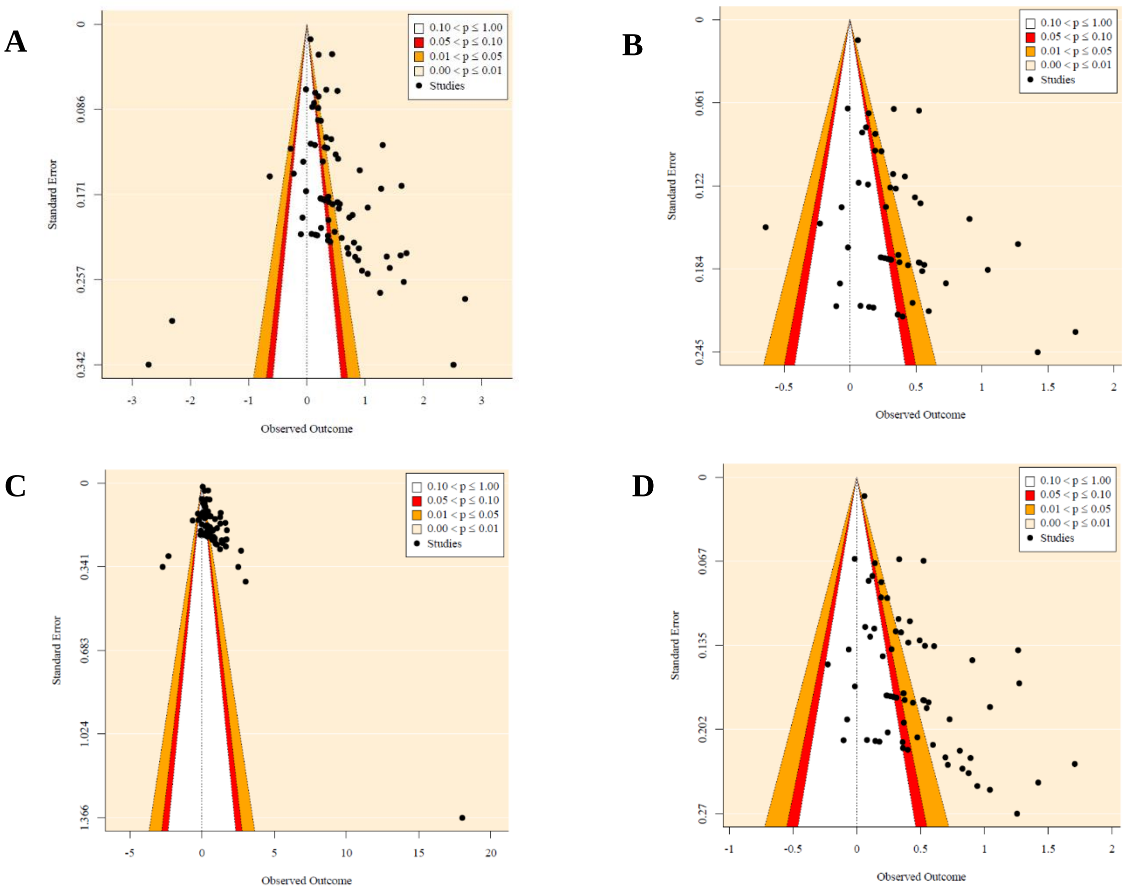
The regression test was significant (8.78, z=9.31, p<.001)

Imputed-redux

The regression test was significant (3.99, z=4.40, p<.001)

Figure AJ1

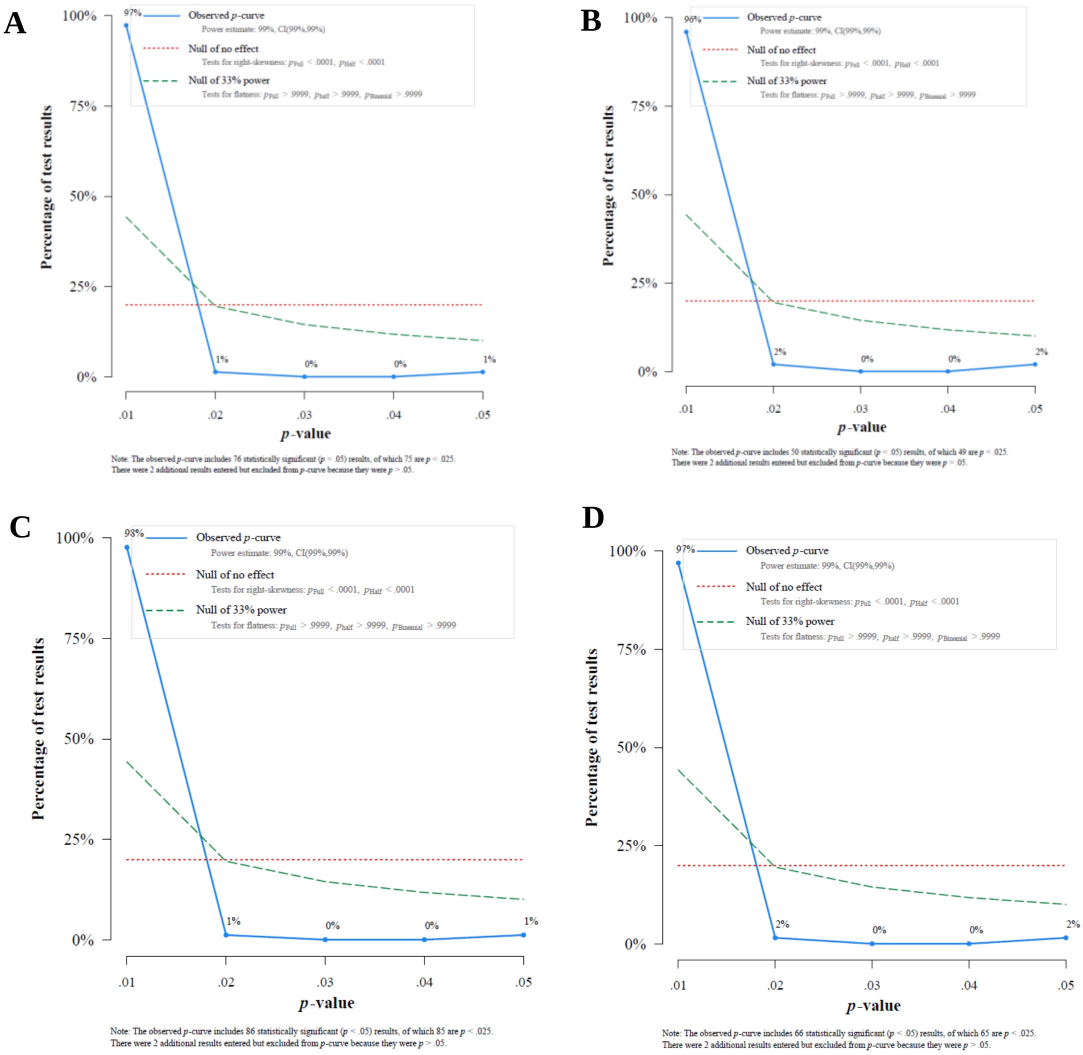
Funnel Plots for Overall Accuracy (discriminability) Before and After Influential Cases and Outliers were Removed



Note. (AJ1-A) Funnel plot testing for publication bias using reported data – 'original-all' model. (AJ1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model. (AJ1-C) Funnel plot testing for publication bias using completed data – 'imputed-all' model. (AJ1-D) Funnel plot testing for publication bias using completed data after influential cases and outliers were removed – 'imputed-redux' model

Figure AJ2

P-curve Analysis for Overall Accuracy (Discriminability) Before and After Influential Cases and Outliers were Removed



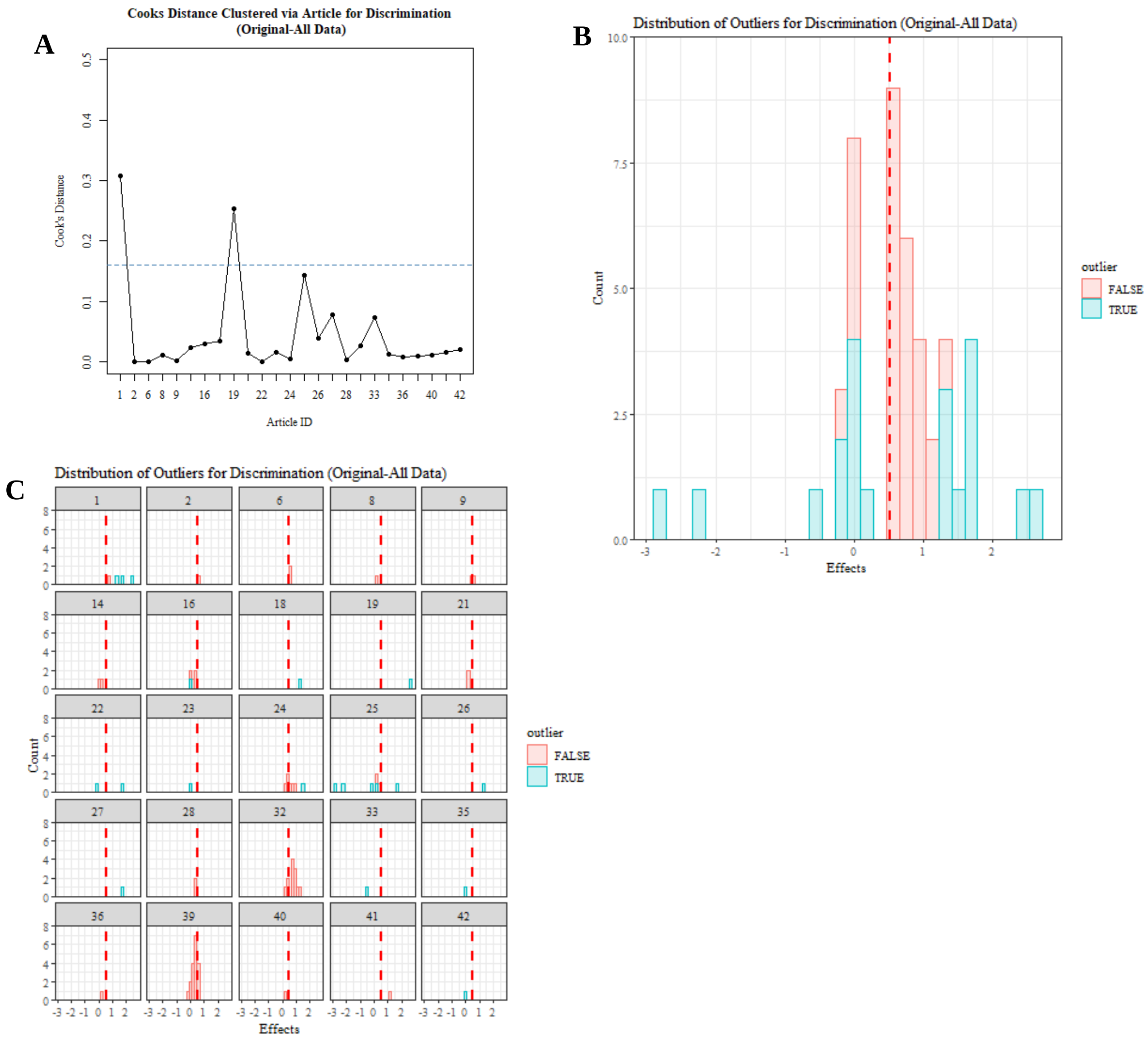
Note. (AJ2-A) P-curve plot testing for publication bias using reported data –'original-all' model. (AJ2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model. (AJ2-C) P-curve plot testing for publication bias using completed data – 'imputed-all' model. (AJ2-D) P-curve plot testing for publication bias using completed data after influential cases and outliers were removed – ‘imputed-redux’ model

Appendix AK

OENE-Discriminability: Diagnostic Plots

Figure AK1

Diagnostic Plots for Reported Data – 'original-all' Model



Note. (AK1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AK1-B) Distribution of effect size outliers. (AK1-C) Distribution of effect size outliers faceted by article ID.

Figure AK2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model

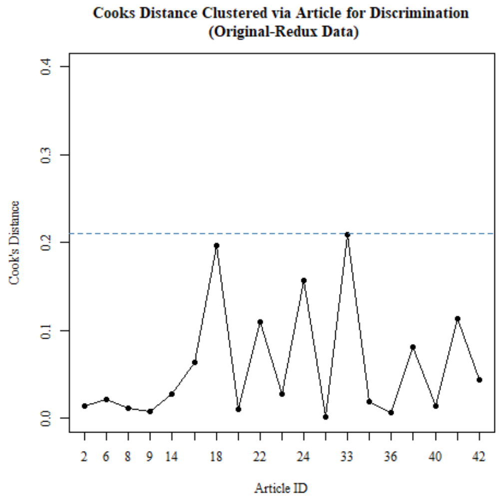
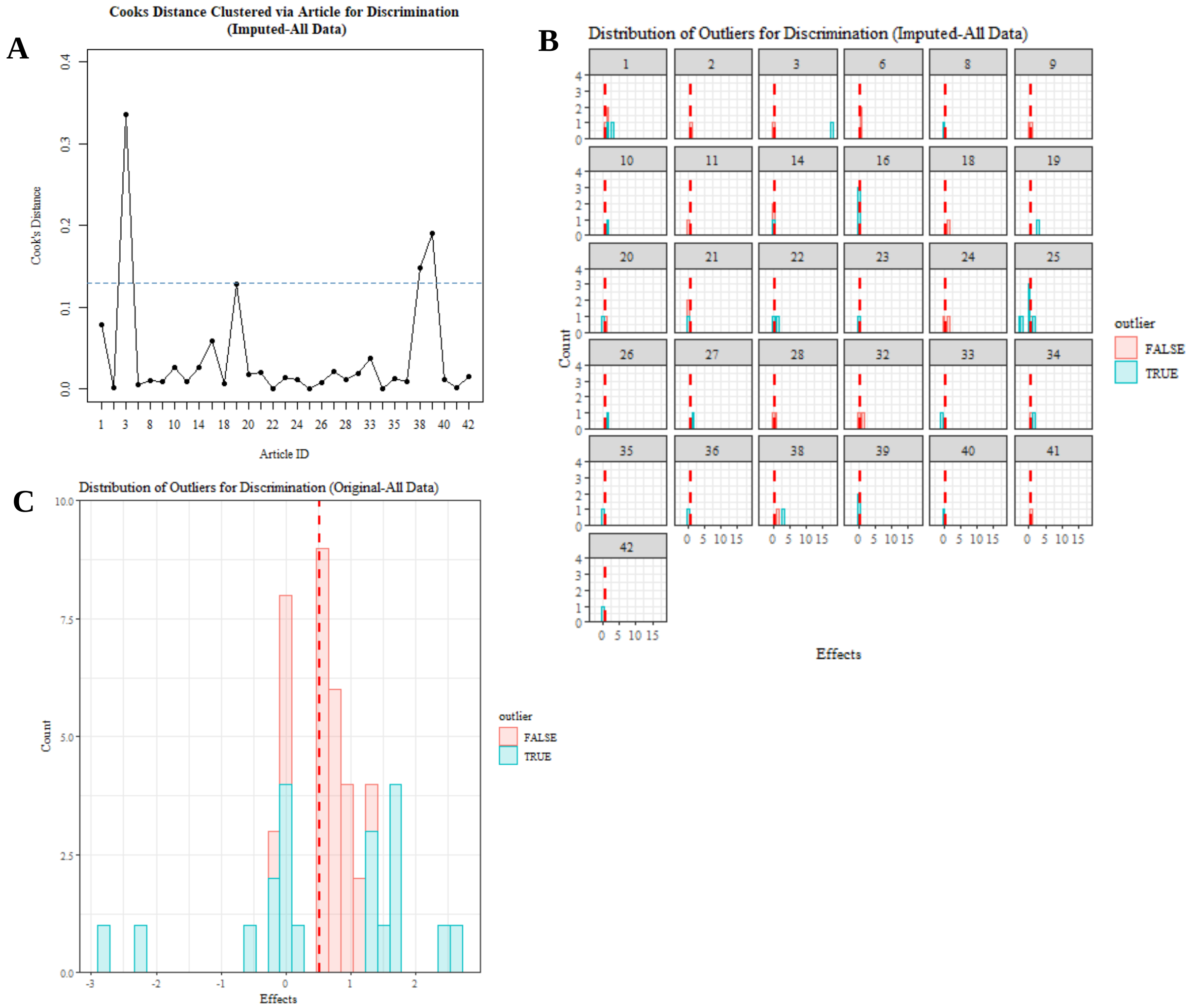


Figure AK3

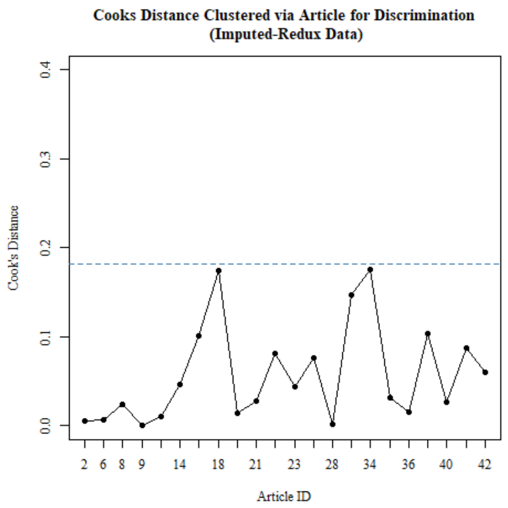
Diagnostic Plots for Completed Data – 'imputed-all' Model



Note. (AK3-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AK3-B) Distribution of effect size outliers faceted by article ID. (AK3-C) Distribution of effect size outliers.

