GROUP EVALUATIONS AMONG MEMBERS OF DISADVANTAGED GROUPS

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften

an der Universität Hamburg

Fakultät für Psychologie und Bewegungswissenschaften, Institut für Psychologie

vorgelegt von

Dipl.-Psych. Iniobong Essien

Hamburg, Juni 2020

Datum der Disputation: 16. Juli 2020

Mitglieder des Promotionsprüfungsausschusses:

Vorsitz: Prof. Jenny Wagner

1. Dissertationsgutachterin: Prof. Juliane Degner-Premraj

2. Dissertationsgutachter: Prof. Dirk Wentura

1. Disputationsgutachter: Prof. Kai Epstude

2. Disputationsgutachter: Prof. Jan Wacker

ABSTRACT

Members of disadvantaged groups sometimes display outgroup favoritism, an evaluative preference for an outgroup over the ingroup. Some scholars argue that such outgroup favoritism is common among members of disadvantaged groups, and that it is caused by an internalization of negative stereotypes. The present dissertation critically reflects upon these ideas. Specifically, this work argues that empirical findings regarding group evaluations among members of disadvantaged groups are more mixed than previously acknowledged, and that this heterogeneity is possibly due to moderators. The first study investigated moderators of group evaluations, derived from system justification theory (SJT; Jost, Banaji, & Nosek, 2004). We used a meta-analytic approach with large samples of online participants, spanning 8 social identities and 14 nations, and including Implicit Association Tests (IAT; Greenwald, McGhee, & Schwartz, 1998) and self-report measures as dependent variables. We observed that disadvantaged groups displayed outgroup favoritism on the IAT, but no group preference or ingroup favoritism on self-report measures. However, effects were highly heterogeneous and exploratory moderator analyses revealed that *social identity* moderated group evaluations in disadvantaged groups: Whereas some disadvantaged groups always displayed ingroup favoritism, other disadvantaged groups always displayed outgroup favoritism, and yet others displayed divergent patterns on IATs and self-report measures. Furthermore, group-based *stigma* and self-reported *conservatism* moderated group evaluations. A second set of studies investigated *ingroup typicality* as another potential moderator of group evaluations. We assumed that members of disadvantaged groups who perceive themselves as less typical for their ingroup may be more likely to demonstrate outgroup favoritism. In Study 1 and 2, Black participants with lighter skin tone more strongly

preferred White relative to Black people. In Study 3, heavyweight participants with lower body weight more strongly preferred thin relative to heavyweight people. In Study 4, participants with less visible disabilities more strongly preferred non-disabled relative to disabled people. A meta-analysis across studies estimated an overall small effect of ingroup typicality on group evaluations. A third set of studies investigated procedural differences between measures as another potential moderator of group evaluations. In Study 1, Turkish-German participants displayed a preference for Turkish relative to German on two IAT variants, but no preference for either group on feeling thermometers. In Study 2, Muslim participants displayed preferences for Arabs and Muslims relative to Whites on two variants of the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) and on self-report measures, suggesting ingroup favoritism. Muslim participants also created classification images of a typical Muslim via a reverse correlation task. Compared with classification images created by a control sample, those created by Muslim participants were rated higher in trustworthiness, but did not differ on other dimensions. Taken together, findings do not suggest ubiquitous patterns of outgroup favoritism, but highlight the importance of moderators for our understanding of group evaluations among members of disadvantaged groups.

ACKNOWLEDGEMENTS

I cannot underestimate my sincere gratitude for Prof. Juliane Degner, who provided wise guidance, feedback, and support throughout this endeavor. Thank you for being such a great mentor, for keeping your door open, and for your kindness.

I would also like to express my appreciation to the members of the examining committee, Prof. Kai Epstude, Prof. Jan Wacker, Prof. Jenny Wagner, and in particular, Prof. Dirk Wentura, who supported the PhD project at various stages throughout the years.

Furthermore, I would like to express my genuine gratitude to the students and student research assistants who supported this project, especially Imen Jelassi, Elisabeth Kreider, Lisa Krichel, Tanja Oschatz, Marta Plonka, and Habiba Schiller.

To my family and especially my mum, for her love, trust, and patience. To my friends, to Doris, Janus, Jakob, Lisi, and especially to Nikolas for so much.

My deep and heartfelt gratitude goes to Marleen Stelter, for her intellectual engagement, for countless insightful conversations, and for her support and unwavering encouragement. I could not wish for a better travel mate on this journey.

And to David, who means the world to me.

TABLE OF CONTENTS

"It is a peculiar sensation, this double-consciousness, this sense of always looking at one's self through the eyes of others, of measuring one's soul by the tape of a world that looks on in amused contempt and pity." (Du Bois, 1897)

CHAPTER 1: INTRODUCTION

Civil-rights icon Jesse Jackson was once famously quoted with the words *"There is nothing more painful to me at this stage in my life than to walk down the street and hear footsteps and start thinking about robbery — then look around and see somebody white and feel relieved"* (Herbert, 1993). He expressed these words at a time when the topic of crime—that is, violent crime commited by young Black men—was at the forefront of the public debate in the United States (Coates, 2015). At the same time, these words pose troubling questions regarding the possible consequences of societal stereotyping of disadvantaged groups: Why would a Black person feel relieved when seeing a White person at night? How would he feel if the follower was Black? Could it be that Jesse Jackson fears Black people? Has he internalized an image of young Black men as threatening and dangerous? More broadly, these questions point to the issue of how society may influence how we think and feel about the groups to which we belong. People are not independent of the judgments of those around them (e.g., Asch, 1951), so could the views of others also affect how people judge their ingroup? One aim of the present dissertation is to examine whether or to what extent members of disadvantaged groups share society's negative beliefs and feelings about their social group.

Ideas about the possible consequences of societal stereotyping and discrimination for those who are targeted have been expressed quite early (e.g., Du Bois, 1897), in different scientific disciplinces (e.g., sociology, psychoanalytic theorizing; Fanon, 1986; Pyke, 2010), and have also permeated popular discourses and writing (e.g., Thurman, 2015). It has been argued, for example, that those who are stereotyped might suffer from an inferiority complex (Fanon, 1986), experience self-hatred (Lewin, 1948), might distance themselves from the devalued

ingroup (Allport, 1954; Tajfel & Turner, 1979), or might display outgroup favoritism (Jost et al., 2004).

One focus of the present dissertation is *outgroup favoritism*—the evaluative preference for the outgroup over the ingroup. The relevance of this topic becomes clear when we think about which outcomes might be affected by outgroup favoritism (see Dasgupta, 2004, for a review). Previous research has linked outgroup favoritism and the endorsement of (negative) ingroup stereotypes to health problems (e.g., Levy, Hausdorff, Hencke, & Wei, 2000; Levy, Zonderman, Slade, & Ferrucci, 2009), such as depressive symptoms (Mouzon & McLean, 2017), maladaptive health-related behaviors (Carels et al., 2010), or cardiovascular events (Krieger et al., 2010; Levy et al., 2009). In addition, studies suggest that outgroup favoritism and negative ingroup stereotypes may negatively impact psychological adjustment and performance in academic domains (e.g., Ashburn-Nardo, Knowles, & Monteith, 2003; Crocker & Major, 1989; Jellison, McConnell, & Gabriel, 2004; Keller, 2002; Major & O'Brien, 2005; Spicer, 1999; Steele, 1997).[1]

[1] Note that a large body of work has also investigated the role of *stereotype threat* (Steele & Aronson, 1995) and its impact on academic performance in members of disadvantaged groups (see Lamont, Swift, & Abrams, 2015; Spencer, Logel, & Davies, 2016, for reviews and meta-analyses). However, the robustness of the effect is currently debated (e.g., Flore & Wicherts, 2015) and a recent large-scale experiment did not demonstrate evidence of stereotype threat (Flore, Mulder, & Wicherts, 2019).

Together, this body of research suggests that outgroup favoritism and the endorsement of negative stereotypes are relevant for both health and psychological adjustment.

One popular perspective on group evaluations among members of disadvantaged groups is System Justification Theory (SJT; Jost et al., 2004), which proposes that members of disadvantaged groups internalize societal negativity directed at their group and, in turn, display (implicit) outgroup favoritism. An important aim of the present dissertation is to critically reflect upon the idea that outgroup favoritism is a relatively general or common phenomenon among members of disadvantaged groups. More specifically, this work argues that the current state of empirical findings regarding group evaluations among members of disadvantaged groups is more mixed than is acknowledged by prominent theorizing (Jost et al., 2004), and that this heterogeneity is possibly due to moderating factors. These moderating factors include: aspects of the intergroup context; society- and system-based processes; group-based processes; self-based processes; and procedural aspects of the measurement. These moderating factors of group evaluations complicate assumptions about a general pattern of outgroup favoritism among members of disadvantaged groups and point to the need for more nuanced theorizing in this domain.

Theoretical Perspectives on Ingroup Favoritism and Outgroup Favoritism in Disadvantaged Groups

The following review will focus on relevant theoretical perspectives in social psychology, detailing whether and to what extent group evaluations among members of disadvantaged groups might differ from those who are advantaged. This review has two aims. The first aim is to

synthesize from extant theories whether members of disadvantaged groups are expected to display ingroup favoritism or outgroup favoritism. The second aim is to describe psychological and contextual factors potentially moderating group evaluations among members of disadvantaged groups.

System Justification Theory

System Justification Theory (SJT; Jost & Banaji, 1994) was introduced in part to account for the fact that negative stereotypes of disadvantaged groups are often shared across group boundaries, and sometimes even endorsed by members of the afflicted disadvantaged group. SJT is based on the premise that people are motivated to accept the status quo as legitimate and fair (Jost et al., 2004). In its first formulation, SJT (Jost & Banaji, 1994) proposes that members of disadvantaged groups participate in perpetuating the existing, and often unjust, social hierarchy, by adopting negative stereotypes that justify their disadvantaged position. For example, disadvantaged group members might explain their disadvantaged position through internal attributions (e.g., low competence or low effort), thus blaming themselves instead of appealing to discriminatory treatment or to societal inequality (e.g., poor access to resources). According to a more recent formulation of SJT (Jost et al., 2004), members of disadvantaged groups internalize culturally communicated negative stereotypes, potentially leading to outgroup favoritism, a process Jost and colleagues termed "false consciousness"—endorsing beliefs that are detrimental to the self or the ingroup. Importantly, outgroup favoritism is not only expected to occur in stereotype- or status-related domains (e.g., achievement, intelligence), but also in terms of evaluations and affective responses (e.g., the liking of the ingroup), as well as in behavioral tendencies (e.g., the preference to interact with an advantaged outgroup member over interacting

with a disadvantaged ingroup member; Jost et al., 2004). In sum, whereas earlier formulations of SJT (Jost & Banaji, 1994) propose that members of disadvantaged groups endorse negative ingroup stereotypes, more recent formulations of SJT (Jost et al., 2004) propose that the internalization of cultural stereotypes leads to outgroup favoritism among members of disadvantaged groups.

Crucially, SJT (Jost et al., 2004; Jost, Pelham, & Carvallo, 2002) argues that members of disavantaged groups may be largely unaware or unconscious about their attitudes that favor advantaged outgroups over their own group (i.e., outgroup favoritism; Jost et al., 2002). Jost and colleagues (Jost et al., 2004, 2002) provided a variety of speculations for their notion of unconscious outgroup favoritism among members of disadvantaged groups. For example, they argued that attitudes that favor advantaged groups become "over-learned and unconscious" during socialization (Jost et al., 2002, p. 588). According to this view, stereotypes are repeatedly learned in childhood until they operate outside of awareness, and thus "become imperceptible — like water to the fish" (Jost, 2001, p. 91).

Jost and colleagues (Jost et al., 2002) also reasoned that outgroup favoritism may cause cognitive dissonance among members of disadvantaged groups, because such attitudes are not consistent with consciously held beliefs about the ingroup. Furthermore, outgroup favoritism may induce negative emotions (Jost et al., 2004, 2002). Lastly, Jost and colleagues (2004) speculated that members of disadvantaged groups often feel normative pressures within their communities to display ingroup pride. Such normative pressures among disadvantaged groups would foster open displays of ingroup favoritism among group members. Taken together, SJT (Jost et al., 2004, 2002) assumes that members of disadvantaged groups internalize attitudes that favor advantaged

groups, but avoid explicit displays of outgroup favoritism due to cognitive dissonance, emotional discomfort, or perceived normative pressures. Yet, while members of disadvantaged groups consciously reject outgroup favoritism, they might be less able to reject outgroup favoritism at a level that is introspectively not accessible or "unexamined" (Jost et al., 2002, p. 588). Therefore, SJT proposes that outgroup favoritism is more likely to be observed on implicit or non-obtrusive measures, as compared to explicit measures (Jost et al., 2004).[2]

---

[2] Note that Jost and colleagues (Jost et al., 2004, 2002) seem to interpret implicit measures as measures of unconscious cognition. However, there is ongoing debate about whether implicit measures tap into unconscious cognitions (e.g., Gawronski & Bodenhausen, 2006; Hahn & Gawronski, 2014; Hahn, Judd, Hirsh, & Blair, 2013), and challenges of conflation and lack of process purity of implicit measures (see subchapter on current debates regarding the "implicit" concept) become relevant in this context. Claims that the IAT might be interpreted as a measure of unconscious cognition are difficult to reconcile with current understandings of implicit measures generally, and the IAT specifically. First, implicit measures are not process-pure and inferences regarding features of automaticity based solely on the type of task (i.e., the IAT) are unwarranted (e.g., Sherman, 2009). Second, the notion that people are unaware about mental content captured by the IAT stands in contrast to empirical evidence. To investigate whether people have introspective access into their evaluations assessed by the IAT, Hahn and colleagues (Hahn et al., 2013; Rivers & Hahn, 2018) asked participants to predict their IAT scores. They found that participants were indeed able to accurately predict their IAT scores. This line of

To support the notion of outgroup favoritism on implicit measures, Jost and colleagues (Jost et al., 2004) mostly point to studies that used the Implicit Association Test (IAT; Greenwald et al., 1998). In one study, for example, Black participants provided evaluations of Black Americans and White Americans on a self-report measure of perceived warmth and completed an evaluative IAT (Livingston, 2002). Whereas participants displayed ingroup favoritism on the self-report measure, they displayed outgroup favoritism on the IAT, a finding which is consistent with SJT (Jost et al., 2004). However, whereas some studies suggest that members of disadvantaged groups on average display outgroup favoritism on implicit measures (e.g., Ashburn-Nardo et al., 2003; Nosek, Banaji, & Greenwald, 2002), other studies document ingroup favoritism (e.g., Livingston, 2002, Experiment 1b), and yet other studies suggest no group preference in either direction (e.g., Jost et al., 2002; Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002).

In addition to the hyothesis that outgroup favoritism registers on implicit rather than explicit measures, SJT hypothesized that a number of factors moderate outgroup favoritism among members of disadvantaged groups. Outgroup favoritism among members of

---

research suggests that people have introspective access into mental content captured by the IAT and are able to report on them. Taken together, current understandings of implicit measures and evidence by Hahn and colleagues (Hahn et al., 2013; Rivers & Hahn, 2018) are at odds with the notion that members of disadvantaged groups are unconscious of the mental content captured by the IAT.

disadvantaged groups is assumed to be more likely to the extent that they: perceive the system as legitimate; endorse system-justifying; or endorse conservative beliefs. Lastly, a common interpretation of SJT is that members of disadvantaged groups are expected to display outgroup favoritism to the extent that they a stigmatized (Lane, Mitchell, & Banaji, 2005; Livingston, 2002; Rudman, Feinberg, & Fairchild, 2002, and @Dasgupta.2004). The empirical evidence regarding SJT's proposed moderators of outgroup favoritism is mixed, with studies documenting inconsistent effects of conservatism (e.g., Hoffarth & Jost, 2017; Jost et al., 2004) and system justifying beliefs (e.g., Jost & Thompson, 2000; Jost et al., 2002) on outgroup favoritism among disadvantaged groups. Furthermore, although stigma seems to be related to outgroup favoritism among members of disadvantaged groups, studies document inconsistent effects for implicit and explicit measures (Livingston, 2002; Rudman et al., 2002).

Taken together, SJT (Jost et al., 2004) hypothesizes a discrepancy between (favorable) explicit and (unfavorable) implicit evaluations: While members of disadvantaged groups may explicitly display positive views, they implicitly harbor negative attitudes regarding their own groups. Furthermore, SJT argues that outgroup favoritism among disadvantaged groups is moderated by perceptions of legitimacy of the social system, by the endorsement of system justifying and conservative beliefs, and by stigma.

Social Identity Theory

The arguably most prominent account with implications for ingroup and outgroup favoritism among disadvantaged groups comes from Social Identity Theory (SIT; Tajfel & Turner, 1979). SIT is based on the assumption that people generally want to feel good about themselves; they want to maintain or achieve a positive view of themselves. Besides striving for a

positive self-concept, people also belong to social groups. Yet, different social groups are differentially evaluated by society. Some social groups are evaluated more positively, whereas other social groups are evaluated less positively (or even negatively). Thus, SIT presumes that status hierarchies exist within society: Depending on their social group memberships, people tend to have more or less positive (or negative) social identities.

Based on these assumptions, SIT proposes that people are motivated to achieve or preserve positive social identities. People strive for *positive ingroup distinctiveness* in that they want to belong to social groups that are positively evaluated. However, because some social groups are negatively evaluated, members of these groups are thought to resort to a number of strategies to cope with their threatened social identity. One way to achieve a positive social identity is through *social creativity*. Social creativity refers to (strategic) favorable comparisons between the ingroup and socially relevant outgroups. For example, one can: make comparisons to an outgroup to which the ingroup fares well; choose a comparison dimension on which the ingroup excels; redefine the value of a given comparison dimension and so forth. A person may, for example, believe that their group is depicted negatively in terms of competence, leading them to shift the subjective importance to a dimension in which the ingroup does comparably well, such as warmth (*change of evaluative dimension*; e.g., Oldmeadow & Fiske, 2010). Or the person might reject the validity of negative ingroup evaluations by endorsing beliefs such as “Black is beautiful” (i.e., *change of valence of ingroup evaluation*). In sum, social creativity serves the goal to achieve positively distinctive social identities.

SIT (Tajfel & Turner, 1979) proposes that positive ingroup distinctiveness is to a great extent based on social creativity. If social creativity is not a viable option, however, people may

resort to *individual mobility* to achieve positive ingroup distinctiveness. For example, people may dissociate, that is, disidentify, from or leave the group in order to join a group with better reputation (i.e., an advantaged outgroup). Such individual mobility is largely determined by the permeability of intergroup boundaries. In short, based on SIT, to the extent that people are motivated and/or are able to achieve positive distinctiveness through social creativity, they are expected to display ingroup favoritism.

Over the past decades, SIT has informed a vast number of empirical studies and its basic tenets as well as assumptions regarding social creativity and social mobility have been supported (Bettencourt, Charlton, Dorr, & Hume, 2001; Jackson, Sullivan, Harnish, & Hodge, 1996; Mullen, Brown, & Smith, 1992; van Zomeren, Postmes, & Spears, 2008). A meta-analysis combining 42 studies documented that all investigated social groups—real and artificial—display ingroup favoritism. Nonetheless, the strength of ingroup favoritism varied as a function of group status "depending on the reality of group categorization" (Mullen et al., 1992, p. 116). Among real groups, disadvantaged and advantaged groups did not differ in terms of their level of ingroup favoritism; only among artificially created groups, ingroup favoritism was weaker in disadvantaged experimental conditions than in advantaged experimental conditions. Although these results were only observed in artificially created groups, they add a caveat to the assumption of ingroup favoritism as an almost general principle, as it can be derived from SIT. *Group status*, thus, may be a relevant factor potentially moderating the extent to which people favor their ingroup.

The reviewed theorizing and empirical findings suggest that membership in advantaged groups may provide better chances to achieve positive ingroup distinctiveness. However, a

number of reasons may explain why members of disadvantaged groups do often not resort to individual mobility strategies, such as disidentification with the ingroup. Some group memberships are inescapable in the real world (Wright & Lubensky, 2009). For example, some phenotypic group characteristics, such as skin tone, may create boundaries that are difficult or impossible to cross. Another possible explanation is that identification with a disadvantaged group may provide the necessary fuel in the fight for equal status (Doosje, Spears, & Ellemers, 2002; Rubin, Badea, & Jetten, 2014; Tajfel & Turner, 1979; van Zomeren et al., 2008) and motivate people to act on behalf of their group (e.g., Wright & Lubensky, 2009). It may thus paradoxically be the striving for status that motivates members of disadvantaged groups to affiliate with their ingroups.

Rejection Identification Model

More recent theorizing also suggests that ingroup identification with a disadvantaged group can serve a compensatory function, like a "social cure" (Greenaway et al., 2015, p. 53). In providing members of disadvantaged groups with a sense of belonging and support, social identification is thought to have beneficial effects for well-being and health (Greenaway et al., 2015). In a similar vein, the Rejection-Identification Model (RIM; Branscombe, Schmitt, & Harvey, 1999) understands social identification among members of disadvantaged groups as resulting from perceptions of group disadvantage and inequality. The RIM proposes that members of disadvantaged groups who tend to attribute negative life-outcomes to discrimination also tend to feel hostility towards dominant groups. According to the RIM, members of disadvantaged groups who perceive pervasive discrimination are motivated to identify even stronger with their ingroup. Thus, while perceptions of discrimination are thought to foster

negative outgroup attitudes and decrease personal well-being, the same (negative) perceptions are thought to increase ingroup identification. Thus, the RIM model proposes that ingroup identification may provide a strategy to buffer against social identity threats, eventually fostering affiliation and attachment with fellow ingroup members. Given that stronger ingroup identification tends to coincide with more positive evaluations of the ingroup (cf. Branscombe et al., 1999; Brewer, 1999), the RIM allows for the somewhat counter-intuitive prediction that higher perceived discrimination might increase ingroup favoritism among disadvantaged groups. In short, the RIM conceptualizes social identification in disadvantaged groups as a compensatory process resulting from perceptions of discrimination. Thus, perceiving pervasive discrimination might paradoxically cause members of disadvantaged groups to display ingroup favoritism. Empirical studies investigating the proposed processes in advantaged and disadvantaged groups support the basic tenets of the RIM (Branscombe et al., 1999; Ramos, Cassidy, Reicher, & Haslam, 2012; for a review see also Schmitt, Branscombe, Postmes, & Garcia, 2014).

Social Dominance Theory

Social Dominance Theory (SDT; Pratto, Sidanius, & Levin, 2006; Sidanius & Pratto, 1999) is based on the premise that human societies are organized in group-based hierarchies. These group-based hierarchies provide conditions under which some (advantaged) groups enjoy a variety of privileges—e.g., wealth; good education and health—whereas other (disadvantaged) groups face a variety of negative outcomes—e.g., poverty; incarceration; poorer health. SDT was developed to explain how societal hierarchies are established and maintained. One of the mechanisms identified by SDT, which is assumed to contribute to the maintenance of societal hierarchies, and which is especially relevant in the context of the present dissertation, is the

concept of *legitimizing myths* (Pratto et al., 2006). Legitimizing myths are cultural narratives—based on beliefs, stereotypes, ideologies, values etc.—many of which are aimed at justifying the dominance of advantaged groups over disadvantaged groups. Thus, legitimizing myths serve the function of justifying social hierarchies and inequality, thereby suppressing intergroup conflict and keeping societal stability (Pratto, Sidanius, Stallworth, & Malle, 1994). Legitimizing myths are widely shared in society, and are assumed to shape decision-making and behavior at the individual level (e.g., individual acts of discrimination) as well as at institutional levels (e.g., policies; laws).