Figure AK4

Cooks Distance for Completed Data after Influential Cases and Outliers were Removed – 'imputed-redux' Model



Appendix AL

Quantity of Outgroup Contact – Identification Difference Scores: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

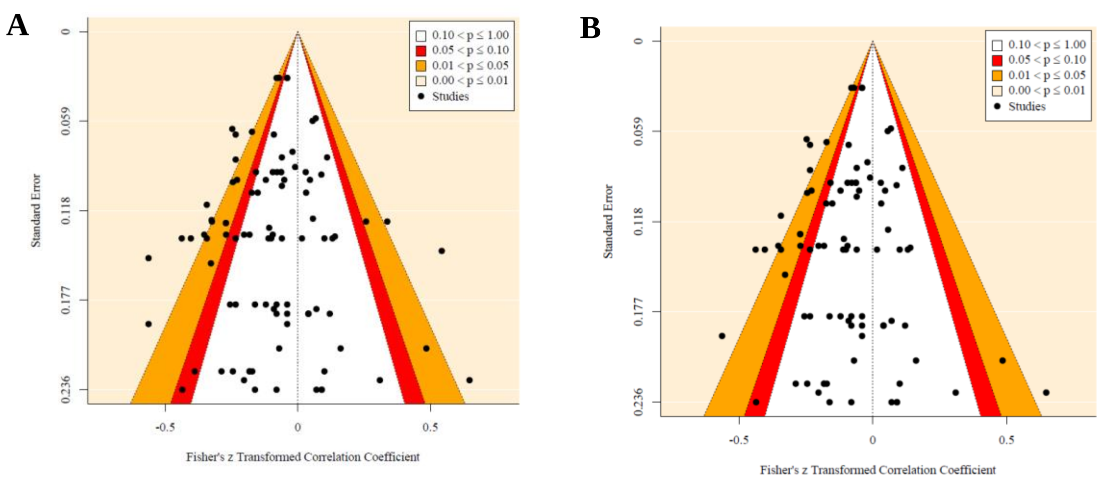
The regression test was not significant (.06, z=.13, p>.05)

Original-redux

. The regression test was not significant (.23, z=.61, p>.05)

Figure AL1

Funnel Plots for Quantity of Outgroup Contact – Identification Difference Scores Before and After Influential Cases were Removed

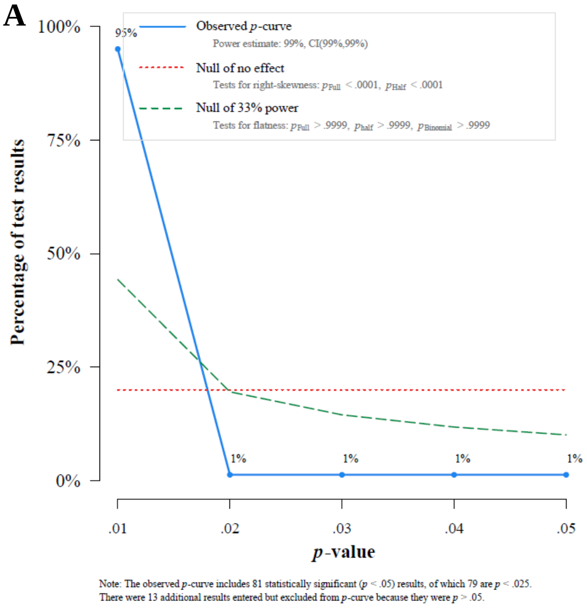


Note. (AL1-A) Funnel plot testing for publication bias using reported data – 'original-all' model. (AL1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model.

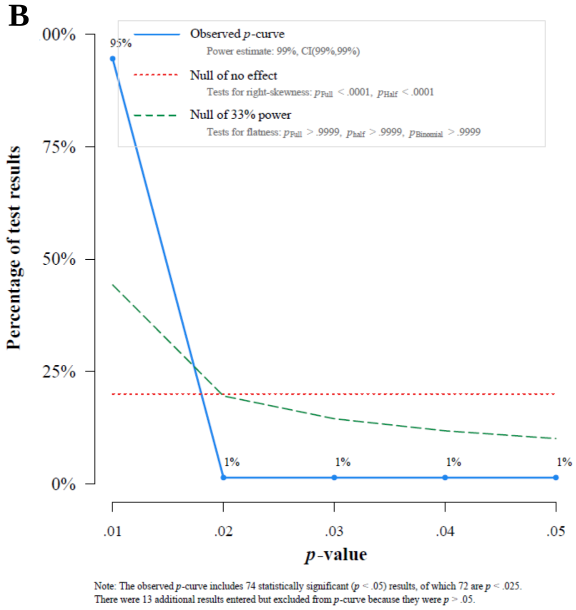
Figure AL2

P-curve Analysis for Quantity of Outgroup Contact – Identification Difference Scores Before and After Influential Cases and Outliers were Removed

A



B



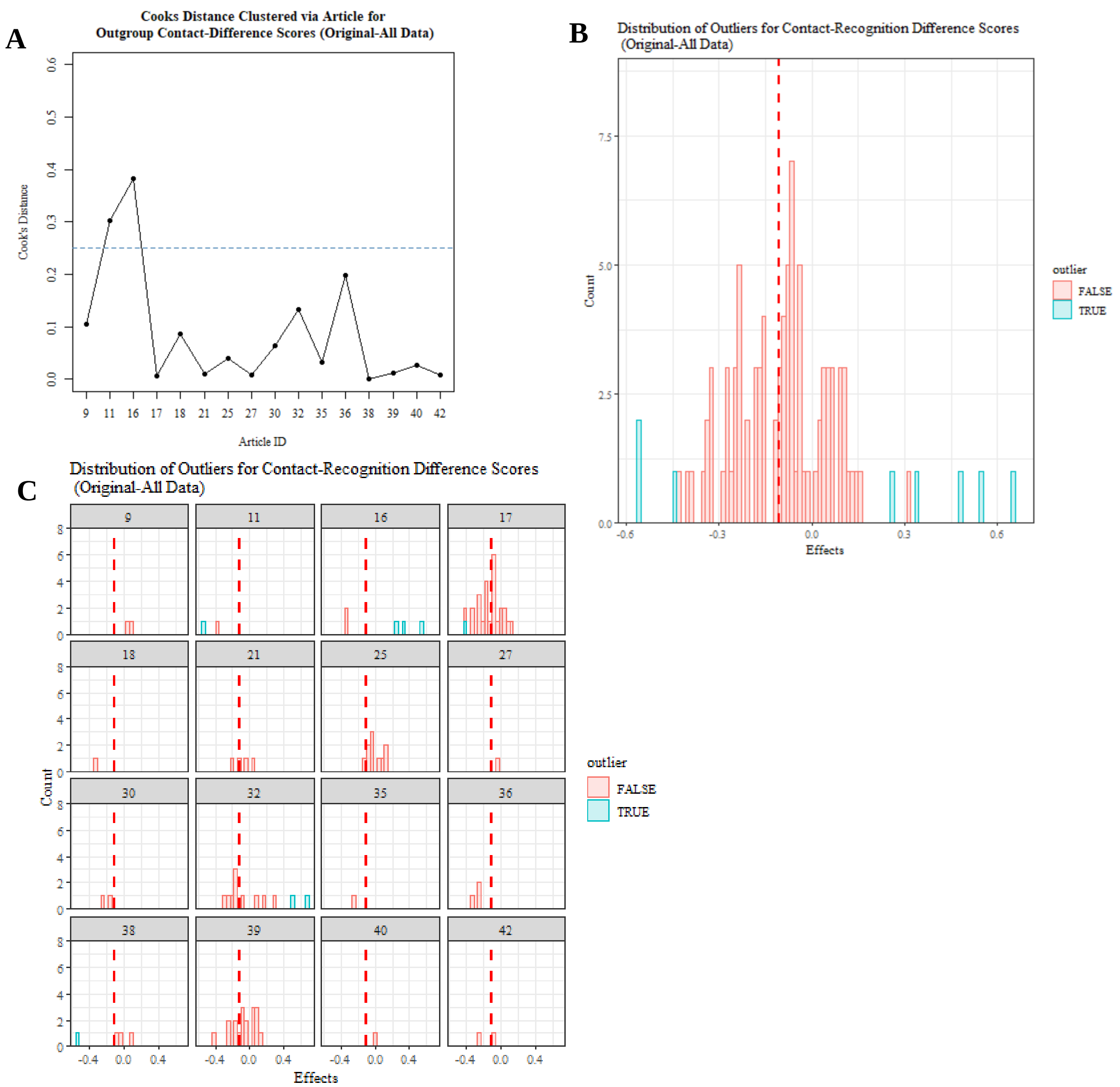
Note. (AL2-A) P-curve plot testing for publication bias using reported data – ‘original-all’ model. (AL2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model.

Appendix AM

Quantity of Outgroup Contact – Identification Difference Scores: Diagnostic Plots

Figure AM1

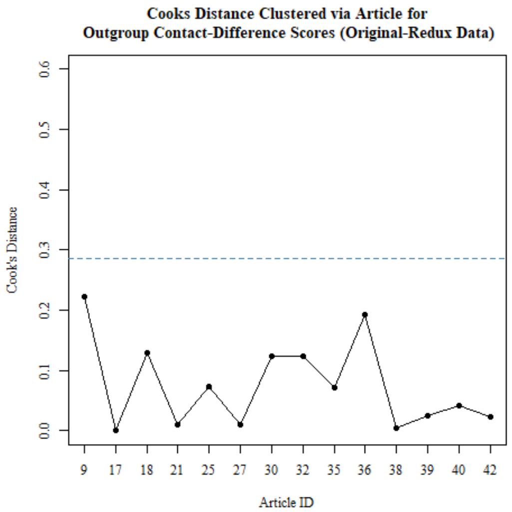
Diagnostic Plots for Reported Data – 'original-all' Model



Note. (AM1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AM1-B) Distribution of effect size outliers. (AM1-C) Distribution of effect size outliers faceted by article ID.

Figure AM2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model



Appendix AN

Quantity of Outgroup Contact – Outgroup Discriminability: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

The regression test was not significant (1.39, z=1.73, p>.05)

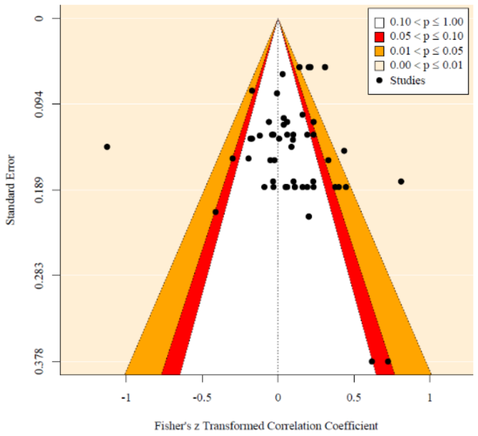
Original-redux

. The regression test was not significant (1.06, z=1.69, p>.05)

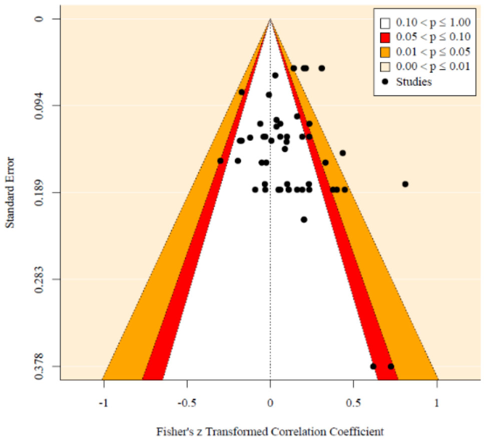
Figure AN1

Funnel Plots for Quantity of Outgroup Contact – Outgroup Discriminability Before and After Influential Cases were Removed

A



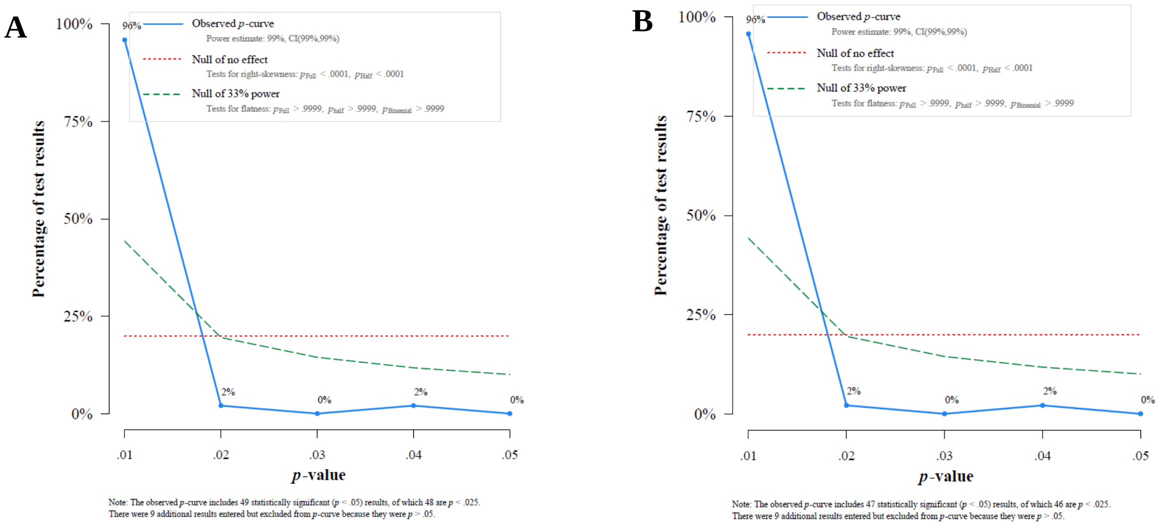
B



Note. (AN1-A) Funnel plot testing for publication bias using reported data – 'original-all' model. (AN1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model.

Figure AN2

P-curve Analysis for Quantity of Outgroup Contact – Outgroup Discriminability Before and After Influential Cases and Outliers were Removed



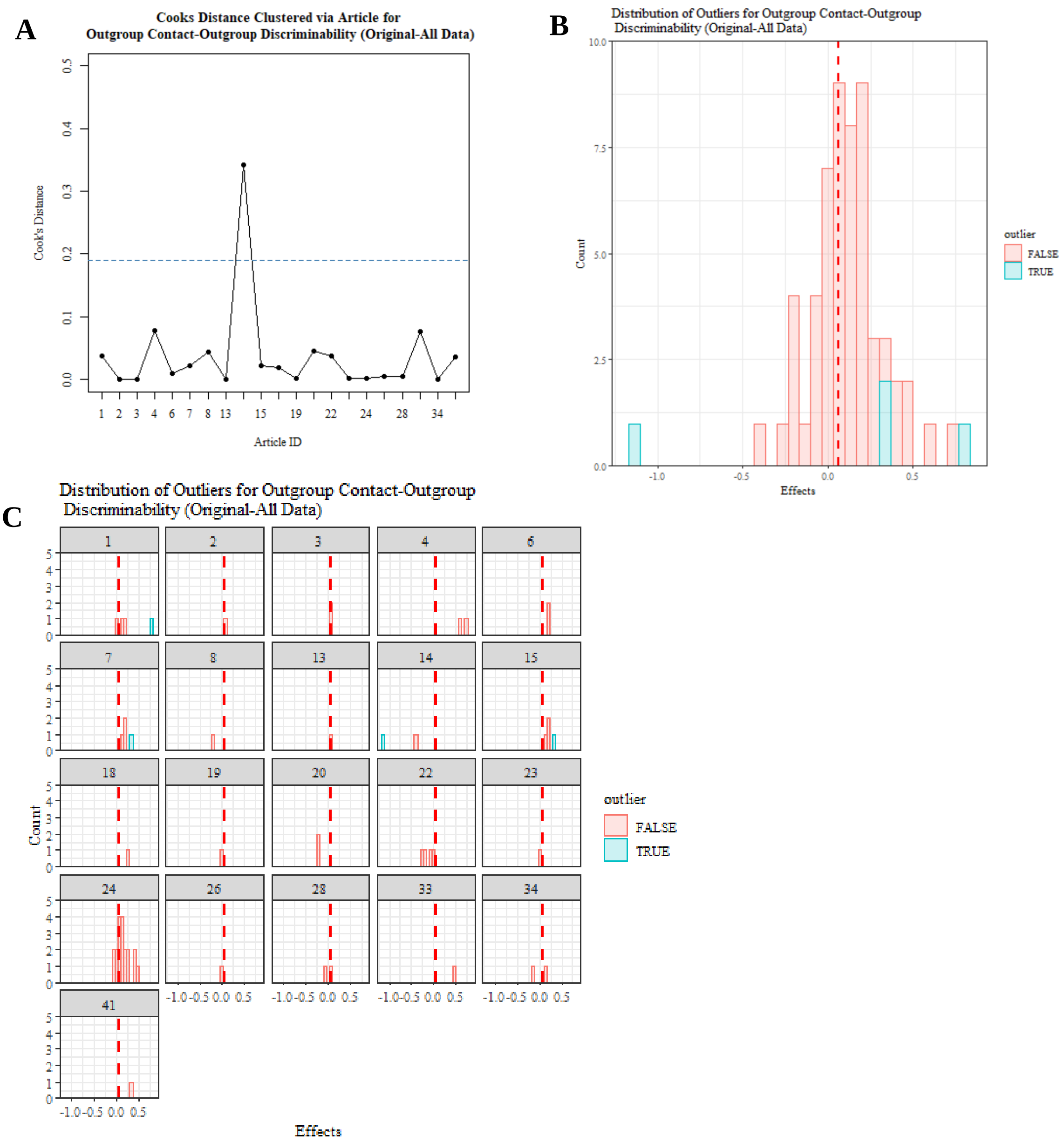
Note. (AN2-A) P-curve plot testing for publication bias using reported data – ‘original-all’ model. (AN2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model.

Appendix AO

Quantity of Outgroup Contact – Outgroup Discriminability: Diagnostic Plots

Figure AO1

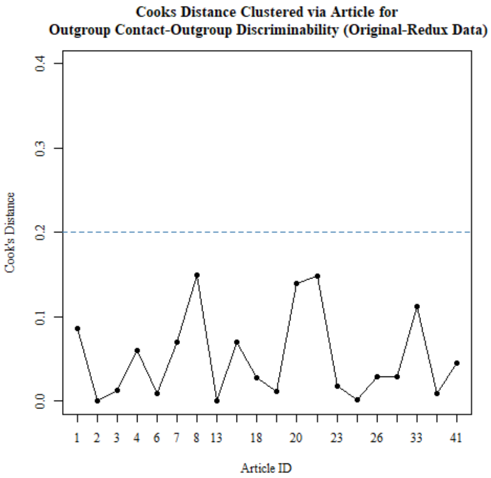
Diagnostic Plots for Reported Data – 'original-all' Model



Note. (AO1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AO1-B) Distribution of effect size outliers. (AO1-C) Distribution of effect size outliers faceted by article ID.

Figure AO2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model



Appendix AP

Quantity of Outgroup Contact – Outgroup Hits: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

The regression test was not significant (-.09, z=-.09, p>.05)

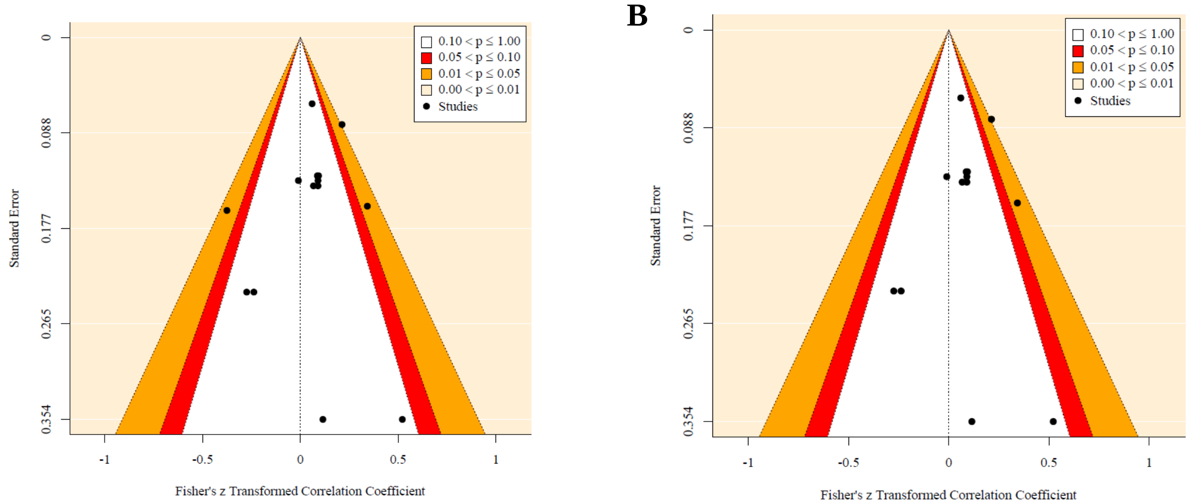
Original-redux

. The regression test was not significant (-.21, z=-.31, p>.05)

Figure AP1

Funnel Plots for Quantity of Outgroup Contact – Outgroup Hits Before and After Influential Cases were Removed

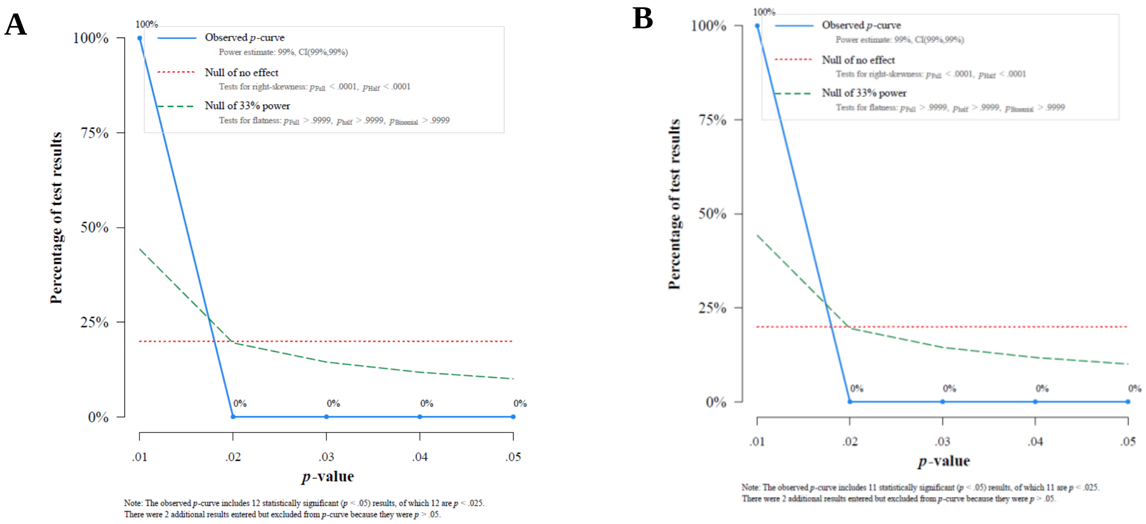
A B



Note. (AP1-A) Funnel plot testing for publication bias using reported data – 'original-all' model. (AP1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model.

Figure AP2

P-curve Analysis for Quantity of Outgroup Contact – Outgroup Hits Before and After Influential Cases and Outliers were Removed



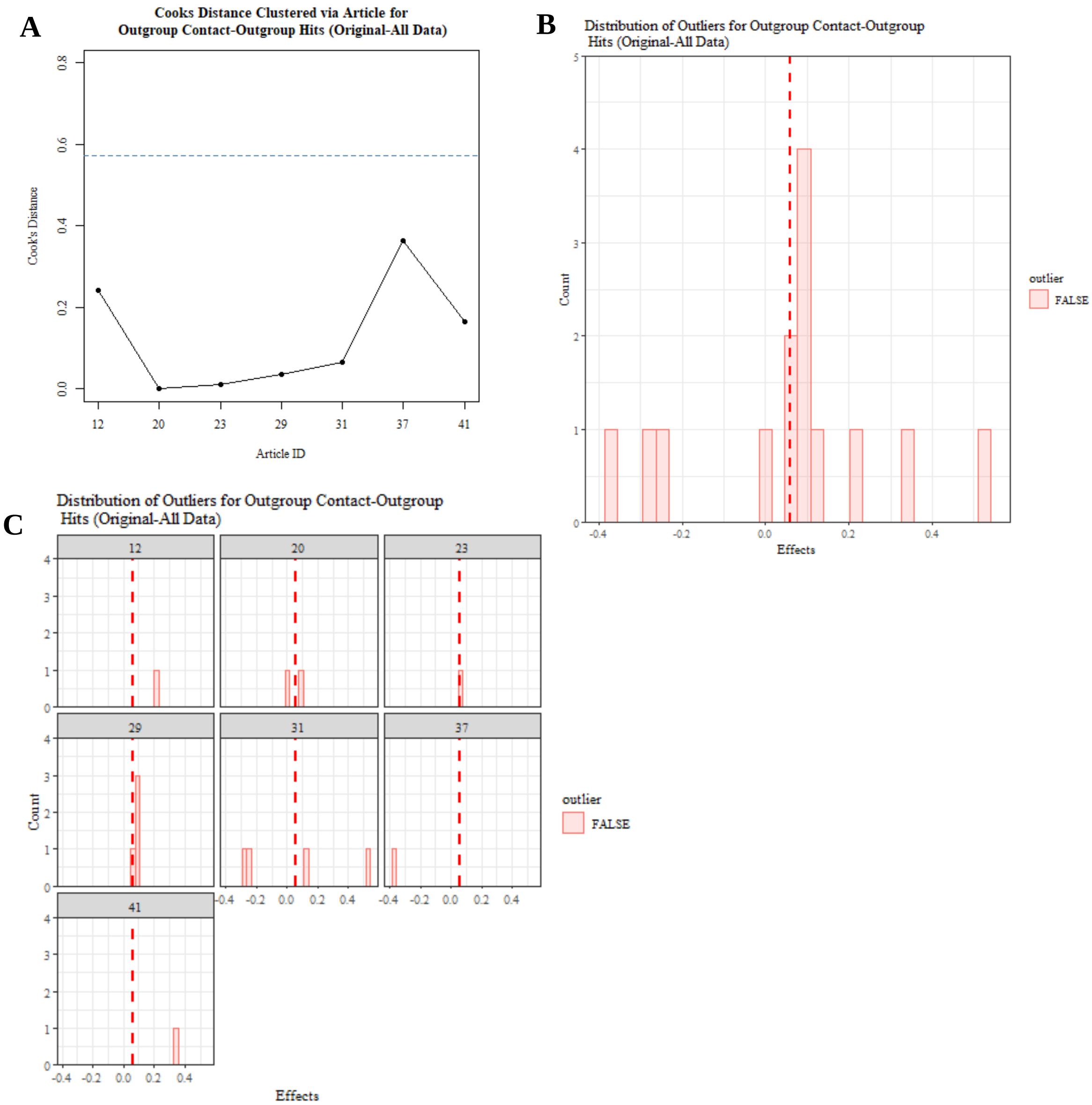
Note. (AP2-A) P-curve plot testing for publication bias using reported data – ‘original-all’ model. (AP2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model.

Appendix AQ

Quantity of Outgroup Contact – Outgroup Hits: Diagnostics Plots

Figure AQ1

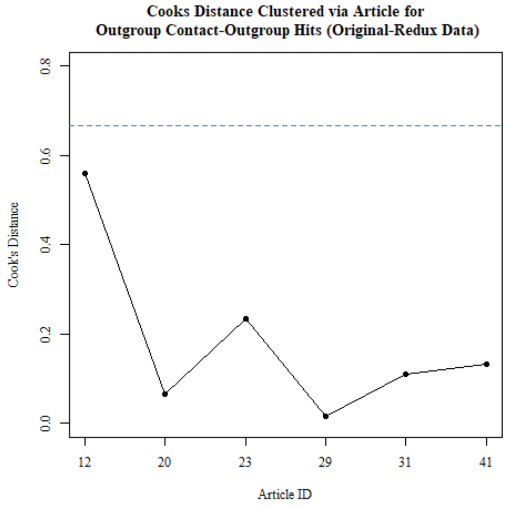
Diagnostic Plots for Reported Data – 'original-all' Model



Note. (AQ1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AQ1-B) Distribution of effect size outliers. (AQ1-C) Distribution of effect size outliers faceted by article ID.

Figure AQ2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model



Appendix AR

Quality of Outgroup Contact – Identification Difference Scores: Publication Bias

Eggers Test

This is a statistical test of funnel plot asymmetry. A significant test suggests publication bias may be present.

Original-all

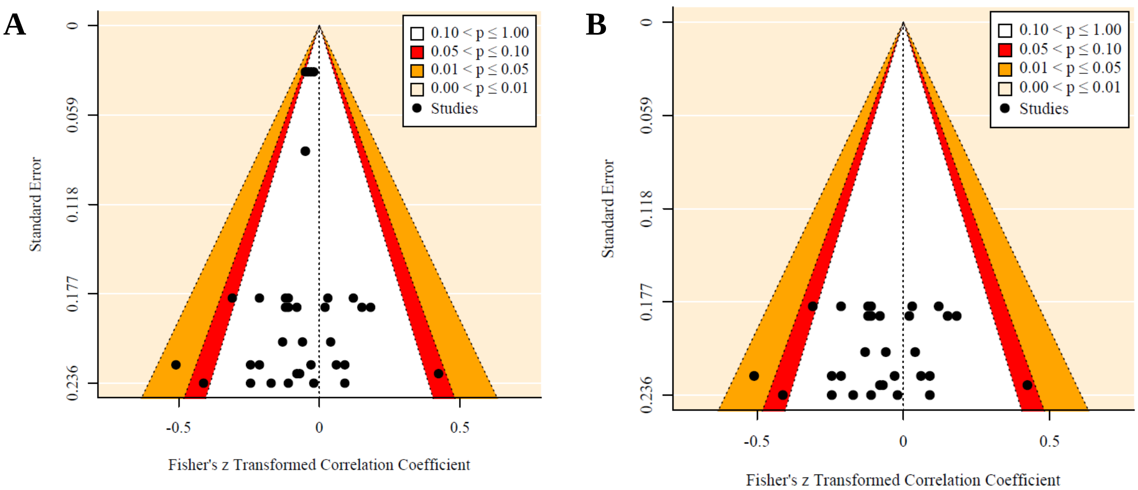
The regression test was not significant (-.20, z=-.88, p>.05)

Original-redux

. The regression test was not significant (-.80, z=-.48, p>.05)

Figure AR1

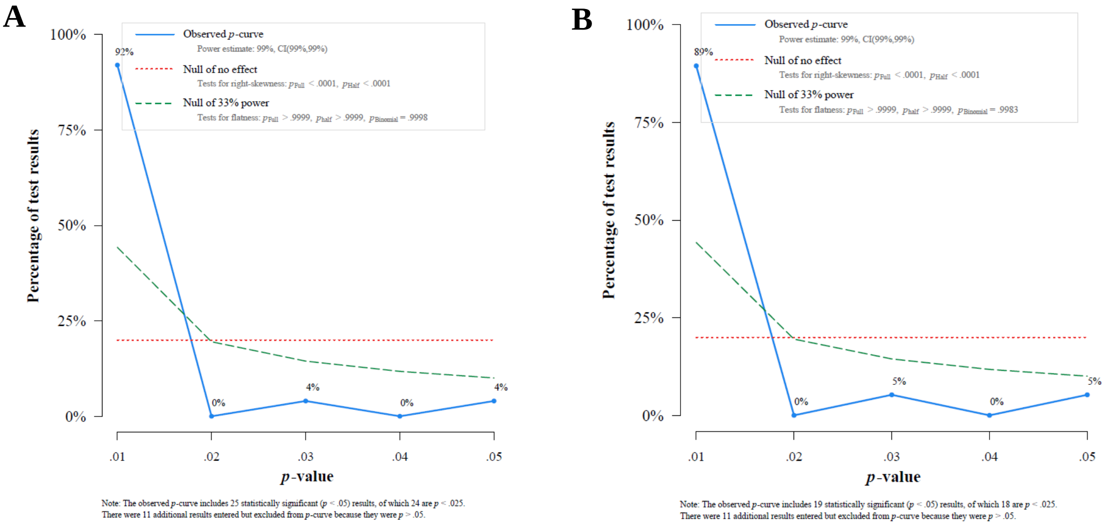
Funnel Plots for Quality of Outgroup Contact – Identification Difference Scores Before and After Influential Cases were Removed



Note. (AR1-A) Funnel plot testing for publication bias using reported data – ‘original-all’ model. (AR1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model.