Examples of legitimizing myths are: the attribution of negative societal outcomes (e.g., poverty) to internal person or group characteristics (e.g., laziness); stereotypes that justify the harsher treatment of ethnic groups in terms of criminal justice; beliefs entailing that people get what they deserve (e.g., Lerner, 1980); values or ideologies emphasizing that success is the result of individual merit and hard work (e.g., meritocratic beliefs). What these cultural narratives have in common is that they provide a justification for why groups in society are where they are in terms of status and power. Legitimizing myths thus provide "moral and intellectual justification" (Pratto et al., 2006, p. 275) for the lower status and power of disadvantaged groups (e.g., "because they did not work hard enough") and the higher status and power of advantaged groups (e.g., "because they place higher importance on education").

Importantly, SDT (Pratto et al., 2006; Sidanius & Pratto, 1999) posits that legitimizing myths are often agreed upon across group boundaries, by members of both advantaged and disadvantaged groups (Pratto et al., 2006). Especially in societies with stable social hierarchies, advantaged and disadvantaged groups are more likely to agree about legitimizing myths than to

disagree. Consistent with the notion that members of disadvantaged groups tend to endorse legitimizing myths, SDT (Pratto et al., 2006; Sidanius & Pratto, 1999) proposes an *asymmetry in ingroup favoritism* between disadvantaged and advantaged groups. According to SDT, ingroup favoritism varies as a function of power and status. Members of disadvantaged groups display less ingroup favoritism than members of advantaged groups, because it is more effortful for them than for their advantaged counterparts. In extreme cases, SDT (Pratto et al., 2006; Sidanius & Pratto, 1999) posits, that this asymmetry might even register as outgroup favoritism among members of disadvantaged groups.

Relatedly, SDT proses that ingroup favoritism is moderated by the endorsement of legitimizing myths. More specifically, to the extent that members of advantaged groups endorse legitimizing myths—which often work in favor of their group—they also display ingroup favoritism (Pratto et al., 2006). Conversely, although not explicitly stated by SDT, to the extent that members of disadvantaged groups endorse legitimizing myths—which may involve negative ingroup stereotypes—they might display lower levels of ingroup favoritism. Thus, the endorsement of legitimizing myths increases the asymmetry in ingroup favoritism between advantaged and disadvantaged groups (Pratto et al., 2006).

Taken together, according to SDT (Pratto et al., 2006; Sidanius & Pratto, 1999) members of disadvantaged groups are expected to display lower levels of ingroup favoritism than members of advantaged groups, and in extreme cases even outgroup favoritism. Furthermore, SDT proposes that group status, power, and the endorsement of legitimizing myths moderate ingroup favoritism among disadvanged groups.

Self Anchoring

Another relevant perspective with implications for group evaluations among members of disadvantaged groups is the self-anchoring account (Cadinu & Rothbart, 1996). The self-anchoring account provides a framework about how people form psychological connections with social groups. Whereas SIT proposes that people ascribe characteristics of the ingroup to the self (i.e., *self stereotyping*; Tajfel & Turner, 1979), the self-anchoring account (Cadinu & Rothbart, 1996) provides a *bottom-up explanation* for links between the self and the ingroup. Here, people form psychological connections with social groups by applying characteristics of the self to the group, eventually resulting in ingroup favoritism. More specifically, given that most people have a positive image of themselves (Greenwald, 1980; Taylor & Brown, 1988), things that are associated with the self are also positively evaluated. Self anchoring is assumed to be a rather spontaneous or automatic process, whereby positive evaluations of the self *automatically* generalize to things, which are associated with the self (Cadinu & Rothbart, 1996). These things can be objects in the environment (e.g., belongings), but also the social groups and organizations a person belongs to. Thus, according to the self anchoring account, people also display ingroup favoritism based on positive self evaluations (van Veelen, Otten, Cadinu, & Hansen, 2016).[3]

---

[3] However, there is still an ongoing debate about whether self anchoring equally applies to different contexts (e.g., organizational contexts; intergroup contexts) or social groups (e.g., disadvantaged vs. advantaged groups; see Latrofa, Vaes, Cadinu, & Carnaghi, 2010), and whether this pathway to ingroup favoritism might be especially likely when people have

Adding to our understanding of processes moderating group evaluations, the self anchoring account suggests that inter-individual differences in self evaluations (e.g., self-esteem) should be related to ingroup favoritism (van Veelen et al., 2016). Thus, it would seem plausible that group evaluations among members of disadvantaged groups should also be moderated by differences in self evaluations. Taken together, according to the self-anchoring account (Cadinu & Rothbart, 1996), the self serves as a reference point for evaluation of social groups and ingroup favoritism follows from applying positive characteristics of the self to the group.

Synthesizing the Ideas from Theoretical Perspectives on Intergroup Processes

The review of social psychological perspectives on intergroup processes highlights that it is unlikely that members of disadvantaged groups display uniform patterns of outgroup favoritism. Instead, a more nuanced picture emerges, suggesting that a variety of factors potentially moderate group evaluations among members of disadvantaged groups. These factors include: aspects of the *intergroup context*, such as the permeability of group boundaries, group status, and power (Sidanius & Pratto, 2004; Tajfel & Turner, 1979); *society- and system-based processes*, such as the acceptance of legitimizing myths (Sidanius & Pratto, 1999) or the endorsement of system justifying and conservative beliefs (Jost et al., 2004); *group-based processes*, such as social identification, social creativity, individual mobility (Tajfel & Turner,

---

relatively few knowledge about their (soon to-be) ingroup (see van Veelen, Otten, & Hansen, 2011; van Veelen et al., 2016).

1979), or perceptions of discrimination (Branscombe et al., 1999); *self-based processes*, such as inter-individual differences in self-esteem (Cadinu & Rothbart, 1996).

Together, the reviewed social psychological perspectives suggest that it is unlikely that all members of all disadvantaged groups display outgroup favoritism all the time. Besides documenting whether members of disadvantaged groups display ingroup or outgroup favoritism, it is thus important to examine moderating factors of group evaluations.

The Present Dissertation

The present dissertation critically reflects on the notion that members of disadvantaged groups tend to display outgroup favoritism on implicit measures (Jost et al., 2004). An aim of this work was thus to examine whether disadvantaged groups indeed display a uniform pattern of implicit outgroup favoritism. We investigated this idea in a meta-analysis of online studies, by re-analyzing data from Project Implicit and the American National Election Studies, and in a series of lab studies in the German context (Chapter 2, 3, and 4). In addition, as the previous section illustrates, a variety of processes possibly moderate group evaluations among disadvantaged groups. Thus, a second aim was to examine moderators of group evaluations on implicit and explicit measures (Chapter 2, 3, and 4).

The present work examines group evaluations among disadvantaged groups on both implicit and explicit measures of group evaluations. Consequently, the type of measurement used to assess group evaluations (i.e., the distinction between “implicit” and “explicit” measures) plays a central role in the context of the present work. The following section briefly summarizes current debates based on a taxonomy proposed by a recent review (Corneille & Hütter, 2020). An

important goal of this excursion is to disambiguate the use of the term "implicit" in the present work. In addition, this excursion helps clarify what kind of inferences are warranted based on different conceptualizations of the "implicit" concept.

Excursion: Current Debates regarding the "Implicit" Concept

The term "implicit" has acquired different meanings among social-cognitive researchers. For example, "implicit" can refer to *mental processes* (e.g., Gawronski & Bodenhausen, 2006), to *features of automaticity* (e.g., Bargh, 1994; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009), or to *procedural aspects of the measurement* (e.g., De Houwer & Moors, 2010; Gawronski & Hahn, 2019). The section briefly discusses these different perspectives and closes with a definition of the term "implicit" in the context of the present dissertation.

Mental processes. One conceptualization of the term "implicit" refers to the question of how attitudes that underlie participants' evaluations on implicit measures are acquired and represented in memory. The *implicit-as-associative* conceptualization uses the term "implicit" to refer to the mental processes underlying implicit measures (Corneille & Hütter, 2020). For example, some theoretical accounts propose that attitudes are represented in the form of associations in memory (e.g., Fazio, 1990). Other theoretical accounts challenge the notion of mere associations in memory and propose that attitudes are rather represented in the form of propositions (e.g., Houwer, van Dessel, & Moran, 2020). Yet other theoretical accounts, for instance the associative-propositional evaluation model (APE; Gawronski & Bodenhausen, 2006, 2011), propose that implicit evaluations are the outcomes of spontaneous affective reactions. Explicit evaluations, in contrast, are the outcomes of propositional reasoning processes, which do or do not validate spontaneous affective reactions. When propositional reasoning processes

validate affective reactions, implicit and explicit evaluations are assumed to converge; when propositional reasoning processes reject affective reactions, implicit and explicit evaluations are assumed to dissociate. Critically, which affective reactions become activated depends on preexisting associations in memory as well as on the situational context (Gawronski & Bodenhausen, 2006, 2011).

A more recent account, the Bias of Crowds model (BoC; Payne, Vuletich, & Lundberg, 2017), highlights the role of context for implicit evaluations even more. In short, the BoC model proposes that implicit evaluations reflect situational activation of concepts rather than personalized attitudes (i.e., interindividual differences). Payne and colleagues argue that "most of the systematic variance in implicit biases appears to operate at the level of situations" (Payne et al., 2017, p. 236). According to this view, implicit evaluations should vary to a large extent based on situations and places, because some places cue certain concepts (e.g., evaluations; stereotypes) more readily than in others. However, the BoC model is formulated at the macro (i.e., regional or geographical) level and does not specify how attitudes underlying implicit measures are represented. In sum, a variety of theoretical accounts haven been proposed to conceptualize how attitudes that underlie implicit measures are represented in memory. However, there is still considerable debate about the empirical support regarding the different accounts (e.g., Corneille & Hütter, 2020).

Features of automaticity. Another conceptualization of the term "implicit" refers to features of automaticity (e.g., Bargh, 1994). The *implicit-as-automatic* conceptualization uses the term "implicit" to refer to the conditions under which participants provide their responses when completing a measure (Corneille & Hütter, 2020; De Houwer & Moors, 2010). De Houwer and

colleagues define implicit measures as "outcomes of measurement procedures that are caused in an automatic manner by psychological attributes" (De Houwer et al., 2009, p. 347). For example, participants might complete an implicit measure under conditions that limit their ability to exert control over responses. If the to-be-measured attitude affects participants' responses in spite of their attempts to exert control, the measure would classify as automatic (in terms of uncontrollability). Whether an implicit measure classifies as automatic in terms of a specific feature of automaticity needs to be established empirically, and a variety of features of automaticity have been studied, most notably *efficiency*, *awareness*, *intention*, and *control* (e.g., Bargh, 1994; De Houwer et al., 2009).

Implicit-as-automatic conceptualizations face challenges. For example, because different features of automaticity exist, it is imprecise and potentially misleading to use umbrella terms such as "implicit" without further qualifications regarding the specific feature of automaticity (Corneille & Hütter, 2020). Corneille and Hütter also argue that the semantic proximity to concepts such as "unexpressed" or "unrevealed" increase the risk that researchers conflate "implicit" with unconscious processes. Moreover, it is important to highlight that implicit measures are *not process-pure* (Corneille & Hütter, 2020; De Houwer et al., 2009; Gawronski & Hahn, 2019; Sherman, 2009). They are not entirely automatic regarding any feature of automaticity. Instead, automatic, controlled, and construct-unrelated processes jointly contribute to implicit (and explicit) measurement outcomes (e.g., Sherman et al., 2008; Teige Mocigemba, Klauer, & Sherman, 2010). Lastly, it is unclear beyond which threshold of automaticity an implicit measure should classify as automatic (Corneille & Hütter, 2020), which complicates clear-cut differentiations between implicit and explicit measures. Given these limitations,

generalized claims about implicit measures as solely capturing a specific feature of automaticity (e.g., unawareness) should be viewed with caution.