Figure AR2

P-curve Analysis for Quality of Outgroup Contact – Identification Difference Scores Before and After Influential Cases and Outliers were Removed



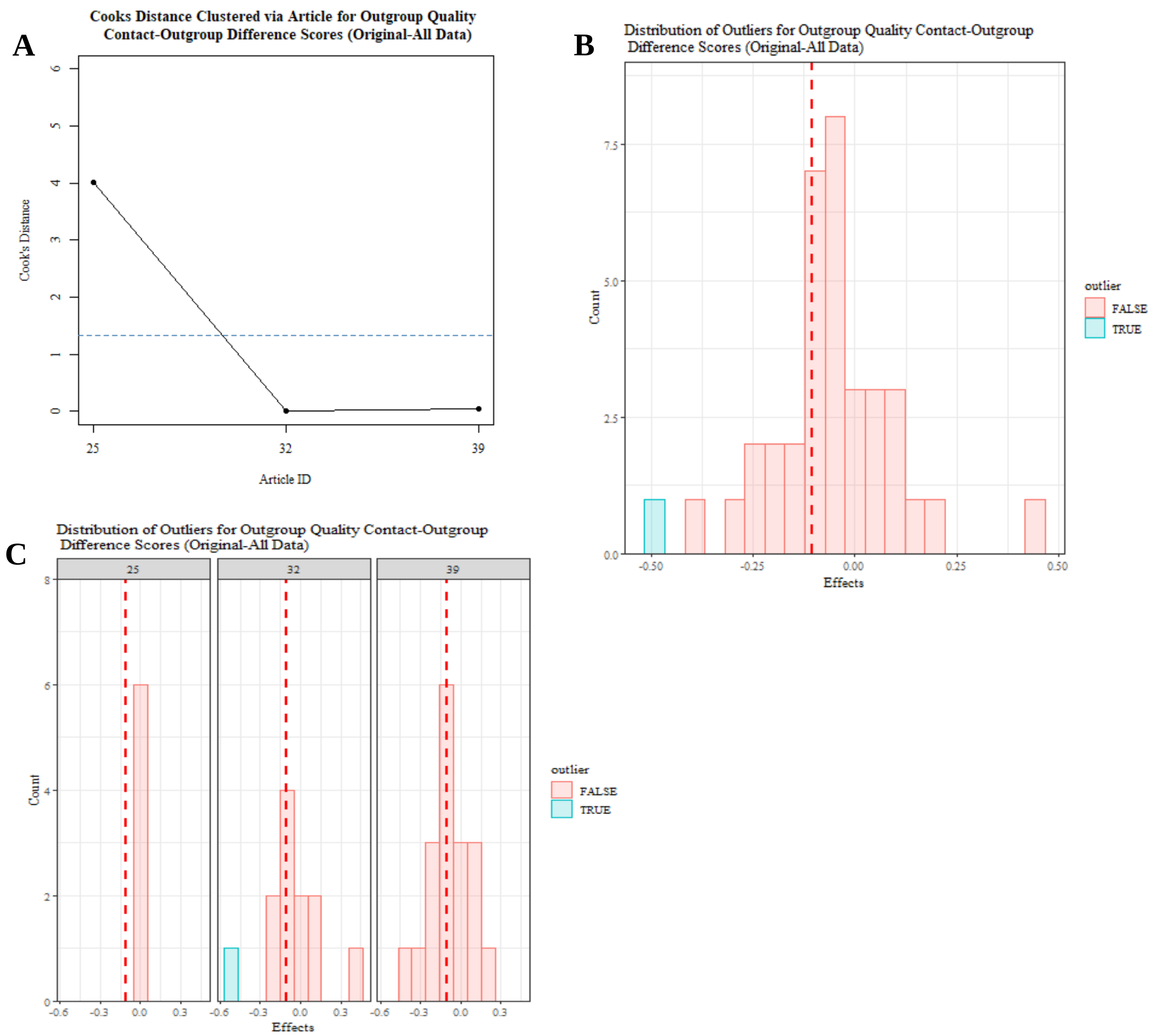
Note. (AR2-A) P-curve plot testing for publication bias using reported data – ‘original-all’ model. (AR2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – ‘original-redux’ model.

Appendix AS

Quality of Outgroup Contact – Identification Difference Scores: Diagnostic Plots

Figure AS1

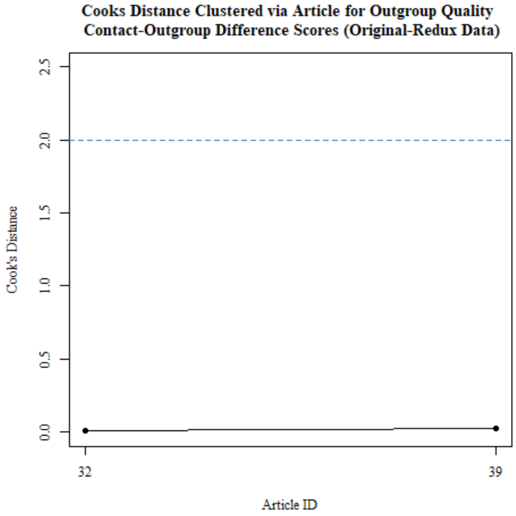
Diagnostic Plots for Reported Data – 'original-all' Model



Note. (AS1-A) Plot depicting *Cooks Distance* after nesting effects within parent articles. (AS1-B) Distribution of effect size outliers. (AS1-C) Distribution of effect size outliers faceted by article ID.

Figure AS2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model



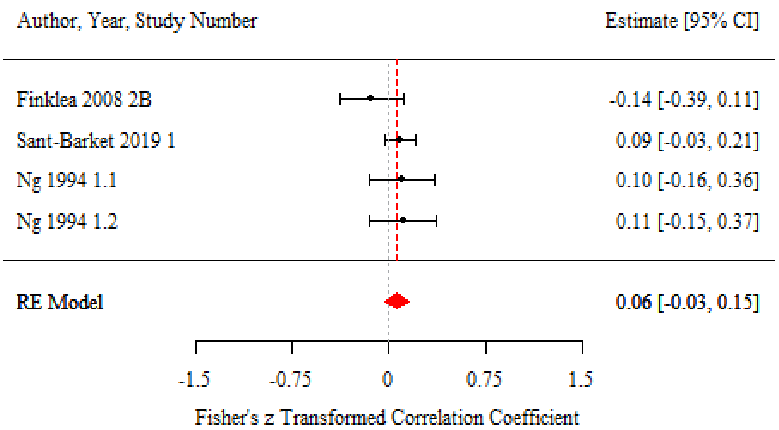
Appendix AT

Quantity of Outgroup Contact – Outgroup False Alarms

The aggregate effect (*r*) suggests a positive relationship between outgroup contact and outgroup false-alarms (*r*=.06). This is counter to expectations and therefore, hypothesis 2e is not supported within the given sample which contains a limited number of data points. The aggregate effect is non-significant (p<.05) and due to the limited sample size and no outliers, only one model was assessed.

Figure AT1

Forest Plot for Quantity of Outgroup Contact – Outgroup False Alarms using Reported Data



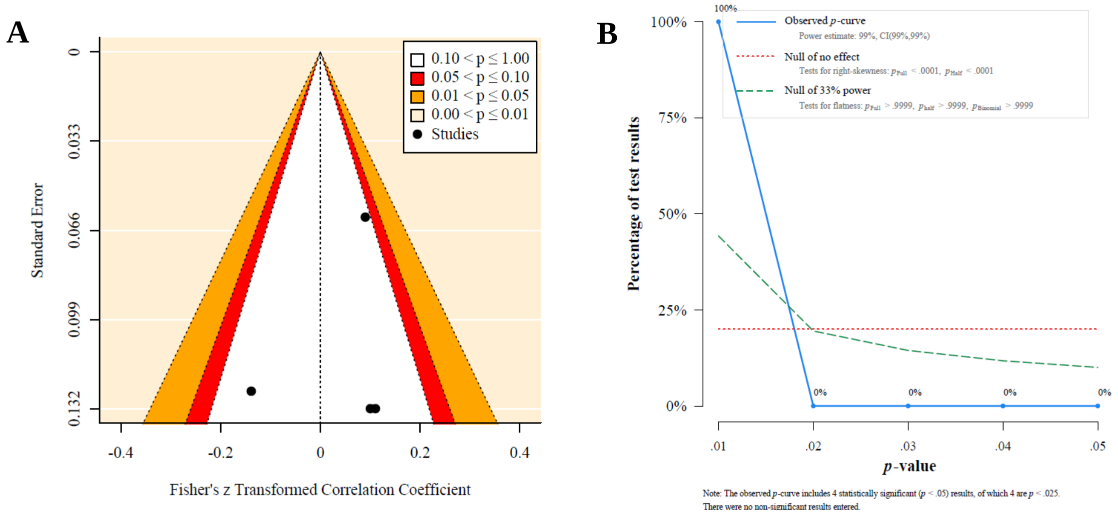
*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – ‘original-all’ model.

Publication Bias

Owing to a limited sample, no assertions were made regarding publication bias

Figure AT2

Diagnostics Plots for Quantity of Outgroup Contact – Outgroup False Alarms using Reported Data



Note. (AT2-A) Funnel Plot testing for publication bias in the 'original-all' model. (AT2-B) P-curve analysis testing for publication bias in the 'original-all' model.

Moderators

The model yielded no significant moderators for the outgroup contact – outgroup false alarm relationship.

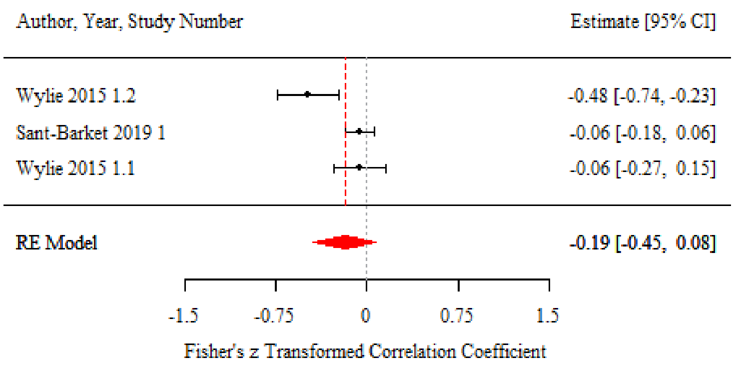
Appendix AU

Quantity of Outgroup Contact – Outgroup Response Bias

The aggregate effect (*r*) suggests a relationship between outgroup contact and outgroup response bias that is counter to expectations. Due to a limited sample size, a single '-all' model was run which found that as outgroup contact increased it was associated with a lower response bias i.e. a more liberal decision criterion. Hypothesis 2d is therefore not supported. The aggregate effect was non-significant (*r*=-.19, p>.05)

Figure AU1

Forest Plot for Quantity of Outgroup Contact – Outgroup Response Bias using Reported Data



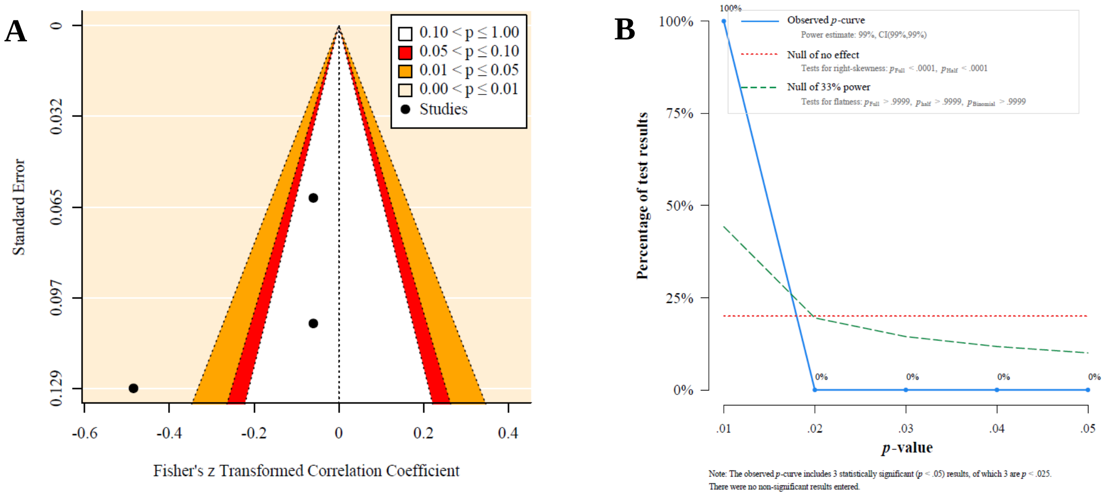
*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model.

Publication Bias

Owing to a limited sample, no assertions were made regarding publication bias

Figure AU2

Diagnostics Plots for Quantity of Outgroup Contact – Outgroup Response Bias using Reported Data



Note. (AU2-A) Funnel Plot testing for publication bias in the 'original-all' model. (AU2-B) P-curve analysis testing for publication bias in the 'original-all' model.

Moderators

Implicit Prejudice

Combined option 1. When participants have greater implicit outgroup prejudice, the negative relationship between outgroup contact and outgroup response bias is strengthened (Original-all: -.79, z=-3.00, p<.01). Therefore, when implicit prejudice is demonstrated, outgroup contact is associated with even poorer outgroup recognition. In the case of response bias this is an even greater liberal decision strategy.

Combined option 2. The same pattern of results is observed (Original-all: -.79, z=-3.00, p<.01).

Merged. The same pattern of results is observed (Original-all: -.79, z=-3.00, p<.01)

Completed. The same pattern of results is observed (Original-all: -.79, z=-3.00, p<.01). Even within a small sample, implicit prejudice was an important variable to account for as it explained 100% of the proportional reduction of total variance.

Sample Age

When participants age, higher outgroup contact is associated with an even greater liberal decision strategy (Original-all: -.01, z=-1.99, p<.01). This is counter to expectations. Despite the critical period for contact, later life outgroup contact should still result in a reduction in response bias even if the reduction is not as strong as it would be had contact occurred within the critical period.

Table AU1

Significant Moderators for Quantity of Outgroup Contact- Outgroup Response Bias

			Model 2 (Original -all)										
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	z	p	LB	UB	Sig	Prop. Variance reduction
Implicit prejudice for outgroup members (combined option one)	Core	-	3	2	-0.79	Yes	-3.00	0.003	-1.30	-0.27	**	100.00
Implicit prejudice for outgroup members (combined option two, using merged data completed with imputation)	Core	-	3	2	-0.79	Yes	-3.00	0.003	-1.30	-0.27	**	100.00
Harvard's 'Project Implicit' values for outgroup implicit prejudice	Core	-	3	2	-0.79	Yes	-3.00	0.003	-1.30	-0.27	**	100.00
Harvard's 'Project Implicit' values for outgroup implicit prejudice (completed)	Core	-	3	2	-0.79	Yes	-3.00	0.003	-1.30	-0.27	**	100.00
Sample Age	Methodological / Task	-	3	2	-0.01	No	-1.99	0.046	-0.02	0.00	*	55.39

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

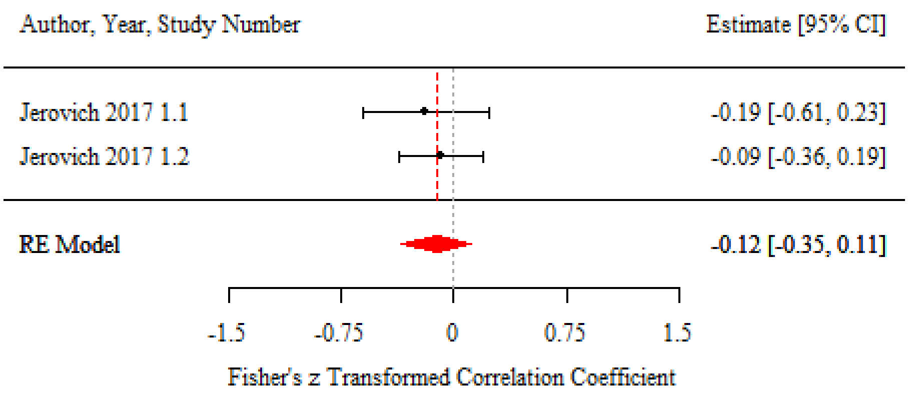
Appendix AV

Quality of Outgroup Contact – Outgroup Discriminability

Due to limited sample size, only one model was run. The aggregate effect (*r*) demonstrated a slight negative, non-significant relationship between outgroup quality of contact and outgroup discrimination (*r*=-.12, p>.05). This is counter to expectations therefore, hypothesis 2b was not supported for quality of outgroup contact.

Figure AV1

Forest Plot for Quality of Outgroup Contact – Outgroup Discriminability using Reported Data



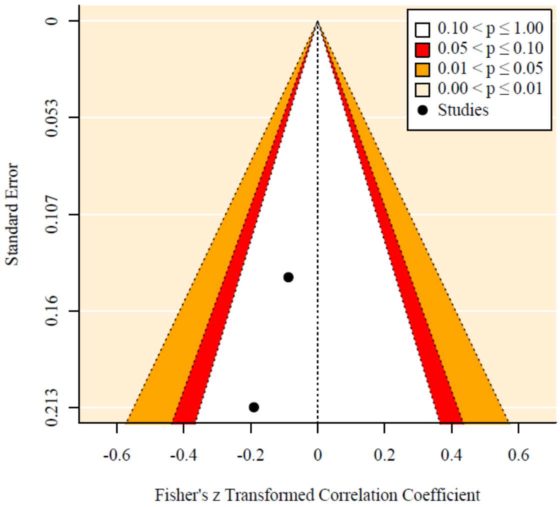
*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model.

Publication Bias

Owing to a limited sample, no assertions were made regarding publication bias

Figure AV2

Funnel Plot for Quality of Outgroup Contact – Outgroup Discriminability using Reported Data



Moderators

There were no statistically significant moderators for the outgroup quality of contact – outgroup discrimination relationship.

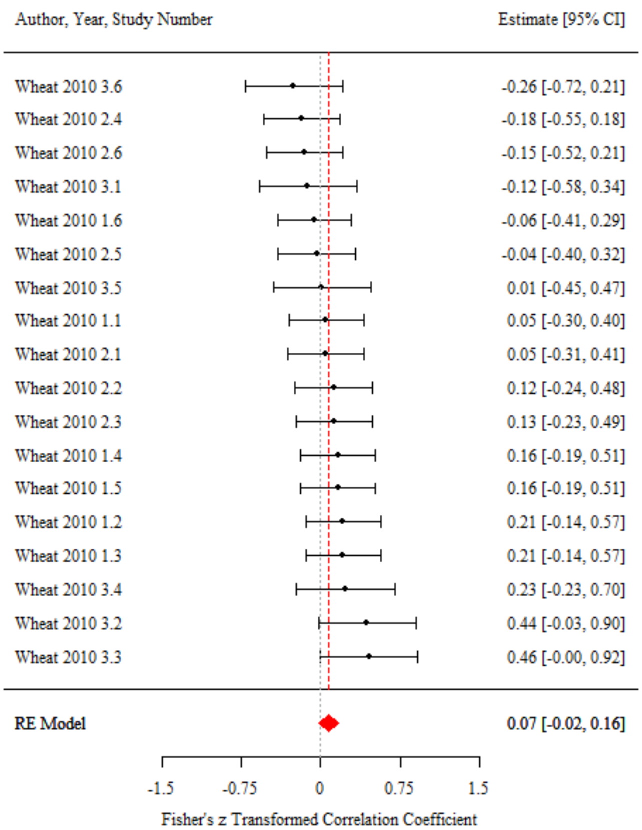
Appendix AW

Implicit Outgroup Prejudice – Identification Difference Scores

The aggregate effect (*r*) demonstrates a positive relationship between implicit outgroup prejudice and identification differences scores (*r*= .07). Increases in implicit outgroup prejudice are therefore associated with a larger OENE. Hypothesis 3a is therefore supported. The aggregate effect was however non-significant (p>.05).

Figure AW1

Forest Plot for Implicit Outgroup Prejudice – Identification Difference Scores using Reported Data



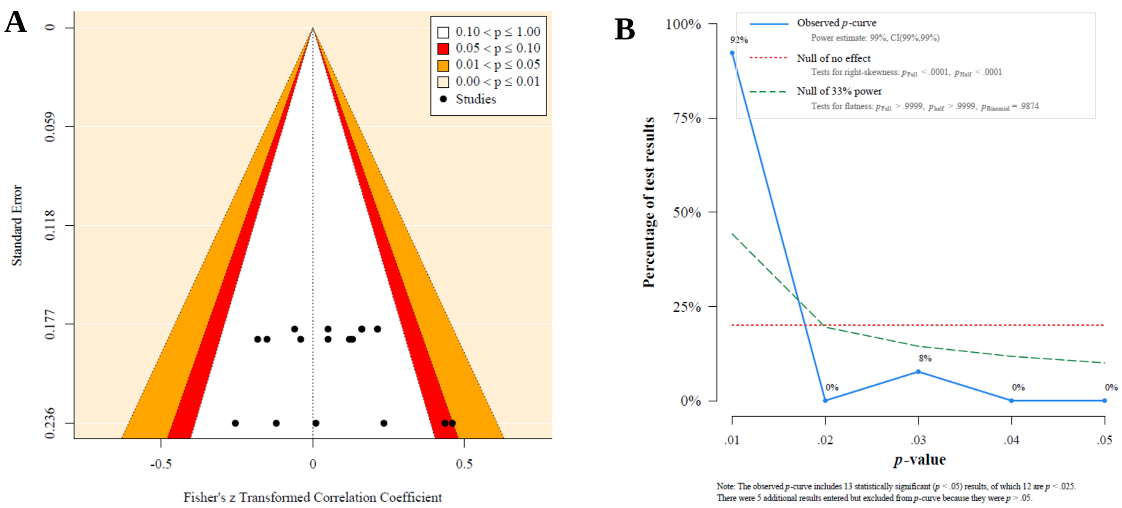
*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model.

Publication Bias

Owing to a limited sample size, comprised of a single article, assertions were not made regarding publication bias.

Figure AW2

Diagnostics Plots for Implicit Outgroup Prejudice – Identification Difference Scores using Reported Data



Note. (AW2-A) Funnel Plot testing for publication bias in the 'original-all' model. (AW2-B) P-curve analysis testing for publication bias in the 'original-all' model.

Moderators

Quantity of outgroup contact

Reported. As quantity of outgroup contact increases, implicit outgroup prejudice is associated with an even greater OENE, or larger identification difference scores (Original-all: 1.28, z=2.33, p<.05). This is consistent with expectations

Completed. The same pattern of results was observed (Original-all: 1.28, z=2.33, p<.05). For both reported and merged variables, the importance of accounting for outgroup contact was underscored by the 10.02% reduction in variance once included within the model.

Table AW1

Significant Moderators for Implicit Outgroup Prejudice- Identification Difference Scores

			Model 2 (Original -all)										
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction
Quantity of outgroup contact (reported)	Core	-	18	1	1.28	Yes	2.33	0.020	0.20	2.37	*	10.02
Quantity of outgroup contact (completed)	Core	-	18	1	1.28	Yes	2.33	0.020	0.20	2.37	*	10.02

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

Appendix AX

Explicit Outgroup Prejudice – Identification Difference Scores

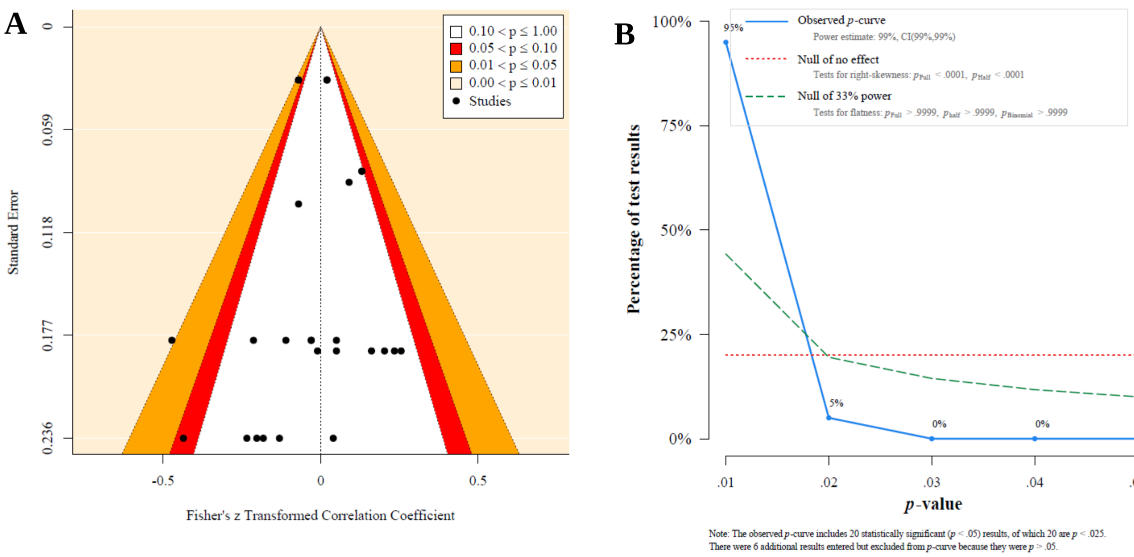
The aggregate effects (*r*) for both models demonstrates a negative relationship between explicit outgroup prejudice and identification difference scores. Lower explicit outgroup prejudice is associated with a decrease in the OENE, or the size of the identification difference scores. Hypothesis 3a is therefore supported for explicit outgroup prejudice. Both aggregate effects were slight and non-significant (p>.05).

Publication Bias

Publication bias was not present in the sample

Figure AX1

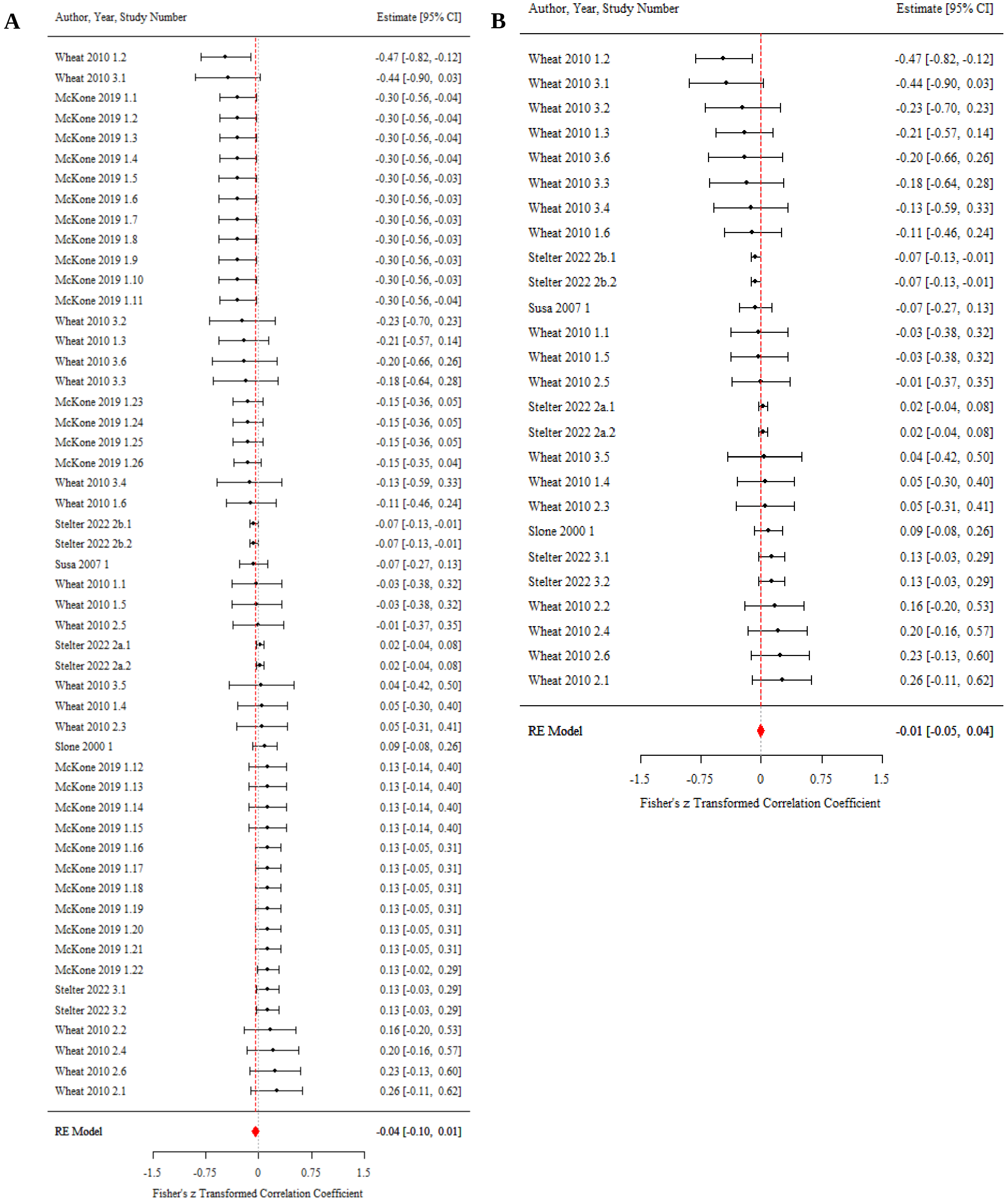
Diagnostics Plots for Explicit Outgroup Prejudice – Identification Difference Scores After Influential Cases and Outliers were Removed from Reported Data



Note. (AX1-A) Funnel Plot testing for publication bias in the 'original-redux' model. (AX1-B) P-curve analysis testing for publication bias in the 'original-redux' model.

Figure AX2

Forest Plots for Explicit Outgroup Prejudice – Identification Difference Scores Before and After Influential Cases and Outliers were Removed



*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. (AX2-A) Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model. (AX2-B) Forest plot for original data after influential cases and outliers were removed – 'original-redux' model.

Moderators

Implicit Outgroup Prejudice

Combined Option 1. When implicit outgroup prejudice increases, lower outgroup explicit prejudice is associated with a less effective reduction in the size of the OENE (Original-all: .85, $z=7.93$, $p<.001$). This is consistent with expectations.

Combined Option 2. The same pattern of results was observed (Original-all: .84, $z=7.92$, $p<.001$)

Merged. The same pattern of results was observed (Original-all: .65, $z=6.02$, $p<.001$)

Merged completed. The same pattern of results was observed (Original-all: .64, $z=6.37$, $p<.001$). Implicit prejudice is an important moderator which accounted for a large proportion of model variance (between 61.32 – 75.71% across implicit variables)

Quantity of outgroup contact

Reported. When quantity of outgroup contact is increased, lower explicit outgroup prejudice is associated with a greater reduction in the observed OENE (Original-all: -.51, $z=-4.58$, $p<.001$). Quantity of outgroup contact accounted for 30.48% of the model variance, supporting the importance of accounting for outgroup contact.

Completed. The same pattern of results was observed (Original-all: -.48, $z=-4.36$, $p<.001$, 27.15%)

Quality of outgroup contact (completed)

When quality of outgroup contact is increased, lower explicit outgroup prejudice is associated with an even greater reduction in the observed OENE (Original-all: -.35, $z=-3.05$, $p<.01$, 22.77%).

Quantity of outgroup contact in or out of critical period

When outgroup contact occurs in high levels within the critical period, lower explicit outgroup prejudice is associated with a bigger reduction in the OENE (Original-all: -.17, $z=-2.67$, $p<.01$). Comparatively, when outgroup contact occurs at low levels in the critical period or low levels later in life, lower explicit outgroup prejudice is associated with smaller reductions in the OENE (Original-all: .21, $z=2.53$, $p<.05$; Original-all: .19, $z=2.67$, $p<.01$

respectively). This categorical interaction term is important (43.18% of variance is attributed to both level and period in which outgroup contact occurred).

Positionality of participants relative to outgroup members

When minority members are tested on majority outgroup members within an identification task, lower explicit outgroup prejudice is associated with a greater reduction in the observed OENE (Original-all: -.29, $z=-6.98$, $p<.001$).

Task

When a CFMT is used, lower explicit outgroup prejudice is associated with a bigger reduction in the OENE (Original-all: -.08, $z=-2.21$, $p<.05$)

Sample Age

As participants age, lower explicit outgroup prejudice is associated with a smaller reduction in the OENE (Original-redux: -.00, $z=-3.44$, $p<.001$).