Procedural aspects of the measurement. Another conceptualization of the term "implicit" refers to procedural aspects of the measurement. The *implicit-as-indirect* conceptualization uses the term "implicit" to refer to indirect measurement procedures and their outcomes (e.g., evaluations, stereotypes, bias). This conceptualization does not assume that implicit and explicit measurement outcomes differ with respect to underlying mental representations or features of automaticity. For example, following implicit-as-indirect conceptualizations, it is not warranted to equate implicit measures with unconscious mental processes and explicit measures with conscious mental processes. Instead, the term implicit simply refers to the type of *measurement procedure* used to assess (group) evaluations. According to Corneille and Hütter's (2020) implicit-as-indirect conceptualization, explicit measures refer to procedures where attitudes are inferred from verbal responses of participants (i.e., self-reports); in contrast, implicit measures refer to procedures where evaluations are inferred from observed behavior (e.g., differences in response times and/or errors).

De Houwer and Moors (De Houwer & Moors, 2010) provided a different conceptualization, suggesting that indirect and direct measures should be differentiated based on (a) whether participants provide responses in form of self-assessments, and (b) whether participants' self-assessed responses are the to-be-used outcomes. Following this conceptualization, direct measures of group evaluations are measures where participants provide self-assessments, which are the to-be-used outcomes; they do not require additional steps of interpretation by researchers. In contrast, indirect measures of group evaluations are measures

that require an addtional step of interpretation by the researcher. Here, participants' responses are not directly taken as the to-be-used outcomes; instead, participants' responses (e.g., reaction times, errors) are interpreted by the researcher. This conceptualization deviates from that of Corneille and Hütter (2020) in that self-report measures can also be indirect measures, as long as their outcomes are not directly derived from participants responses. In fact, De Houwer and Moors (2010) note that direct measures can be seen as a "subclass of self-report measures" (p. 183; for an extended discussion of conceptualizations of indirect and direct measures see De Houwer & Moors, 2010). In sum, a variety of definitions have been proposed to distinguish implicit from explicit measures based on the measurement procedure (for a review see also Gawronski & Hahn, 2019).

Besides the question of how to define indirect and direct measures, implicit-as-indirect conceptualizations face other challenges. For example, it has been argued that even research that uses such conceptualizations (e.g., Greenwald & Banaji, 2017) tends to slip into mental-process-oriented interpretations (Corneille & Hütter, 2020). Thus, research that uses implicit-as-indirect conceptualizations needs to take care not to conflate procedural aspects of the measurement with mental processes or features of automaticity. Consequently, implicit-as-indirect conceptualizations face the challenge that "process-free" interpretations cannot inform theories with regard to mental processes (Corneille & Hütter, 2020).

Use of the term "implicit" in the present work. Taken together, social-cognitive and intergroup research uses the term "implicit" with reference to various mental processes and procedural aspects. Implicit-as-associative conceptualizations refer to the mental processes underlying implicit measures (e.g., representations in memory). Implicit-as-automatic

conceptualization refer to the conditions under which participants provide their responses with regard to features of automaticity. Lastly, implicit-as-indirect conceptualizations refer to procedural aspects of the measurement, but are agnostic to mental processes or features of automaticity. The present dissertation uses this latter definition, which refers to indirect measurement procedures and to their outcomes (e.g, evaluations, stereotypes, bias). The aforementioned limitations associated with this definition apply to the present dissertation as well. Consequently, in presenting and discussing findings regarding group evaluations among members of disadvantaged groups, this work refrains from making strong assumptions in terms of underlying mental representations (e.g., associations) or in terms of features of automaticity (e.g., awareness; intentionality; control). Applied to the topic of this dissertation, this implies, for example, that effects on implicit measures are not interpreted as revealing unconscious mental content. Instead, similar to a multimethod approach, different implicit and explicit measures are primarily seen as alternative ways to operationalize group evaluations among disadvantaged groups.

Overview of Chapters

The present dissertation consists of individual publications and manuscripts submitted for publication in peer-reviewed journals. Chapter 2 and 3 have been accepted for publication and Chapter 4 has been submitted for publication:

Chapter 2: Essien, I., Calanchini, J., & Degner, J. (in press). Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *Journal of Personality and Social Psychology*.

Chapter 3: Essien, I., Otten, S., & Degner, J. (in press). Group evaluations as self-group distancing: Ingroup typicality moderates evaluative intergroup bias in stigmatized groups. *European Journal of Social Psychology*.

Chapter 4: Essien, I., & Degner, J. (Manuscript submitted for publication). Do members of disadvantaged groups display increased ingroup favoritism on personalized measures of group evaluations?

Chapter 2 presents a study on moderators of intergroup evaluations in disadvantaged groups, testing hypotheses derived from SJT (Jost et al., 2004), specifically that outgroup favoritism: is more likely to manifest on implicit than on explicit measures; is more likely to the extent that disadvantaged groups endorse conservative beliefs; is more likely to the extent that a group is stigmatized. We used a meta-analytic approach with large samples of online respondents from Project Implicit, spanning 8 social identities and 14 nations, and including IATs and self-report measures as dependent variables. We observed that disadvantaged groups displayed a medium-sized positive effect on the IAT, indicating outgroup favoritism, but average null effects and small negative effects on self-report measures, indicating no group preference or ingroup favoritism. However, effects were highly heterogeneous and additional exploratory moderator analyses revealed that social identity moderated intergroup evaluations in disadvantaged groups: Whereas some disadvantaged groups always displayed ingroup favoritism, other disadvantaged groups always displayed outgroup favoritism, and yet other disadvantaged groups displayed divergent patterns on implicit and explicit measures. These moderation effects by social identity are not easily explained by SJT and thus may suggest a boundary condition. Furthermore, self-reported conservatism moderated implicit and explicit intergroup evaluations. In addition, group-

level analyses suggest that intergroup evaluations among disadvantaged groups were also moderated by group-based levels of stigma. Taken together, findings in Chapter 2 align with predictions by SJT. However, the more complex patterns of intergroup evaluations in disadvantaged groups illustrate a need for further theory development and more theory-driven research into the boundary conditions of SJT.

Chapter 3 presents studies examining ingroup typicality as another potential moderator of group evaluations in disadvantaged groups. We assumed that members of disadvantaged groups who perceive themselves—or are perceived—as less typical for their ingroup and thus more similar to an advantaged outgroup may be more likely to demonstrate outgroup favoritism. In Study 1 and 2, Black participants more strongly preferred light-skinned or White relative to dark-skinned or Black individuals the lighter their skin tone. In Study 3, heavyweight participants more strongly preferred normal-weight relative to heavyweight individuals the lower their self-reported body weight. In Study 4, participants with disabilities more strongly preferred non-disabled relative to disabled individuals the less visible they judged their disability. A meta-analysis across studies yielded a small average effect size, indicating an overall small effect of ingroup typicality on group evaluations.

Chapter 4 presents two studies examining procedural differences between measures as another potential moderator of group evaluations in disadvantaged groups. A secondary aim was to examine whether members of disadvantaged groups indeed display outgroup favoritism on implicit measures, as SJT proposes (Jost et al., 2004). Participants completed different variants of implicit and explicit measures of group evaluations. In Study 1, Turkish-German participants completed feeling thermometers and two variants of the IAT—a standard version and a

personalized version. Participants favored Turkish relative to German, suggesting a preference for a stigmatized identity relative to a non-stigmatized identity on average. However, participants did not display a preference for either group on feeling thermometers. In Study 2, Muslim participants completed self-report measures and two variants of the Affect Misattribution Procedure (AMP; Payne et al., 2005)—a normative AMP and a personalized AMP. Participants favored Arab and Muslim targets relative to White targets on the AMP, suggesting ingroup favoritism on average. Participants also displayed preferences for Arab and Muslim individuals relative to White individuals on self-report measures, also suggesting ingroup favoritism on average. Moreover, Muslim participants in this study created classification images of a typical Muslim via a reverse correlation task. Compared with classification images created by a control sample, those created by Muslim participants were rated higher in trustworthiness and lower in Arab stereotypicality. Together, these studies extend meta-analytic findings in Chapter 2 in that members of disadvantaged groups displayed ingroup favoritism on implicit measures, although their respective ingroups are severely stigmatized in their societal context.

CHAPTER 2: MODERATORS OF INTERGROUP EVALUATION IN DISADVANTAGED GROUPS: A COMPREHENSIVE TEST OF PREDICTIONS FROM SYSTEM JUSTIFICATION THEORY[4]

Abstract

We examined hypotheses proposed by System Justification Theory (SJT; Jost, Banaji, & Nosek, 2004) regarding intergroup evaluation in disadvantaged groups, using large samples of online participants (total N = 715,721), spanning eight intergroup domains and 14 nations. Using a meta-analytic approach, we tested these hypotheses at the individual level (as SJT is generally articulated), as well as at the social group level. Consistent with SJT, individual-level analyses revealed that disadvantaged groups demonstrated outgroup favoritism on IATs (i.e., implicit measures), but demonstrated ingroup favoritism or no intergroup preference on self-report (i.e., explicit) measures. Additionally, these average effects were characterized by high heterogeneity, and follow-up exploratory analyses revealed that intergroup evaluation in disadvantaged groups was moderated by the intergroup domain: Whereas some disadvantaged groups consistently displayed outgroup favoritism (e.g., age; weight), others consistently displayed ingroup favoritism (e.g., sexual orientation; religion), and yet others displayed diverging patterns on implicit and explicit measures (e.g., race; ethnicity). Consistent with SJT, intergroup evaluation on all measures was moderated by self-reported conservatism. Furthermore, the magnitude of

---

[4] This chapter is based on Essien, I., Calanchini, J., & Degner, J. (in press). Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *Journal of Personality and Social Psychology*.

these relationships depended on the level of analysis, with small effects emerging at the individual level and medium-sized effects emerging at the social group level. Social group-level analyses also indicated that intergroup evaluation in disadvantaged groups was moderated by stigma. Overall, these findings support and extend the predictions of SJT, but the relatively complex patterns of intergroup evaluation in disadvantaged groups identified here illustrate a need for further theory development and more theory-driven research in this domain.

Introduction

"Ask yourself what would happen to your own personality if you heard it said over and over again that you were lazy, a simple child of nature, expected to steal, and had inferior blood. Suppose this opinion were forced on you by the majority of your fellow-citizens. And suppose nothing that you could do would change this opinion-because you happen to have black skin." (Allport, 1954, p. 142)

In his classic work *The Nature of Prejudice*, Allport (1954) speculated about the consequences of being socialized in a society with a predominantly negative attitude towards one's own social group. He reasoned that societal devaluation of one's ingroup might—among other things—lead to feelings of insecurity, rejection of ingroup membership, identification with dominant outgroups, self-hate, or even aggression against the ingroup. Similar ideas were formulated by Clark and Clark (1950) who reasoned that Black children, early in their development, become aware of the inferior status position of their ingroup in society, leading to "feelings of inadequacy and inferiority" (Clark & Clark, 1950, p. 350). In their seminal doll studies, they observed that a substantial number of Black children displayed *outgroup favoritism*, preferring White dolls over Black dolls. More recently, System Justification Theory (SJT; Jost et

al., 2004) posited that such feelings of inferiority are nursed by a general motive to accept the current state of affairs, which in turn gives rise to *false consciousness*: an internalization of stigma by members of disadvantaged groups.

The present research investigates intergroup evaluations across a wide range of disadvantaged groups[5]. Specifically, we provide a large-scale test of hypotheses proposed by Jost and colleagues (2004), who suggested that intergroup evaluation in disadvantaged groups should be moderated by a number of factors, including aspects of the measurement procedure, system justification tendencies, political ideology, and stigma. The present research advances the literature by testing these hypotheses at the individual level—as SJT is primarily articulated—as well as at the level of the social group, thereby extending SJT with novel tests of generalizability and boundary conditions.

System Justification Theory

Decades of social psychology research have shown that people often think, feel, and act in self-interested ways (e.g., Miller, 1999) or in ways that serve the interests of the social groups they belong to (Greenwald & Pettigrew, 2014; Tajfel & Turner, 1979). Additionally, according to

---

[5] The SJT key publication (Jost et al., 2004) uses the terms *disadvantaged* and *low-status* interchangeably. Here, we use *disadvantaged* because it can be broadly applied to groups that are stigmatized, numerical minorities, of lower socio-economic status, and lacking access to resources.

SJT (Jost & Banaji, 1994; Jost et al., 2004), people are also generally motivated to justify and defend the social systems in which they live. However, many societies are characterized by high levels of inequality, and some social groups are more advantaged than other social groups in terms of status, power, and access to resources. Consequently, for members of advantaged groups, system justification motives (e.g., to support the status quo) are congruent with motivations to see themselves and the ingroup positively (e.g., self-esteem; group pride); but for members of disadvantaged groups, system justification motives are at odds with motivations to see themselves or the ingroup positively (Jost, 2019; Jost, Gaucher, & Stern, 2015; Jost & van der Toorn, 2012). SJT was developed, in part, as an advancement of Social Identity Theory (Tajfel & Turner, 1979), to better account for findings of outgroup favoritism in disadvantaged groups (Jost, Burgess, & Mosso, 2001). According to SJT (Jost & Banaji, 1994), system justifying beliefs often contain stereotypes that offer ostensible explanations why social groups inhabit their positions in society. Stereotypes not only describe the characteristics of social groups, but also justify the higher status of dominant groups (e.g., "they hold high positions by virtue") and the lower status of disadvantaged groups (e.g., "they are poorly off because they did not work hard enough").

In SJT's initial formulation, Jost and Banaji (1994) emphasized that stereotypes about social groups help to rationalize the social order and legitimize the positions that social groups hold in society. Although stereotypes of disadvantaged groups are often negative, members of disadvantaged groups are assumed to accept and internalize beliefs that are detrimental to one's own or the ingroup's interests but legitimize the status quo (i.e., false consciousness; Jost & Banaji, 1994). In the course of socialization, stereotypical beliefs become deeply entrenched and

highly accessible to both advantaged and disadvantaged group members (Jost et al., 2002; Mentovich & Jost, 2008).

SJT posits that even those who are clearly disadvantaged by their positions in society are motivated to justify the status quo because system justifying beliefs serve a palliative function and make people feel better about how things are (Jost & Hunyady, 2002). Acknowledging that the social system is unfair or unjust is emotionally taxing, especially for those who are disadvantaged by the system (Jost, Wakslak, & Tyler, 2008). Consequently, SJT assumes that disadvantaged group members internalize society's negative perceptions of their ingroup to regulate negative emotions about the unfairness of society.

Predictions Regarding Outgroup Favoritism

In articulating SJT as a framework for understanding outgroup favoritism in disadvantaged groups, Jost and colleagues (2004) proposed a number of hypotheses specifying the conditions under which members of disadvantaged groups should be especially likely to display outgroup favoritism. Of these hypotheses, we test two in the context of the present research. First, outgroup favoritism in members of disadvantaged groups is more likely observed on implicit than explicit measures. Second, outgroup favoritism is more likely when disadvantaged group members endorse system justifying beliefs. Additionally, a typical reading of SJT (Jost & Banaji, 1994) suggests a third, testable hypothesis: that disadvantaged group members' evaluations of their own group are more negative to the extent that their group is viewed negatively (i.e., stigmatized) by society (e.g., Dasgupta, 2004; Lane, Mitchell, & Banaji, 2005; Livingston, 2002; Rudman, Feinberg, & Fairchild, 2002). In the following section, we

elaborate on each of these hypotheses and review existing empirical evidence for each hypothesis.

Implicit versus explicit measures. Previous research has revealed that disadvantaged groups' intergroup evaluations are moderated by measurement method. Intergroup evaluations can be measured *explicitly*, often through direct self-report, as well as *implicitly*, based on the interpretation of speed or accuracy of responses rather than the contents of the response, *per se*.[6] Implicit measures were introduced, in part, because explicit measures have been shown to be susceptible to socially desirable responding (Gawronski & Hahn, 2019). Implicit measures are thought to circumvent self-presentation and social desirability through task procedures designed to minimize the extent to which people can deliberately feign responses (but see Czellar, 2006; Fiedler & Bluemke, 2005; Steffens, 2004). Nosek, Banaji, and Greenwald (2002) examined intergroup evaluation among members of disadvantaged groups who completed both implicit and explicit measures at the Project Implicit demonstration website. Black participants displayed a large effect of ingroup favoritism on the explicit measure, but a small effect of outgroup favoritism on the implicit measure. Older participants demonstrated an even more pronounced divergence by measurement method, with a small effect of outgroup favoritism on the explicit measure but a large effect of outgroup favoritism on the implicit measure.

---

[6] Throughout this paper, we use the term "implicit" in reference to indirect measurement tools ("implicit measures") and their behavioral outcomes ("implicit evaluations" and "implicit bias"). Thus, our use of the term "implicit" does not make assumptions about underlying mental representations or process characteristics (Corneille & Hütter, 2020).

In response to such demonstrations of implicit outgroup favoritism but explicit ingroup favoritism among disadvantaged groups, Jost and colleagues (2004) reasoned that outgroup favoritism in disadvantaged groups should be more likely to be observed on implicit than on explicit measures (Hypothesis 6'; Jost et al., 2004, p. 893). They argued that members of disadvantaged groups feel intense social pressures to show ingroup pride, so they should be reluctant to openly endorse beliefs that disfavor the ingroup. Because disadvantaged group members would not want to be seen as identifying with the dominant outgroup, they display ingroup favoritism under conditions that readily allow for such deliberate responding, i.e., explicit measures. However, disadvantaged group members' internalized negativity may be more readily expressed under conditions that constrain deliberate responding, i.e., implicit measures. Taken together, SJT (Jost et al., 2004) predicts that implicit measures rather than explicit measures should more readily reveal outgroup favoritism in disadvantaged groups because they minimize socially-desirable responding.

As evidence of implicit outgroup favoritism in disadvantaged groups, a substantial body of work in the SJT tradition (Jost et al., 2004, 2015; Jost & van der Toorn, 2012; Mentovich & Jost, 2008) relies upon research in which disadvantaged group members complete an Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) measuring their evaluative preferences for their own group relative to an advantaged outgroup (Ashburn-Nardo, Knowles, & Monteith, 2003; Jost & Hunyady, 2002; Livingston, 2002; Nosek et al., 2002; Rudman et al., 2002; Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002). In line with the predictions of SJT, some studies provide evidence of outgroup favoritism among disadvantaged groups. For example, Black participants displayed a preference for White relative to Black targets (i.e., outgroup favoritism) on an evaluative IAT (Ashburn-Nardo et al., 2003; see also Nosek et al.,

2002). Another study reported the opposite result: Black participants displayed a preference for Black relative to White targets (i.e., ingroup favoritism) on an evaluative IAT, albeit to a lower degree than on an explicit feeling thermometer measure (Livingston, 2002, Experiment 1b). Still other studies have found more nuanced results in other intergroup domains. Students from a prestigious university demonstrated implicit ingroup favoritism, but students from a less prestigious university demonstrated no evaluative preference, i.e., neither ingroup nor outgroup favoritism (Jost et al., 2002, Study 1). A similar pattern was observed among Hispanic participants, who did not display any intergroup preference on two evaluative Hispanic-White IATs (Uhlmann et al., 2002). Taken together, some studies provide evidence that disadvantaged groups display outgroup favoritism on implicit measures (Jost et al., 2004), but other studies report either no preference or ingroup favoritism among disadvantaged group members on implicit measures.

System justifying tendencies and ideological beliefs. The second prediction articulated by Jost and colleagues is that disadvantaged group members are more likely to display outgroup favoritism to the extent that they exhibit system justifying tendencies (Hypothesis 8; Jost et al., 2004, p. 901). System justifying tendencies have been operationalized in a number of ways. For example, Jost and Thompson (2000) developed a scale to assess *economic system justification*, which measures beliefs about (in)equality and the (un)fairness of the economic system. System

justifying tendencies have also been operationalized in terms of social dominance orientation (SDO; Pratto, Sidanius, Stallworth, & Malle, 1994).[7]

System justifying tendencies can also manifest as ideological beliefs (Jost, 2019). Specifically, SJT predicts that disadvantaged group members should be more likely to display outgroup favoritism to the extent that they hold conservative beliefs (Hypothesis 8'; Jost et al., 2004, p. 901). Jost and colleagues (2003) reasoned that conservatism is comprised of two potentially interrelated core aspects: the tendency to oppose change and maintain the status quo, and the preference for inequality. These two core aspects of conservatism—opposition to change and preference for inequality—are often correlated because, in unequal societies, opposing change usually implies sustaining inequality and keeping traditionally advantaged groups in power (Jost et al., 2003). As such, conservatism can be regarded as a system justifying belief because it provides an intellectual basis for rationalizing the current state of affairs (see Jost, 2019, for an overview of studies on the relationship between system justification and ideological beliefs). In short, because conservatism is regarded as a system justifying ideology, disadvantaged group members should display outgroup favoritism to the extent that they hold conservative beliefs.

Previous research has examined relations between conservatism and intergroup evaluations in dominant groups (e.g., Cunningham, Nezlek, & Banaji, 2004) but, to our

---

[7] See Jost and Hunyady (2005) for a discussion of constructs related to system justifying tendencies.

knowledge, only a handful of studies have directly investigated these relationships in disadvantaged groups. Hoffarth and Jost (2017) re-analyzed data from homosexual and bisexual participants and discovered that higher conservatism was related to less favorable evaluations of gay people relative to straight people on an IAT. Similarly, Jost and colleagues (2004) found that gay and lesbian participants displayed more outgroup favoritism on both IATs and self-report measures to the extent that they self-identified as conservative. However, a different pattern of results emerged for Black and older participants. Black participants demonstrated a positive relationship between conservatism and explicit outgroup favoritism, but no relationship between conservatism and implicit outgroup favoritism, and intergroup evaluations were unrelated to conservatism among older participants. These findings illustrate heterogeneous effects of conservatism on intergroup evaluation in disadvantaged groups. Consequently, open questions remain regarding whether conservatism is related to outgroup favoritism in some groups but not others, as well as whether these effects might differ for implicit versus explicit measures of intergroup evaluation.

Relative group status and stigma. An idea frequently attributed to SJT (Jost & Banaji, 1994) is that disadvantaged groups are more likely to display outgroup favoritism to the extent that their group is stigmatized by society (e.g., Dasgupta, 2004; Lane et al., 2005; Livingston, 2002; Rudman et al., 2002). This notion is also inherent in Jost and colleagues' (2004) proposal that those who "suffer the most from the system are also those who have the most to explain, justify, and rationalize" (p. 909). Based on these ideas, intergroup evaluation in disadvantaged groups should depend on a social group's relative status and/or level of stigma.