Table AX1

Significant Moderators for Explicit Outgroup Prejudice- Identification Difference Scores across

			Model 2 (Original - All)											Model 1 (Original Redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	z	p	LB	UB	Sig	Prop. Variance reduction	No. effects	No. studies	Est.	Expected	z	p	LB	UB	Sig	Prop. Variance reduction
Implicit prejudice for outgroup members (combined option one)	Core	-	51	4	0.85	Yes	7.93	<.0001	0.64	1.05	***	70.77	-	-	-	-	-	-	-	-	-	-
Implicit prejudice for outgroup members (combined option two, using merged data completed with imputation)	Core	-	52	5	0.84	Yes	7.92	<.0001	0.63	1.04	***	75.71	-	-	-	-	-	-	-	-	-	-
Harvard’s 'Project Implicit' values for outgroup implicit prejudice	Core	-	45	4	0.65	Yes	6.02	<.0001	0.44	0.86	***	61.32	-	-	-	-	-	-	-	-	-	-
Harvard’s 'Project Implicit' values for outgroup implicit prejudice (completed)	Core	-	52	5	0.64	Yes	6.37	<.0001	0.45	0.84	***	70.48	-	-	-	-	-	-	-	-	-	-
Quantity of outgroup contact (reported)	Core	-	50	3	-0.51	Yes	-4.58	<.0001	-0.72	-0.29	***	30.48	-	-	-	-	-	-	-	-	-	-
Quantity of outgroup contact (completed)	Core	-	52	5	-0.48	Yes	-4.36	<.0001	-0.69	-0.26	***	27.15	-	-	-	-	-	-	-	-	-	-
Quality of outgroup contact (completed)	Core	-	52	5	-0.35	Yes	-3.01	0.003	-0.57	-0.12	**	22.77	-	-	-	-	-	-	-	-	-	-
Quantity of outgroup contact, in or out of critical period (Interaction term)	Core	Critical-high	31	2	-0.17	Yes	-2.67	0.008	-0.29	-0.04	**	43.18	-	-	-	-	-	-	-	-	-	-
		Critical-low vs Critical-high			0.21	Yes	2.53	0.011	0.05	0.37	*		-	-	-	-	-	-	-	-	-	-
		Later life-low vs Critical-high			0.19	Yes	2.67	0.008	0.05	0.33	**		-	-	-	-	-	-	-	-	-	-
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Minority-Majority vs Majority-Minority	52	5	-0.29	Yes	-6.98	<.0001	-0.37	-0.21	***	78.83	-	-	-	-	-	-	-	-	-	-
Task	Methodological / Task	CFMT	52	5	-0.08	Yes	-2.21	0.027	-0.14	-0.01	*	EMV	-	-	-	-	-	-	-	-	-	-
Sample age	Methodological / Task	-	-	-	-	-	-	-	-	-	-	-	26	4	0.00	-	-3.44	0.001	-0.01	0.00	***	EMV

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance

Appendix AY

Explicit Outgroup Prejudice – Identification Difference Scores: Publication Bias

Eggers Test

Original-all

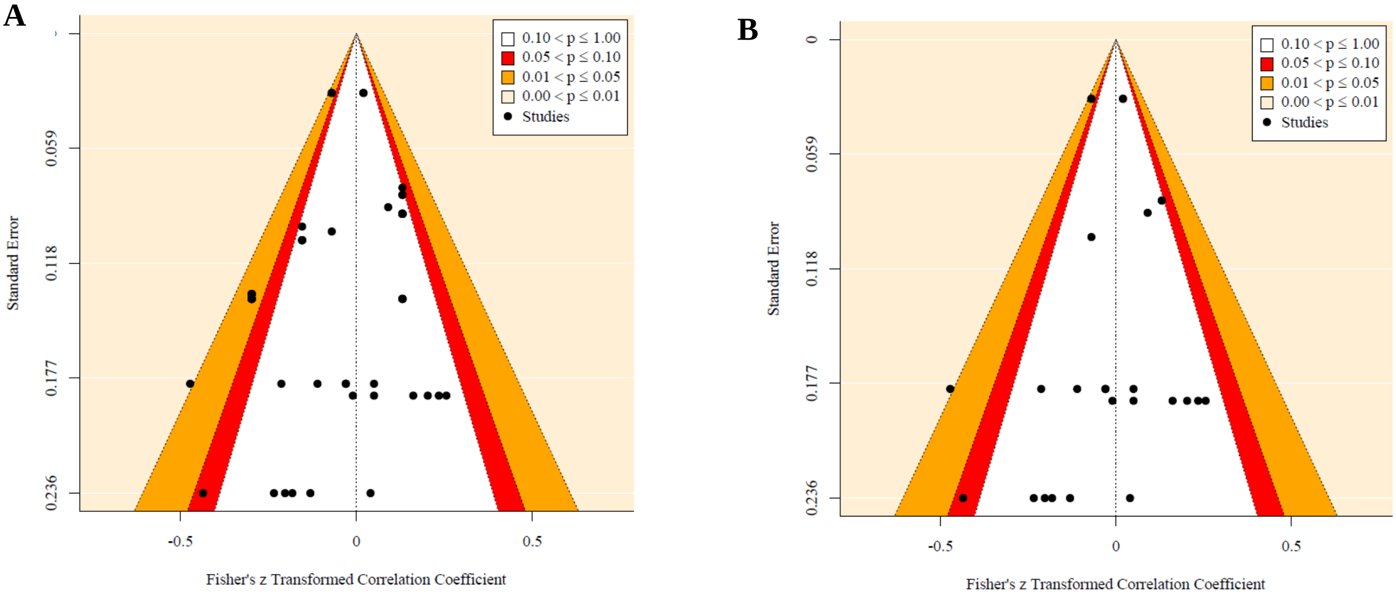
The regression test was not significant (-1.14, z=-1.95, p>.05)

Original-redux

. The regression test was not significant (-.20, z=.-.55, p>.05)

Figure AY1

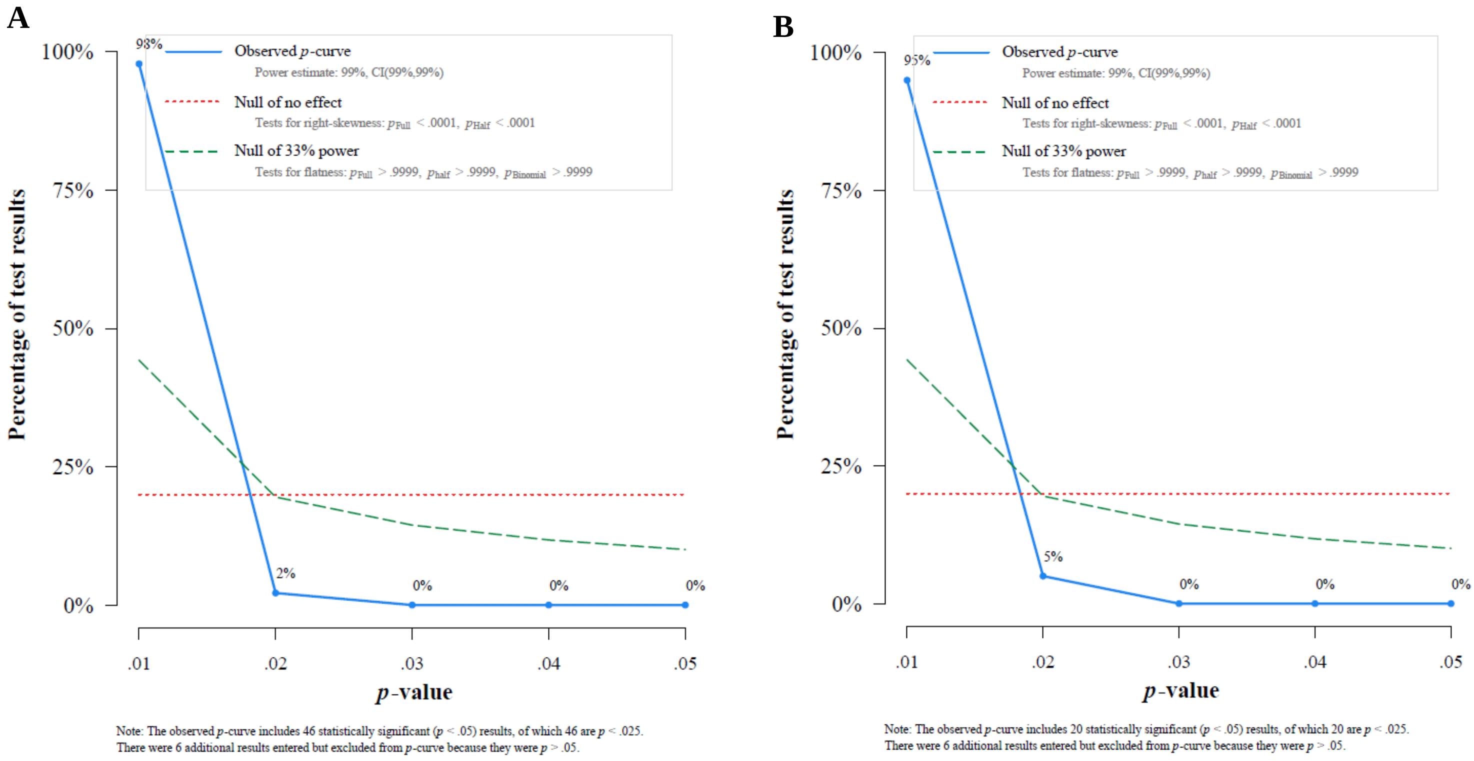
Funnel Plots for Explicit Outgroup Prejudice – Identification Difference Scores Before and After Influential Cases were Removed



Note. (AY1-A) Funnel plot testing for publication bias using reported data – 'original-all' model. (AY1-B) Funnel plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model.

Figure AY2

P-curve Analysis for Explicit Outgroup Prejudice – Identification Difference Scores Before and After Influential Cases and Outliers were Removed



Note. (AY2-A) P-curve plot testing for publication bias using reported data – 'original-all' model. (AY2-B) P-curve plot testing for publication bias using reported data after influential cases and outliers were removed – 'original-redux' model

Appendix AZ

Explicit Outgroup Prejudice – Identification Difference Scores: Diagnostic Plots

Figure AZ1

Cooks Distance for Reported Data – 'original-all' Model

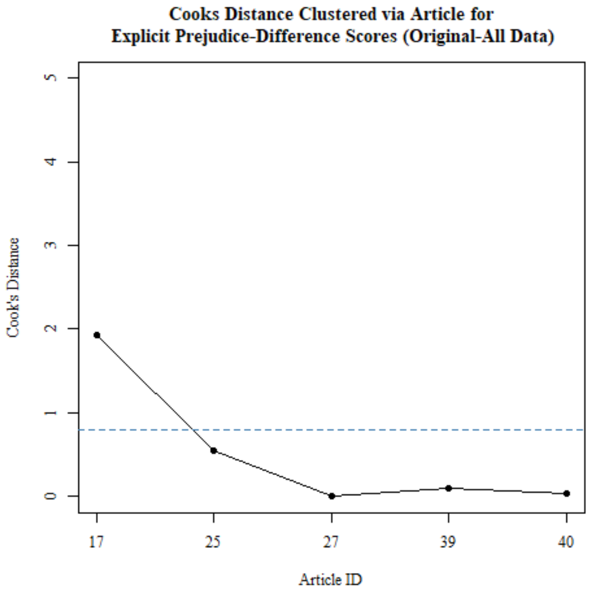
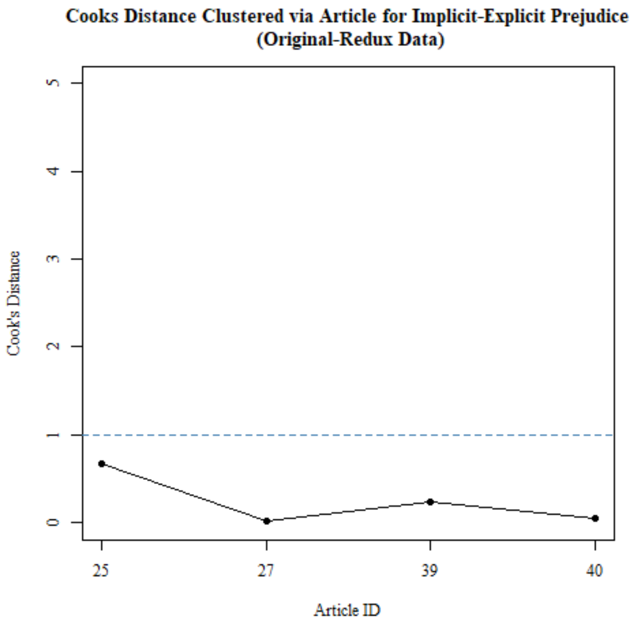


Figure AZ2

Cooks Distance for Reported Data after Influential Cases and Outliers were Removed – 'original-redux' Model



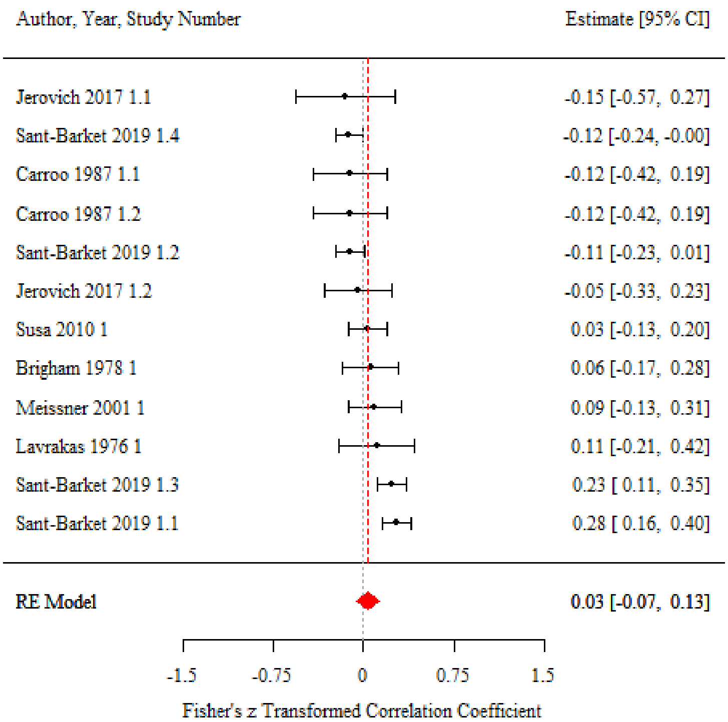
Appendix AAA

Explicit Outgroup Prejudice – Outgroup Discriminability

The aggregate effect demonstrated a positive relationship between explicit outgroup prejudice and outgroup discrimination. Lower explicit outgroup prejudice, or conversely higher favourable outgroup attitudes, are therefore associated with increased outgroup discrimination (r=.03). Hypothesis 3b was supported. The aggregate effect was non-significant (p>.05)

Figure AAA1

Forest Plot for Explicit Outgroup Prejudice – Outgroup Discriminability using Reported Data



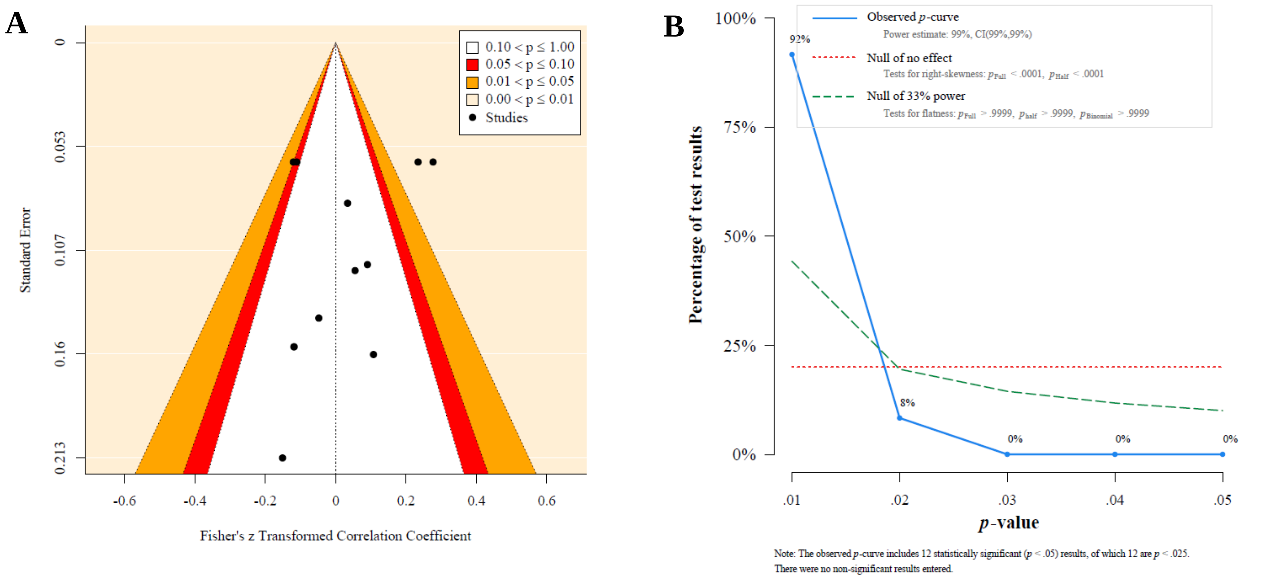
*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model.