Evidence for a relationship between stigma and intergroup evaluation in disadvantaged groups is mixed. For example, in one comparative study disadvantaged groups with higher perceived group status (Jewish and Asian participants) displayed ingroup favoritism on implicit measures and those with lower perceived group status (overweight and poor participants) displayed outgroup favoritism on implicit measures, but both groups' intergroup evaluations on explicit measures were unrelated to group status (Rudman et al., 2002). Similarly, Black participants who believed that Black Americans were disliked by White Americans demonstrated less implicit ingroup favoritism, but more explicit ingroup favoritism, relative to Black participants who believed that Black Americans were liked by White Americans (Livingston, 2002). In other words, to the extent that Black people believed that their social group was disliked by White people, they demonstrated ingroup favoritism on an explicit measure but outgroup favoritism on an implicit measure. In sum, extant findings support the relationship between intergroup evaluations and group status/stigma predicted by SJT (Jost et al., 2004), but only on implicit measures.

Open Questions Regarding Intergroup Evaluation in Disadvantaged Groups

As the reviewed findings illustrate, extant empirical evidence supports some of the predictions derived from SJT (Jost et al., 2004) regarding outgroup favoritism in disadvantaged groups, but evidence for other predictions is mixed. Additionally, the accumulated evidence highlights two related gaps in the SJT literature. First, many of the studies reviewed thus far focused on one disadvantaged group at a time (e.g., Ashburn-Nardo et al., 2003; Hoffarth & Jost, 2017; Livingston, 2002; Uhlmann et al., 2002), so it remains an open question whether or to what extent these results are specific to the investigated social groups. This narrow focus provides a

streamlined experimental design to test specific predictions derived from SJT, but offers only limited insight into general processes among disadvantaged groups. Given that SJT (Jost et al., 2004) is articulated as a generalized theory, analyses that incorporate multiple social identities are better positioned to test the claims made by SJT, as well as examine their boundary conditions. The second, related gap in the SJT literature is that a number of studies have examined intergroup attitudes across multiple social groups, but generally focus on documenting main effects of ingroup versus outgroup favoritism (e.g., Axt, Ebersole, & Nosek, 2014; Devos & Banaji, 2005; Nosek et al., 2002; Rudman & Ashmore, 2007), to the exclusion of psychological processes directly related to SJT, such as system justification tendencies, ideological beliefs, or stigma. To our knowledge, only two studies examining intergroup evaluation in disadvantaged groups fulfill both of these criteria and measure psychological processes related to SJT across multiple social groups (Jost et al., 2004; Rudman et al., 2002). Thus, SJT as a generalized theory of intergroup processes among disadvantaged groups would be strengthened by more process-level evidence from more groups.

Taken together, SJT (Jost et al., 2004) has articulated or inspired a number of hypotheses regarding intergroup evaluation in disadvantaged groups. Namely, SJT predicts that outgroup favoritism is more likely: to manifest on implicit than on explicit measures; to the extent that disadvantaged groups exhibit system justifying tendencies or endorse conservative beliefs; and to the extent that a social group is stigmatized by society. However, extant evidence provides varying levels of support for these predictions. These gaps in the SJT literature have motivated the present research into the relationships among system justifying tendencies, conservative beliefs, and stigma in the context of the implicit and explicit intergroup evaluations of large samples of many disadvantaged groups.

The Present Research

The primary aim of our research is to directly test hypotheses derived from SJT (Jost et al., 2004) regarding intergroup evaluations in disadvantaged groups. To do so, the present research uses very large samples drawn from 14 countries reflecting eight distinct social identities, which is a broader and more diverse sample than has been examined in any previous SJT research. These data were collected by Project Implicit (https://implicit.harvard.edu/implicit/), a demonstration website where visitors can take different online versions of the IAT. Furthermore, Project Implicit has established a wide variety of international collaborations, setting up websites in many countries that additionally conduct country-specific studies with translated and adapted measures. One major advantage of these data is that the methodology is highly similar across countries, providing high levels of internal validity for between-country comparisons. Project Implicit data were made available by Xu and colleagues (Xu et al., 2017, 2018) at the Open Science Framework (OSF).

The present research focuses on intergroup evaluation in disadvantaged groups, including both implicit and explicit measures as dependent variables. As implicit measures, we use the IAT with participants' ingroup and outgroup as target categories. As explicit measures, we use two: one-item preference measures, which ask participants to judge how much they prefer the ingroup relative to the outgroup; and feeling thermometers, which ask participants to (separately) report their felt warmth or coldness towards the ingroup and the outgroup.

Using the Project Implicit data, we first test the prediction that disadvantaged group members "will be more likely to exhibit outgroup favoritism on implicit measures than on explicit measures" (Jost et al., 2004, p. 893) by examining the magnitude and direction of

disadvantaged groups' intergroup evaluation, comparing effect sizes on an implicit measure with two explicit measures. Second, we test the prediction that "[a]s political conservatism increases, members of low-status groups will exhibit increased outgroup favoritism" (Jost et al., 2004, p. 901) by examining whether implicit and explicit intergroup evaluation in disadvantaged groups is moderated by self-reported political ideology. Third, we test the assumption that those who "suffer the most from the system are also those who have the most to explain, justify, and rationalize" (Jost et al., 2004, p. 909) by examining whether stigma is related to outgroup favoritism among disadvantaged groups. One way to think about stigma is that a social group is stigmatized to the extent that the rest of society views that social group negatively (cf. Link & Phelan, 2001; Pinel, 1999). In our analyses, we thus conceptualize stigma in terms of the attitude measures taken from all Project Implicit visitors who are not members of the disadvantaged (i.e., stigmatized) group, and specifically operationalize stigma as non-disadvantaged group members' average evaluations of the disadvantaged group.

Importantly, the analyses reported here extend previous research in two key ways. First, we take a meta-analytic approach to test the hypotheses derived from SJT across a wide variety of intergroup domains. Previous SJT research has generally reported between-group comparisons of intergroup evaluations measured from one relatively advantaged group and one relatively disadvantaged group (e.g., Jost & Hunyady, 2002). Extending upon these relatively more focused comparisons, the present research compares intergroup evaluations measured from a wide variety of disadvantaged groups and, thus, assesses the generalizability of the predictions of SJT across different intergroup domains.

The second way in which the present research extends upon previous work is that we test SJT's predictions—whenever possible—at two levels of analysis: the individual and the social group. The predictions derived from SJT are formulated at the individual level, as applying to members of disadvantaged groups, so we test these predictions at the individual level: e.g., examining whether members of disadvantaged groups who endorse conservative beliefs also prefer the outgroup relative to the ingroup. We also test the predictions of SJT at the level of the social group by using group-level aggregates of measures of intergroup evaluation: e.g., examining whether disadvantaged groups who endorse more conservative beliefs also prefer the outgroup relative to the ingroup. In these analyses, the sample (i.e., social group) rather than the participant is treated as the unit of observation. This social-group level approach has three benefits. The first benefit is that it provides the opportunity to test the moderating influence of social group *per se* on intergroup evaluation. The second benefit to this approach is that, relative to analyses based on individual-level measures, group-level aggregates yield more precise estimates of intergroup evaluation (Rushton, Brainerd, & Pressley, 1983). The issue of measurement precision is especially relevant in the context of the IAT, which has been criticized for having low measurement reliability relative to explicit measures of the same construct (Gawronski, Morrison, Phills, & Galdi, 2017). The third benefit of using the social group as the unit of observation is that it can reveal qualitatively different psychological processes than can analyses that use the individual as the unit of observation.

Rushton et al. (1983) highlight the utility of aggregation from the perspective of amplifying signal (i.e., the construct of interest) by canceling out noise (i.e., measurement error). However, aggregation also cancels out the influence of other psychological constructs that might vary between individuals that are not measurement error, but also are not specific to group

identity. Consequently, individual-level analyses can be interpreted to reflect the influence of a variety of processes. In contrast, group-level analyses cancel out the influence of processes that vary among group members, thereby amplifying the influence of the common trait(s) shared among group members, i.e., the defining feature(s) of group membership. Taken together, individual-level analyses can be interpreted to reflect individual differences, and group-level analyses to reflect group processes. Because SJT is articulated at the level of the individual, these group-level analyses represent a novel extension of the theoretical perspective. By including both individual- and group-level analyses, the present research examines the extent to which the hypotheses proposed by SJT persist at both the individual level and the social group level.

To assume without empirical support that phenomena at one level of analysis persist at other levels is to commit the ecological fallacy (Selvin, 1958). In the classic demonstration of the ecological fallacy, English literacy rates were higher in regions of America with higher proportions of foreign-born (i.e., non-native English speaking) residents (Robinson, 1950). Follow-up analyses of this seemingly paradoxical finding revealed that foreign-born individuals were less likely than native Americans to be English-literate, but largely settled in regions where the population is more literate (e.g., where there are more employment and educational opportunities). From this perspective, it would be premature to assume that any of the predictions of SJT that are based on individual-level data necessarily persist at the social group level. Thus, the present research is both statistically and theoretically positioned to extend SJT and the intergroup relations literature more broadly.

Method

Study Selection

Figure 1 depicts an adapted PRISMA flow diagram (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009), visualizing the process of identifying datasets and assessing their eligibility. At the time of data analysis and compilation of this report on April 5, 2019, $N = 110$ datasets were publicly available at OSF. Datasets can be accessed via osf.io/kaqi5 and osf.io/y9hiq[8]

[8] The two OSF projects differ in that one hosts only datasets collected on the US-based website whereas the other project hosts datasets collected on country-specific, non-US sites.

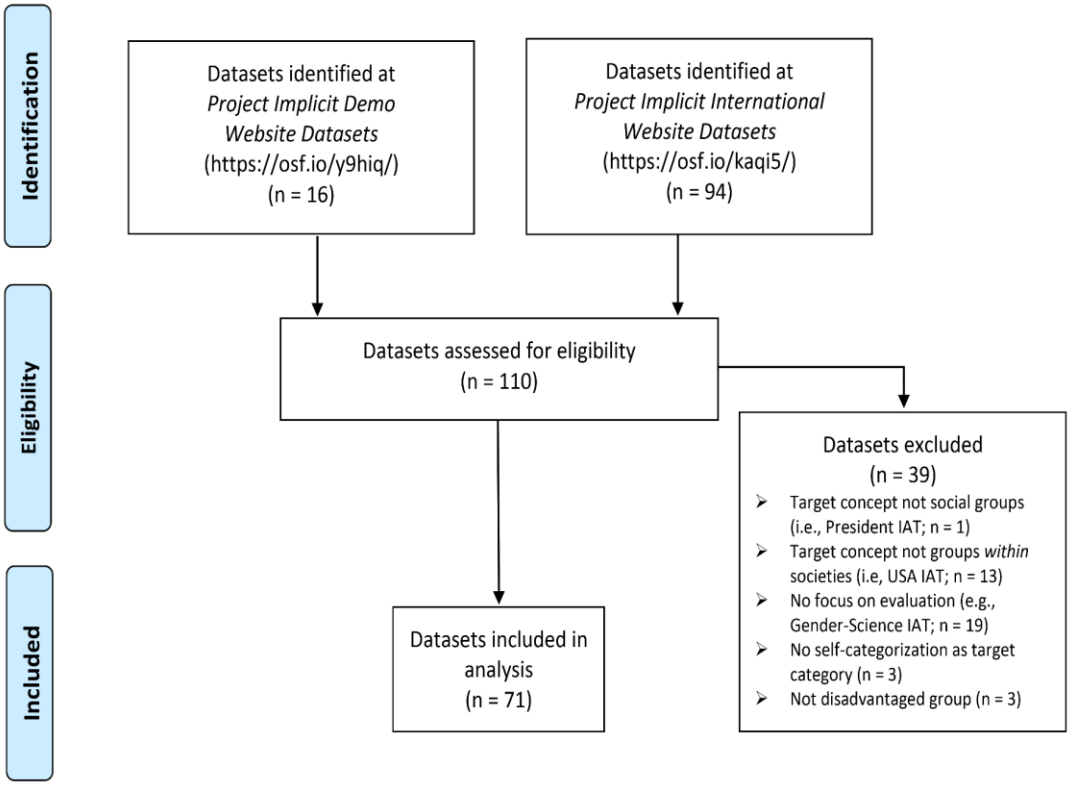

Figure 1. Adapted PRISMA flow diagram (Moher et al., 2009) showing the process of identifying, assessing for eligibility, and selecting datasets. In contrast to the standard PRISMA 2009 flow diagram, the screening phase is omitted here because all datasets were drawn from the same source: Project Implicit.

Inclusion and exclusion criteria. We included only studies with social groups or categories as target concepts. Studies measuring evaluations of individuals, such as evaluations of the President of the US, were outside the scope of the present study and were thus excluded (n = 1 dataset). Because we were interested in how disadvantaged group within societies evaluate their ingroup relative to an outgroup, we excluded studies focusing on nations as social categories (i.e., USA IATs; n = 13 datasets). Furthermore, the present research focused on intergroup *evaluations*, that is, relationships between positive versus negative attributes and the ingroup versus the outgroup. Consequently, we included only evaluative IATs and items and scales assessing liking and preference. Studies that did not focus on evaluations *per se* (e.g., semantic

attributes, or stereotypes, such as American vs. Foreign in the Asian IAT, or Science vs. Arts in the Gender-Science IATs; n = 19 datasets) were outside the scope of this study and were thus excluded. From each dataset we included only data from participants who self-identified as members of the disadvantaged social group that the study focused on (e.g., a person identifying as being older who took the Age IAT). Data of all other participants, who did not self-identify as members of the disadvantaged target category, were excluded from the primary analyses (but were used to compute average stigma scores, see below). Studies were also excluded if one of the target categories did not unambiguously refer to a disadvantaged group (n = 3).[9] Additionally, studies were excluded if participants were not directly asked whether they self-identified as members of either of the target categories (n = 3) .[10] The final number of datasets included in the present study was n = 71, yielding a total of k = 73 independent effect sizes.

---

[9] We excluded two studies, conducted in the US, using the target categories "Arab Muslims" versus "Other People" and "Judaism" versus "Other Religions", because Arab Muslim targets or Jewish targets were not unambiguously disadvantaged relative to *all* other outgroup targets. Furthermore, we excluded data from Jewish participants, who completed a Religion IAT with the target categories "Judaism" versus "Islam", because Judaism was not unambiguously disadvantaged relative to Islam in the US.

[10] We excluded one study that focused on evaluations of dark-skinned versus light-skinned people in Australia, because national/ethnic group membership but not self-reported skin tone was measured. Furthermore, we excluded a study that focused on evaluations of Black people

Datasets reviewed. Over the years, different studies have been added to the Project Implicit demonstration websites. The time period of data collection spans from 2002 to the present, with some studies running for the full time period (e.g., Race IAT in the US) and others starting later (e.g., the Race IAT in Spain) or others discontinued at some point in time (e.g., the Religion IAT in the US). For each country-specific study, we used the dataset that included data for all years of data collection. The present research includes eight intergroup domains with distinct IAT versions and datasets from 14 countries[11]; see Table 1 for an overview of included samples and dependent measures and Table 2 for descriptive statistics of continuous moderators. The total sample size was N = 715,721.

versus White people in The Netherlands, because it assessed ethnic group memberships (e.g., Surinamese) but not self-categorization as Black. Lastly, we excluded one dataset, which used multi-category versions of the IAT to measure evaluations of religious groups, but did not measure self-reported religious group membership.

[11] For data from the US Project Implicit website, we included only participants who indicated US citizenship or who indicated that their current location was in the United States.

Table 1

Overview of Project Implicit samples, dependent variables, and descriptive statistics.

			IAT		Preference		Thermometer	
IAT Version	Group	Country	dz	N	dz	N	dz	N
Age	Older (55+) age	AUS	1.08	1033	0.06	1158	-0.05	1236
Age	Older (55+) age	BEL	1.52	165	0.46	160	0.27	185
Age	Older (55+) age	BRA	1.12	138	0.31	146	0.22	162
Age	Older (55+) age	CAN (EN)	1.07	1387	0.07	1474	-0.06	1536
Age	Older (55+) age	CAN (FR)	1.50	97	0.35	101	0.05	105
Age	Older (55+) age	CHN	1.68	18	0.50	21	0.30	21
Age	Older (55+) age	DEU	1.57	687	0.40	661	0.31	727
Age	Older (55+) age	ESP	1.42	658	0.49	651	-0.02	684
Age	Older (55+) age	FRA	1.67	836	0.44	836	0.25	905
Age	Older (55+) age	GBR	1.17	2485	0.14	2737	0.00	2861

Age	Older (55+) age	KOR	1.25	24	0.67	26	-0.04	28
Age	Older (55+) age	NLD	1.30	652	0.43	704	0.26	767
Age	Older (55+) age	SWE	1.32	773	0.22	663	-0.02	798
Age	Older (55+) age	USA	1.06	35208	0.04	35960	-0.14	37776
Arab	Arab and/or Muslim	FRA	0.33	1455	-0.24	1465	-0.30	125
Disability	Participants with disabilities	USA	0.89	48086	0.23	50589	-0.11	51832
Race	Black or African (American)	AUS	-0.01	325	-0.61	382	-0.55	382
Race	Black or African (American)	BEL	0.08	86	-0.19	96	-0.50	102
Race	Black or African (American)	BRA	0.16	480	-0.22	677	-0.40	691
Race	Black or African (American)	CAN (EN)	0.12	2293	-0.53	2445	-0.56	2532
Race	Black or African (American)	CAN (FR)	0.16	152	-0.46	143	-0.51	158
Race	Black or African (American)	CHN	-0.14	3	1.17	4	-1.79	6
Race	Black or African (American)	DEU	0.34	194	-0.04	202	-0.11	225
Race	Black or African (American)	ESP	0.56	88	0.17	102	-0.21	110

Race	Black or African (American)	FRA	0.23	1022	-0.35	1078	-0.40	1132
Race	Black or African (American)	GBR	0.03	4197	-0.58	4864	-0.55	4988
Race	Black or African (American)	SWE	-0.08	207	-0.33	217	-0.61	229
Race	Black or African (American)	USA	-0.09	272872	-0.76	265536	-0.74	282975
Religion	Jewish	USA	-0.90	780	-2.06	761	-1.46	763
Religion	Muslim	USA	-0.63	405	-1.77	390	-0.97	392
Religion	Muslim	USA	-0.99	426	-1.85	412	-1.06	416
Sexuality	Gay/Lesbian	AUS	-0.31	1880	-0.47	2366	-0.34	2457
Sexuality	Gay/Lesbian	BEL	-0.31	367	-0.71	375	-0.42	435
Sexuality	Gay/Lesbian	BRA	-0.61	1264	-0.70	1732	-0.73	1790
Sexuality	Gay/Lesbian	CAN (EN)	-0.31	2101	-0.47	2328	-0.42	2448
Sexuality	Gay/Lesbian	CAN (FR)	-0.39	305	-0.25	306	-0.24	339
Sexuality	Gay/Lesbian	CHN	-0.53	1855	-0.61	2352	-0.30	2454
Sexuality	Gay/Lesbian	DEU	-0.40	2896	-0.54	2954	-0.40	3276

Sexuality	Gay/Lesbian	ESP	-0.38	1204	-0.52	1371	-0.37	1491
Sexuality	Gay/Lesbian	FRA	-0.49	3068	-0.47	3252	-0.42	3592
Sexuality	Gay/Lesbian	GBR	-0.26	5904	-0.46	6963	-0.32	7198
Sexuality	Gay/Lesbian	KOR	-0.31	1375	-1.01	1818	-0.73	1909
Sexuality	Gay/Lesbian	NLD	-0.27	1020	-0.66	1183	-0.46	1272
Sexuality	Gay/Lesbian	RUS	-0.39	300	-0.62	294	-0.58	353
Sexuality	Gay/Lesbian	SWE	-0.52	962	-0.55	987	-0.56	1143
Sexuality	Gay/Lesbian	USA	-0.30	99669	-0.62	104699	-0.53	108612
Skin Tone	Dark-skinned	BEL	0.25	237	0.20	278	-0.09	319
Skin Tone	Dark-skinned	BRA	0.31	1055	0.27	1577	0.04	1647
Skin Tone	Dark-skinned	CAN (EN)	0.38	1332	0.01	1462	-0.17	1544
Skin Tone	Dark-skinned	CAN (FR)	0.38	91	-0.04	92	-0.19	102
Skin Tone	Dark-skinned	CHN	0.98	1105	0.81	1343	0.01	1414
Skin Tone	Dark-skinned	DEU	0.55	957	0.24	1034	-0.12	1127

Skin Tone	Dark-skinned	ESP	0.62	915	0.33	1103	-0.05	1160
Skin Tone	Dark-skinned	FRA	0.49	1180	-0.01	1310	-0.06	1390
Skin Tone	Dark-skinned	GBR	0.34	1747	-0.03	1974	-0.23	2111
Skin Tone	Dark-skinned	KOR	0.82	1212	0.79	1559	-0.46	1615
Skin Tone	Dark-skinned	NLD	0.18	861	0.08	1055	-0.21	1138
Skin Tone	Dark-skinned	SWE	0.20	834	0.05	919	-0.19	996
Skin Tone	Dark-skinned	USA	0.27	58479	-0.21	59063	-0.31	61888
Weight	Overweight	AUS	0.93	4128	0.72	5280	0.33	5401
Weight	Overweight	BEL	0.83	657	0.70	741	0.26	783
Weight	Overweight	BRA	0.59	1605	0.80	2144	0.55	2164
Weight	Overweight	CAN (EN)	0.89	6233	0.72	6979	0.29	7129
Weight	Overweight	CAN (FR)	1.22	605	1.10	651	0.71	670
Weight	Overweight	CHN	0.12	2774	0.79	3450	-0.42	3541
Weight	Overweight	DEU	1.16	5334	0.65	5666	0.35	5964

Weight	Overweight	ESP	0.96	2778	0.90	3213	0.22	3304
Weight	Overweight	FRA	1.14	4031	0.99	4391	0.59	4568
Weight	Overweight	GBR	0.83	9540	0.65	10859	0.30	11116
Weight	Overweight	KOR	0.60	2363	1.37	3168	-0.22	3264
Weight	Overweight	NLD	0.75	1402	0.84	1727	0.41	1815
Weight	Overweight	SWE	1.19	4692	0.89	5123	0.51	5409
Weight	Overweight	USA	0.85	98172	0.66	100396	0.24	103237

Note. Top row of column labels refer to dependent measures: IAT = IAT; Preference = one-item preference measure; Thermometer = feeling thermometer difference score. Second row of column labels refer to IAT Version, Group = disadvantaged group examined in primary analyses, dz = Cohen's d_z, with positive scores indicating outgroup favoritism and negative scores indicating ingroup favoritism from the perspective of the disadvantaged group; N = sample size. Abbreviations in third column denote country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

Individual-level Measures

Intergroup evaluation. As implicit measures, we used the evaluative variant of the IAT for each sample. Attribute categories of IATs were always "Good" versus "Bad"; target categories differed according to the specific intergroup domain. We included two different explicit measures. One was a one-item preference measure, which asked participants to judge how much they preferred the ingroup relative to the outgroup, with responses made using either a 5-point scale (in the earlier years of Project Implicit) or a 7-point scale (in more recent years). For the other explicit measure, we used feeling thermometers, which asked participants to respond on scales from 0 (*extremely cold*) to 10 (*extremely warm*) how they felt towards the ingroup and the outgroup.

Conservatism. Conservatism was assessed using one-item self-placement measures. Using 6- or 7-point scales, participants were asked to place themselves along a continuum ranging from liberal to conservative.[12]

Group-level Measures

Intergroup evaluation.

[12] Measures were adapted for country-specific websites with some studies using other but similar labels (e.g., left-wing vs. right-wing, conservative vs. progressive).