Publication Bias

Publication bias was not evident within the sample.

Figure AAA2

Diagnostic Plots for Explicit Outgroup Prejudice – Outgroup Discriminability using Reported Data



Note. (AAA2-A) Funnel Plot testing for publication bias in the 'original-all' model. (AAA2-B) P-curve analysis testing for publication bias in the 'original-all' model.

Moderators

There were no statistically significant moderators for the explicit outgroup prejudice – outgroup discrimination relationship.

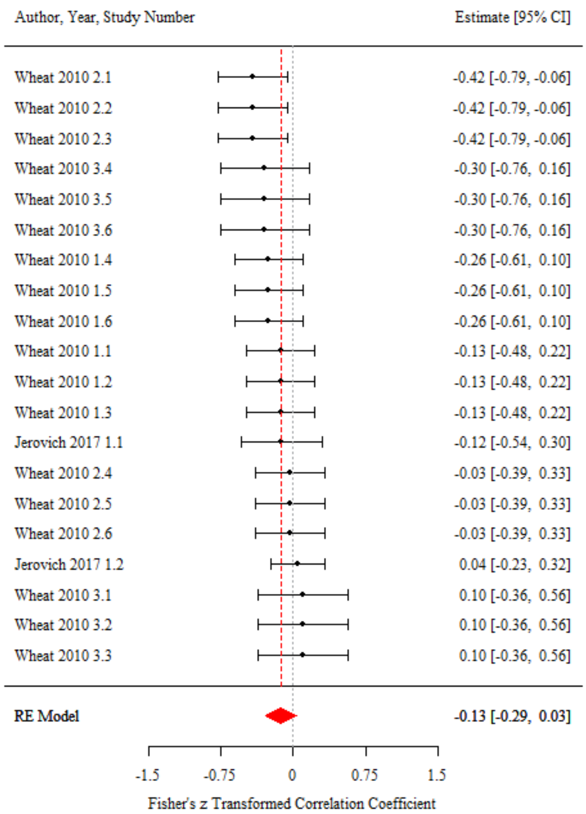
Appendix AAB

Implicit Outgroup Prejudice – Quantity of Outgroup Contact

The aggregate effect (*r*) demonstrated a negative relationship, consistent with expectations (*r*= -.13, p>.05). Higher levels of implicit outgroup prejudice are associated with a reduction in quantity of outgroup contact. Hypothesis 4a is therefore supported.

Figure AAB1

Forest Plot for Implicit Outgroup Prejudice – Quantity of Outgroup Contact using Reported Data



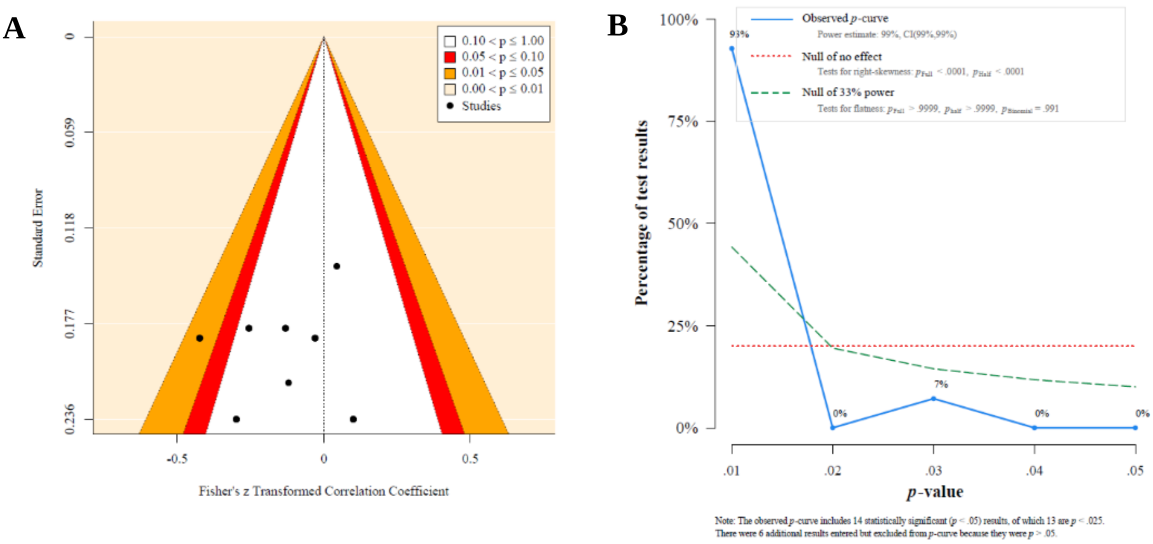
*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model.

Publication Bias

Publication bias was not present in the sample.

Figure AAB2

Diagnostic Plots for Implicit Outgroup Prejudice – Quantity of Outgroup Contact using Reported Data



Note. (AAB2-A) Funnel Plot testing for publication bias in the 'original-all' model. (AAB2-B) P-curve analysis testing for publication bias in the 'original-all' model.

Moderators

Fixed or self-paced encoding

When the length of time allocated to studying a face is self-paced, higher implicit outgroup prejudice is associated with a greater reduction in quantity of outgroup contact (Original-all: -.18, z=-3.29, p<.01).

Motivation

When no motivation manipulations or instructions are utilized, higher implicit outgroup prejudice is associated with a greater reduction in quantity of outgroup contact (Original-all: -.18, z=-2.14, p<.05)

Task

When a delayed matching task is used, higher implicit outgroup prejudice is associated with an even greater reduction in quantity of outgroup contact (Original-all: -.18, z=-3.28, p<.001).

Table AAB1

Significant Moderators for Implicit Outgroup Prejudice - Quantity of Outgroup Contact

			Model 2 (Original -all)										
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	z	p	LB	UB	Sig	Prop. Variance reduction
Fixed or self-paced encoding	Methodological / Task	Self-paced	20	3	-0.18	Yes	-3.28	0.001	-0.29	-0.07	**	100.00
Presence of motivation manipulation or instructions	Methodological / Task	None	20	2	-0.18	Yes	-2.14	0.033	-0.35	-0.02	*	87.14
Task	Methodological / Task	Delayed Matching	20	3	-0.18	-	-3.28	0.001	-0.29	-0.07	***	87.12

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

Appendix AAC

Explicit Outgroup Prejudice – Quantity of Outgroup Contact

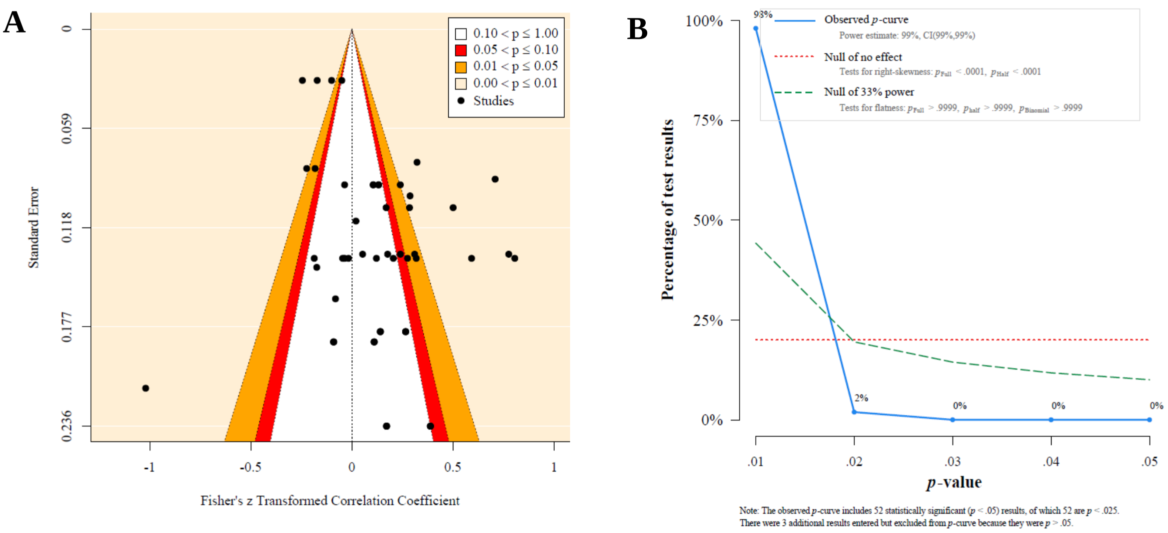
The aggregate effect demonstrates a positive relationship wherein higher favourable outgroup attitudes, or lower explicit outgroup prejudice, is associated with increased outgroup contact (*r*=.05, p>.05). Thereby supporting hypothesis 4a.

Publication Bias

Publication bias was not evident within the sample.

Figure AAC1

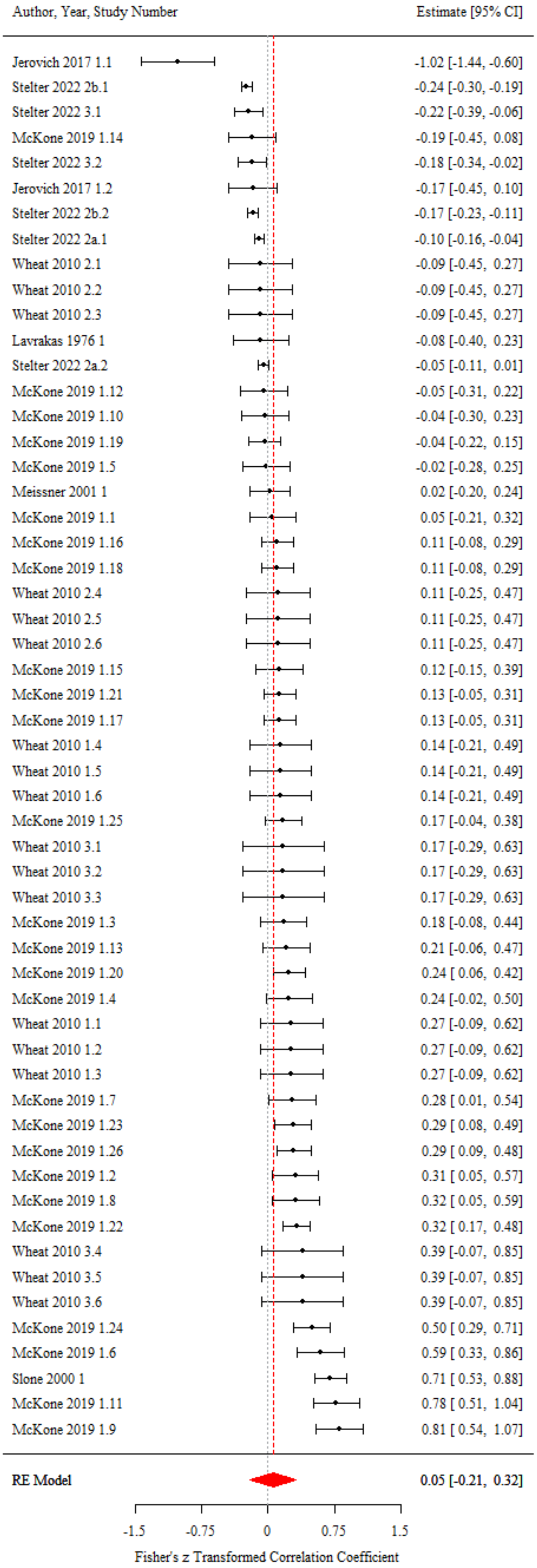
Diagnostic Plots for Explicit Outgroup Prejudice – Quantity of Outgroup Contact using Reported Data



Note. (AAC1-A) Funnel Plot testing for publication bias in the 'original-all' model. (AAC1-B) P-curve analysis testing for publication bias in the 'original-all' model.

Figure AAC2

Forest Plot for Explicit Outgroup Prejudice – Quantity of Outgroup Contact using Reported Data



*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model.

Moderators

Quantity of outgroup contact time bands

When outgroup contact occurs during middle childhood (6-12) as opposed to during adulthood, the relationship between explicit outgroup prejudice and outgroup contact is weakened. If outgroup contact occurs in this time band, lower explicit prejudice is associated with a decrease in outgroup contact. This is counter to expectations of beneficial contact occurring within the critical window (Original-all: -.20, z=-1.98, p<.05).

Table AAC1

Significant Moderators for Explicit Outgroup Prejudice - Quantity of Outgroup Contact

			Model 2 (Original -all)										
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected direction	z	p	LB	UB	Sig	Prop. Variance reduction
Quantity of outgroup contact time bands (multiple)	Core	Middle childhood (6-12) vs Adult (18+)	55	7	-0.20	No	-1.98	0.047	-0.39	0.00	*	EMV

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<. 001.The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as 'EMV' or exceeds model variance

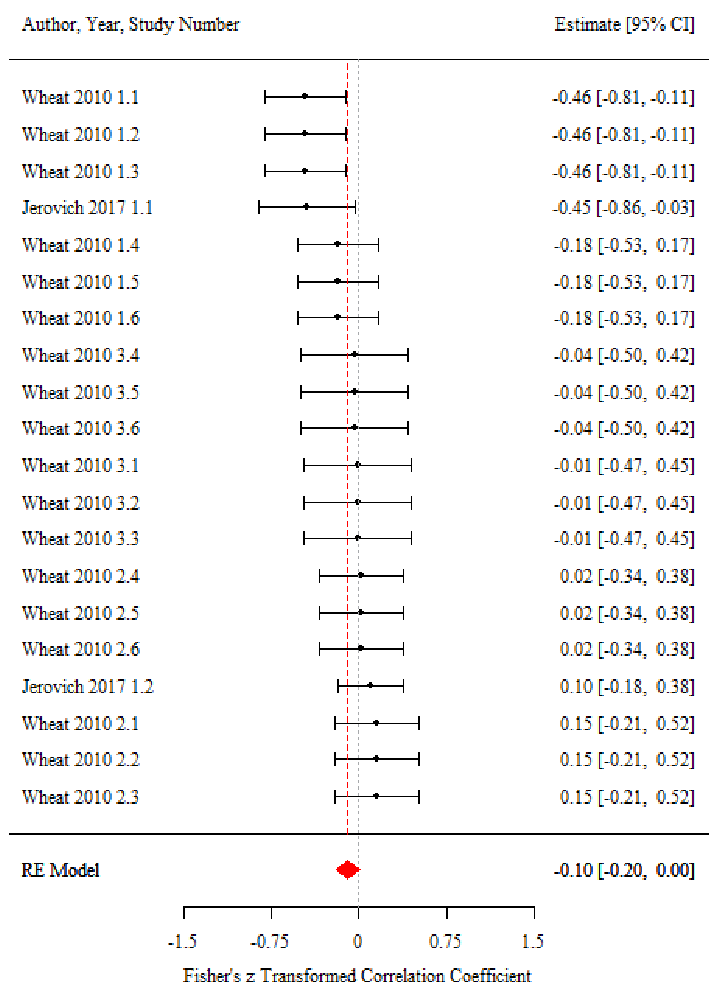
Appendix AAD

Implicit Outgroup Prejudice – Explicit Outgroup Prejudice

Consistent with expectations, the aggregate effect demonstrates that higher implicit outgroup prejudice is associated with lower favourable outgroup attitudes, or conversely higher explicit outgroup prejudice (r=-.10, p>.05).

Figure AAD1

Forest Plot for Implicit Outgroup Prejudice – Explicit Outgroup Prejudice using Reported Data



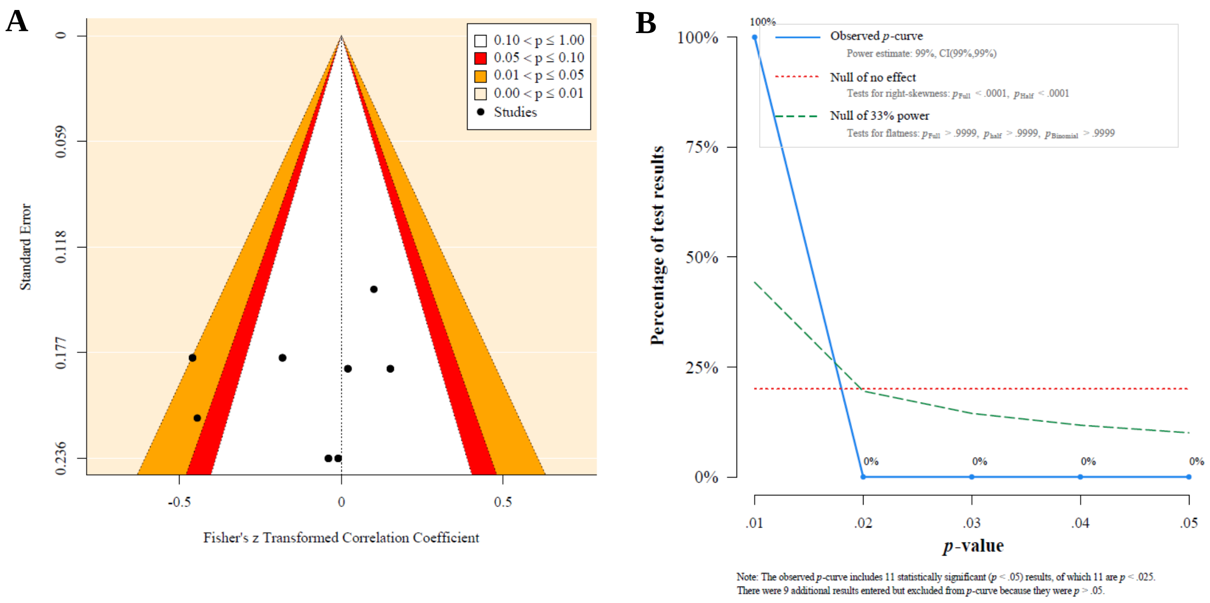
*Note***.** Forest plots depict both the effects used in the analysis and the aggregate effect - reported at the bottom of the plot. The aggregate effect is depicted via the red line and diamond. Forest plot for original or actual data prior to removing influential cases or outliers – 'original-all' model.

Publication bias

Publication bias was not evident in the sample

Figure AAD2

Diagnostic Plots for Implicit Outgroup Prejudice – Explicit Outgroup Prejudice using Reported Data



Note. (AAD2-A) Funnel Plot testing for publication bias in the 'original-all' model. (AAD2-B) P-curve analysis testing for publication bias in the 'original-all' model.

Moderators

Sample Age

As participants become older, the relationship between implicit and explicit outgroup prejudice is weakened. With age, higher implicit outgroup prejudice is associated with lower explicit outgroup prejudice (Original-all: .02, z=3.79, p<.001)

Positionality of participants relative to outgroup members

When majority members are tested on minority outgroup members, the relationship between both types of prejudice is strengthened. Higher outgroup prejudice is associated more strongly with higher explicit outgroup prejudice (Original-all: -.24, z=-3.32, p<.001). By comparison minority members tested on either minority or majority members weaken the relationship between both types of prejudice (p<.05)

Table AAD1

Significant Moderators for Implicit Outgroup Prejudice-Explicit Outgroup Prejudice

			Model 2 (Original-all)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	z	p	LB	UB	Sig	Prop. Variance reduction
Sample age	Methodological / Task	-	20	2	0.12	-	3.79	0.000	0.06	0.18	***	EMV
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Majority-Minority	20	2	-0.24	Yes	-3.32	0.001	-0.38	-0.10	***	54.99
		Minority-Majority vs Majority-Minority			0.25	Yes	2.28	0.022	0.04	0.47	*	
		Minority-Minority vs Majority-Minority			0.23	Yes	2.03	0.043	0.01	0.46	*	

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance

Appendix AAE

Outgroup Prejudice-Outgroup Contact and Implicit-Explicit Outgroup Prejudice: Diagnostic Plots

Figure AAE1

Cooks Distance for Implicit Outgroup Prejudice – Outgroup Quantity of Contact Effects for Reported Data – 'original-all' Model

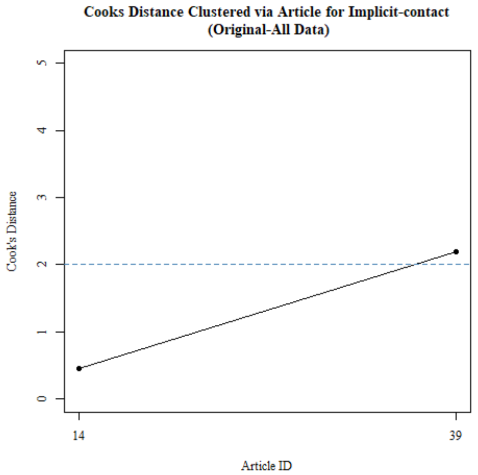


Figure AAE2

Cooks Distance for Explicit Outgroup Prejudice – Outgroup Quantity of Contact Effects for Reported Data – 'original-all' Model

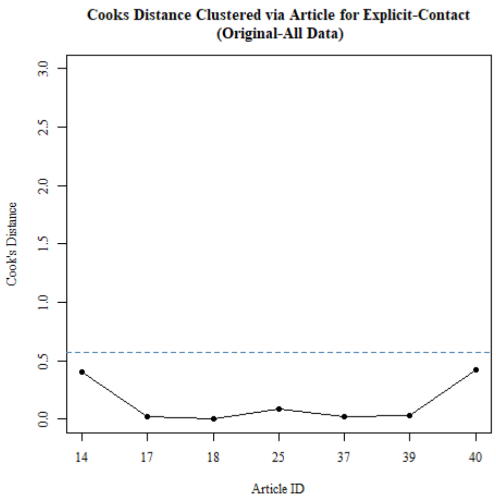
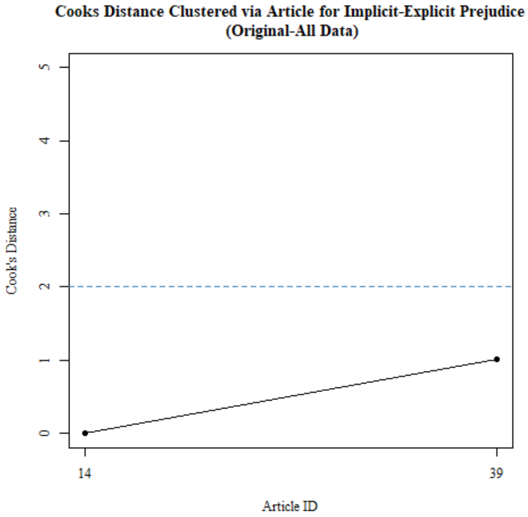


Figure AAE3

Cooks Distance for Implicit Outgroup Prejudice -Explicit Outgroup Prejudice Effects for Reported Data – 'original-all' Model



Appendix AAF

Significant Moderators for Quantity of Outgroup Contact – Identification Difference Scores

			Model 2 (Original - All)										**Model 1 (Original Redux)**									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction
Harvard's 'Project Implicit' values for outgroup explicit prejudice (completed)	Core	-	94	16	-0.59	Yes	-2.50	0.012	-1.04	-0.13	*	7.80	-	-	-	-		-	-	-	-	-
Quantity of contact in or out of critical period	Core	Critical	34	4	-0.19	Yes	-3.25	0.001	-0.31	-0.08	**	44.03	34	4	-0.19	Yes	-3.25	0.001	-0.31	-0.08	**	EMV
		Later life vs Critical			0.06	Yes	1.97	0.048	0.00	0.13	*				0.06	Yes	1.97	0.048	0.00	0.13	*	
Quantity of outgroup contact, in or out of critical period (Interaction term)	Core	Critical-High	34	4	-0.21	Yes	-2.52	0.012	-0.38	-0.05	*	8.57	34	4	-0.21	Yes	-2.52	0.012	-0.38	-0.05	*	EMV
Quantity of outgroup contact time bands (multiple)	Core	Adult (18+)	94	16	-0.15	Yes	-2.81	0.005	-0.26	-0.05	**	EMV	87	14	-0.16	Yes	-3.75	0.000	-0.25	-0.08	***	EMV
		Middle childhood (6 -12) vs Adult (18+)			-0.21	Yes	-2.94	0.003	-0.35	-0.07	**				-0.20	Yes	-3.23	0.001	-0.32	-0.08	**	
		Lifespan vs Adult (18+)			-	-	-	-	-	-	-				0.09	No	2.20	0.028	0.01	0.17	*	
Encoding instructions	Methodological / Task	Directed	46	13	-0.56	Yes	-2.36	0.018	-1.03	-0.10	*	EMV	39	11	-0.59	Yes	-2.88	0.004	-0.99	-0.19	**	EMV
		Standard vs Directed			0.53	Yes	2.22	0.026	0.06	1.00	*				0.51	Yes	2.46	0.014	0.10	0.91	*	
Type of Encoding	Methodological / Task	Basic	46	13	-	-		-	-	-	-	9.44	39	11	-0.06	Yes	-2.03	0.043	-0.13	0.00	*	29.74
		Deep vs Basic			-0.42	Yes	-3.92	<.0001	-0.63	-0.21	***				-0.35	Yes	-3.03	0.002	-0.58	-0.12	**	
Task / Cognitive Demand	Methodological / Task	High	94	16	-0.09	-	-2.33	0.020	-0.16	-0.01	*	0.50	87	14	-0.09	-	-2.65	0.008	-0.15	-0.02	**	5.72
Sample Country's global positionality	Methodological / Task	Global North	91	15	-0.10	Yes	-3.08	0.002	-0.17	-0.04	**	4.86	84	13	-0.10	Yes	-3.64	0.000	-0.16	-0.05	***	24.41
Presence of motivation manipulation or instructions	Methodological / Task	No	64	14	-0.08	Yes	-2.10	0.036	-0.16	-0.01	*	EMV	57	12	-0.09	Yes	-2.99	0.003	-0.15	-0.03	**	34.33
Task	Methodological / Task	AFC	94	16	-0.50	-	-3.21	0.001	-0.81	-0.20	**	28.86	87	14	-	-	-	-	-	-	-	3.67

			Model 2 (Original - All)										Model 1 (Original Redux)									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	z	p	LB	UB	Sig	Prop. Variance reduction	No. effects	No. studies	Est.	Expected	z	p	LB	UB	Sig	Prop. Variance reduction
		CFMT vs AFC			0.33	-	1.98	0.047	0.00	0.66	*				-	-	-	-	-	-	-	
		Delayed Matching vs AFC			0.44	-	2.43	0.015	0.08	0.79	*				-	-	-	-	-	-	-	
		Old-New vs AFC			0.44	-	2.77	0.006	0.13	0.76	**				-	-	-	-	-	-	-	
		CFMT			-	-		-	-	-	-				-0.18	-	-3.08	0.002	-0.29	-0.06	**	
Face format at encoding	Methodological / Task	Dynamic / Multiple View Points	-	-	-	-		-	-	-	-	-	57	12	-0.23	Yes	-2.09	0.036	-0.45	-0.02	*	29.54
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic nationality	Methodological / Task	Majority-Minority	-	-	-	-		-	-	-	-	-	77	10	-0.07	Yes	-2.44	0.015	-0.13	-0.01	*	51.98
Published	Methodological / Task	No	-	-	-	-		-	-	-	-	-	87	14	-0.14	No	-2.41	0.016	-0.26	-0.03	*	-8.53

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance

Appendix AAG

Significant Moderators for Quantity of Outgroup Contact – Outgroup Discriminability (d-prime)

			Model 2 (Original - All)										**Model 1 (Original Redux)**									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction
Type of encoding	Methodological / Task	Basic	45	20	0.13	Yes	3.32	0.001	0.05	0.21	***	73.21	43	19	0.13	Yes	3.16	0.002	0.05	0.21	**	9.13
		Basic & Deep vs Basic			-1.02	No	-5.85	<.0001	-1.36	-0.68	***		-	-	-	-	-	-	-	-	-	-
Sample Country's global positionality	Methodological / Task	Global South vs Global North	56	21	0.25	Yes	2.14	0.033	0.02	0.49	*	2.23	54	20	0.23		2.40	0.016	0.04	0.42	*	EMV
Presence of motivation manipulation or instructions	Methodological / Task	None	58	21	0.11	Yes	2.96	0.003	0.04	0.18	**	74.23	56	20	0.11	Yes	2.91	0.004	0.03	0.18	**	13.19
		Mixed vs None			-1.00	No	-5.75	<.0001	-1.34	-0.66	***					-	-	-	-	-	-	
Length of delay (categories)	Methodological / Task	Brief	-	-	-	-	-	-	-	-	-	-	56	20	0.10	Yes	2.46	0.014	0.02	0.18	*	EMV
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Majority-Minority	-	-	-	-	-	-	-	-	-	-	33	17	0.13	Yes	2.15	0.031	0.01	0.24	*	EMV
Task / Cognitive Demand	Methodological / Task	High	-	-	-	-	-	-	-	-	-	-	56	20	0.11	-	2.83	0.005	0.03	0.18	**	5.54

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance

Appendix AAH

Significant Moderators for Quantity of Outgroup Contact – Outgroup Hits

			Model 2 (Original - All)										**Model 1 (Original Redux)**									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction
Implicit prejudice for outgroup members (combined option one)	Core	-	13	6	-0.42	Yes	-2.19	0.029	-0.80	-0.04	*	80.80	-	-	-	-	-	-	-	-	-	-
Harvard's 'Project Implicit' values for outgroup implicit prejudice	Core	-	13	6	-0.42	Yes	-2.19	0.029	-0.80	-0.04	*	80.80	-	-	-	-	-	-	-	-	-	-
Length of encoding (categories)	Methodological / Task	Brief	5	4	-0.38	Yes	-2.35	0.019	-0.69	-0.06	*	100.00	-	-	-	-	-	-	-	-	-	-
		Longer vs Brief			0.59	Yes	3.29	0.001	0.24	0.94	**		-	-	-	-	-	-	-	-	-	-
		Short vs Brief			0.43	Yes	2.56	0.011	0.10	0.76	*		-	-	-	-	-	-	-	-	-	-
		Longer	-	-	-	-	-	-	-	-	-	-	4	3	0.21	Yes	2.65	0.008	0.06	0.37	***	94.59
Number of lures (new faces) present during testing	Methodological / Task	-	6	5	-0.01	Yes	-2.48	0.013	-0.01	0.00	*	48.53	-	-	-	-	-	-	-	-	-	-
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Minority-Majority vs Majority-Minority	13	6	0.25	Yes	2.06	0.039	0.01	0.49	*	20.87	-	-	-	-	-	-	-	-	-	-
Face format at encoding	Methodological / Task	Dynamic / Multiple Viewpoints	-	-	-	-	-	-	-	-	-	-	5	4	0.21	Yes	2.65	0.008	0.06	0.37	**	83.59
Sample Country's global positionality	Methodological / Task	Global North	-	-	-	-	-	-	-	-	-	-	13	6	0.10	Yes	2.44	0.015	0.02	0.18	*	EMV

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance

Appendix AAI

Significant Moderators for Quality of Outgroup Contact – Identification Difference Scores

			Model 2 (Original - All)										**Model 1 (Original Redux)**									
Moderator	Category	if categorical, levels	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction	No. effects	No. studies	Est.	Expected	**z**	p	LB	UB	Sig	Prop. Variance reduction
Task / Cognitive Demand	Methodological / Task	High	36	3	-0.04	Yes	-2.59	0.010	-0.07	-0.01	**	EMV	-	-	-	-	-	-	-	-	-	-
Sample Country's global positionality	Methodological / Task	Global North	36	3	-0.04	Yes	-2.79	0.005	-0.07	-0.01	**	EMV	-	-	-	-	-	-	-	-	-	-
Presence of motivation manipulation or instructions	Methodological / Task	None	36	3	-0.04	Yes	-2.60	0.009	-0.06	-0.01	**	2.07	-	-	-	-	-	-	-	-	-	-
Positionality of participants (in-group) to out-group targets relative to sample country and ethnic-nationality	Methodological / Task	Majority-Minority	-	-	-	-	-	-	-	-	-	-	28	2	-0.14	Yes	-2.28	0.023	-0.26	-0.02	*	EMV

Note. Abbreviated column headings: Est = estimate, LB = 95% lower bound confidence interval, UB = Upper bound confidence interval, Sig = Level of significance, ED= Expected direction, Var (%) = Proportional reduction in total variance. Level of significance is indicated via *<.05, **<.01, ***<.001.

The proportional reduction in variance calculation, may yield a negative value indicating the variance explained in the model including a moderator, exceeded the total variance explained in the meta-analytic model. Such instances are the result of excluding missing moderator data and in turn reducing the comparative sample size. Such cases are noted as ‘EMV’ or exceeds model variance