We calculated the sample averages of the IAT *D* Score, the one-item preference score, and the feeling thermometer difference score, separately for each sample of disadvantaged groups.

Conservatism. We calculated the sample average of the one-item conservatism measure separately for each sample of disadvantaged groups.

Stigma. We calculated average evaluation scores of each disadvantaged group based on the responses of Project Implicit visitors in each study who self-reported being members of social groups other than the disadvantaged group in the same study. We used these evaluation scores as proxies for the extent to which each disadvantaged group was stigmatized by the rest of society. For each disadvantaged group, we thus calculated three measures of stigma, based on non-disadvantaged group members' IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores.

Table 2

Descriptive statistics for continuous moderators.

				Stigma		
IAT Version	Group	Country	Conservatism	IAT	Preference	Thermometer
Age	Older (55+) age	AUS	-0.25	1.14	0.47	0.17
Age	Older (55+) age	BEL	-0.15	1.39	0.60	0.38
Age	Older (55+) age	BRA	-0.08	0.99	0.51	0.24
Age	Older (55+) age	CAN (EN)	-0.43	1.04	0.41	0.11
Age	Older (55+) age	CAN (FR)	-0.66	1.24	0.53	0.29
Age	Older (55+) age	CHN	-0.82	1.20	0.70	-0.05
Age	Older (55+) age	DEU	-0.62	1.38	0.57	0.28
Age	Older (55+) age	ESP	-0.56	1.20	0.52	0.07
Age	Older (55+) age	FRA	-0.39	1.39	0.61	0.31
Age	Older (55+) age	GBR	-0.05	1.14	0.38	0.12

Age	Older (55+) age	KOR	0.52	1.30	1.15	-0.27
Age	Older (55+) age	NLD	-0.88	1.28	0.61	0.26
Age	Older (55+) age	SWE	-0.01	1.19	0.46	0.08
Age	Older (55+) age	USA	-0.33	1.09	0.42	0.08
Arab	Arab and/or Muslim	FRA	-0.66	1.08	-0.24	0.53
Disability	Participants with disabilities	USA	-0.37	1.16	0.54	0.18
Race	Black or African (American)	AUS	-0.44	0.86	0.49	0.25
Race	Black or African (American)	BEL	-0.78	0.82	0.60	0.41
Race	Black or African (American)	BRA	-0.80	0.86	0.45	0.23
Race	Black or African (American)	CAN (EN)	-0.42	0.89	0.42	0.26
Race	Black or African (American)	CAN (FR)	-0.32	1.08	0.57	0.40
Race	Black or African (American)	CHN	0.88	1.07	0.92	0.27
Race	Black or African (American)	DEU	-0.27	0.88	0.50	0.27
Race	Black or African (American)	ESP	-0.44	1.02	0.66	0.35

Race	Black or African (American)	FRA	-0.43	0.95	0.46	0.29
Race	Black or African (American)	GBR	-0.40	0.87	0.53	0.36
Race	Black or African (American)	SWE	-0.59	0.71	0.54	0.31
Race	Black or African (American)	USA	-0.34	0.84	0.43	0.29
Religion	Jewish	USA	-0.97	0.80	0.57	0.19
Religion	Muslim	USA	-0.53	0.99	0.82	0.42
Religion	Muslim	USA	-0.64	0.50	0.63	0.50
Sexuality	Gay/Lesbian	AUS	-0.83	0.62	0.50	0.46
Sexuality	Gay/Lesbian	BEL	-0.38	0.66	0.72	0.69
Sexuality	Gay/Lesbian	BRA	-0.72	0.47	0.66	0.61
Sexuality	Gay/Lesbian	CAN (EN)	-1.01	0.58	0.48	0.45
Sexuality	Gay/Lesbian	CAN (FR)	-1.00	0.40	0.52	0.54
Sexuality	Gay/Lesbian	CHN	-0.72	0.27	0.88	0.70
Sexuality	Gay/Lesbian	DEU	-0.81	0.46	0.65	0.64

Sexuality	Gay/Lesbian	ESP	-1.11	0.59	0.61	0.51
Sexuality	Gay/Lesbian	FRA	-0.68	0.40	0.54	0.51
Sexuality	Gay/Lesbian	GBR	-0.84	0.65	0.53	0.51
Sexuality	Gay/Lesbian	KOR	-0.55	0.55	0.64	0.25
Sexuality	Gay/Lesbian	NLD	-1.19	0.76	0.65	0.67
Sexuality	Gay/Lesbian	RUS	-0.25	0.85	1.03	0.79
Sexuality	Gay/Lesbian	SWE	-0.40	0.36	0.49	0.48
Sexuality	Gay/Lesbian	USA	-1.21	0.62	0.48	0.47
Skin Tone	Dark-skinned	BEL	-0.47	0.76	0.54	0.31
Skin Tone	Dark-skinned	BRA	-0.52	0.88	0.58	0.38
Skin Tone	Dark-skinned	CAN (EN)	-0.47	0.85	0.43	0.26
Skin Tone	Dark-skinned	CAN (FR)	-0.51	1.09	0.46	0.37
Skin Tone	Dark-skinned	CHN	-0.40	1.12	0.91	0.16
Skin Tone	Dark-skinned	DEU	-0.59	0.89	0.51	0.24

Skin Tone	Dark-skinned	ESP	-0.53	1.12	0.91	0.16
Skin Tone	Dark-skinned	FRA	-0.54	0.96	0.38	0.28
Skin Tone	Dark-skinned	GBR	-0.53	0.85	0.42	0.26
Skin Tone	Dark-skinned	KOR	-0.40	0.96	0.90	-0.39
Skin Tone	Dark-skinned	NLD	-0.67	0.68	0.57	0.30
Skin Tone	Dark-skinned	SWE	-0.33	0.67	0.54	0.35
Skin Tone	Dark-skinned	USA	-0.45	0.75	0.31	0.22
Weight	Overweight	AUS	-0.34	0.96	1.07	0.58
Weight	Overweight	BEL	-0.24	1.03	1.12	0.56
Weight	Overweight	BRA	-0.26	0.84	1.08	0.77
Weight	Overweight	CAN (EN)	-0.46	0.90	0.99	0.56
Weight	Overweight	CAN (FR)	-0.54	1.40	1.26	0.88
Weight	Overweight	CHN	-0.43	0.21	0.77	-0.01
Weight	Overweight	DEU	-0.56	1.28	0.98	0.65

Weight	Overweight	ESP	-0.59	1.05	1.05	0.38
Weight	Overweight	FRA	-0.42	1.21	1.18	0.75
Weight	Overweight	GBR	-0.42	0.83	0.99	0.56
Weight	Overweight	KOR	-0.35	0.75	1.48	-0.04
Weight	Overweight	NLD	-0.66	0.96	1.24	0.58
Weight	Overweight	SWE	-0.09	1.24	1.09	0.67
Weight	Overweight	USA	-0.29	1.09	1.08	0.48

Note. Group = disadvantaged group examined in primary analyses. Conservatism = Cohen's d_z for aggregated mean of disadvantaged group members' self-reported conservatism, with positive scores indicating conservative self-placement and negative scores indicating liberal self-placement. Stigma = Cohen's d_z for the three stigma measures: IAT = IAT *D* Score stigma; Preference = one-item preference score stigma; Thermometer = feeling thermometer difference score stigma. Stigma measures reflect the aggregated mean of non-disadvantaged group members' responses, with positive scores indicating negative evaluations of the disadvantaged group and negative scores indicating positive evaluations of the disadvantaged group. Abbreviations in third column denote country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

Intergroup Domains

Old vs. Young. There were *k* = 14 studies focusing on age-related group evaluations. We included in the primary analyses only participants with a self-reported age of 55 years and older. The threshold of 55 for "older adults" has been used in previous research (e.g., Kite, Stockdale, Whitley, & Johnson, 2005; Neugarten, 1974). Target categories in the Age IAT, one-item preference measures, and feeling thermometers were "Old People" versus "Young People".

Arab vs. French. There was *k* = 1 study focusing on evaluations of Arab people relative to French people. We included in the primary analyses only participants who self-categorized as Arab or Muslim. Target categories in the French Arab IAT were "Maghreb People" versus "French People."[13]

Disabled vs. Abled. There was *k* = 1 study focusing on evaluations of disability relative to non-disability. We included in the primary analyses only participants who indicated that they had a disability. Target categories in the Disability IAT, one-item preference measures, and feeling thermometers were "Disabled People" versus "Abled People".

Black vs. White. There were *k* = 12 studies focusing on evaluations of Black people relative to White people. We included in the primary analyses only participants who self-

[13] Maghreb refers to a region in North and Northwestern Africa. The term is frequently used in French referring to North-African countries with Arabic as an official language, such as Morocco, Algeria, and Tunisia (Oxford Dictionaries, 2019).

categorized as Black or African, or joint identities, such as African American, Black British, or Black Caribbean. Target categories in the Race IAT, one-item preference measures, and feeling thermometers were “Black People” versus “White People”, “Black” versus “White”, or “African American” versus “European American”, depending on the language- or country-specific demonstration website.

Religious groups. There were $k = 3$ studies, all conducted in the USA, focusing on evaluations of religious groups relative to other religious groups. The studies used three variants of the Religion IAT and respective one-item preference measures and feeling thermometers. We included the sample of Jewish participants who completed the “Judaism” versus “Christianity” measures, and the two samples of Muslim participants who completed either the “Judaism” versus “Islam” or the “Islam” versus “Christianity” measures, respectively.

Gay vs. Straight. There were $k = 15$ studies focusing on evaluations of gay people relative to straight people. We included in the primary analyses only participants who self-categorized as homosexual. Target categories in the Sexuality IAT, one-item preference measures, and feeling thermometers were “Gay People” versus “Straight People” or “Homosexual” versus “Heterosexual”, depending on the language- or country-specific demonstration website.

Dark-Skinned vs. Light-Skinned. There were $k = 13$ studies focusing on evaluations of dark-skinned people relative to light-skinned people. We included in the primary analyses only participants who self-categorized as somewhat dark-skinned, dark-skinned, or very dark-skinned. Target categories in the Skin Tone IAT, one-item preference measures, and feeling thermometers were “Dark Skinned People” versus “Light Skinned People”.

Overweight vs. Normal Weight. There were *k* = 14 studies focusing on evaluations of overweight people relative to normal weight people. We included in the primary analyses only participants who self-categorized as being overweight. Target categories in the Weight IAT, one-item preference measures, and feeling thermometers were "Fat People" versus "Thin People" or "Fat" versus "Thin", depending on the language- or country-specific demonstration website.

Data analysis

We calculated average IAT *D* Scores (Greenwald, Nosek, & Banaji, 2003), one-item preference scores, and feeling thermometer difference scores for each sample within each dataset. Effect size estimates for all measures were coded such that positive scores indicated a preference for advantaged groups relative to disadvantaged groups and negative scores indicated a preference for disadvantaged groups relative to advantaged groups. We calculated feeling thermometer difference scores by subtracting ingroup feeling thermometers from outgroup feeling thermometers. We calculated effect size estimates Cohen's d_z for IAT effects by dividing each sample IAT *D* Score by its standard deviation. For one-item preference measures and feeling thermometer difference scores, we took two steps to calculate effect size estimates. First, we performed one-sample *t*-tests, testing one-item preference scores against the scale midpoint and feeling thermometer difference scores against zero. We then calculated for each (sub-)sample Cohen's d_z, using the following formula (see Lakens, 2013; Rosenthal, 1991):

$$d_z = \frac{t}{\sqrt{n}}$$

where *t* is the test statistic obtained from one-item preference scores (versus the scale midpoint) or feeling thermometer difference scores (versus zero) and *n* is the respective sample size.

One-item preference measures were assessed using 5-point scales in earlier years and using 7-point scales in more recent years. Each sample could thus contribute up to two one-item preference scores. For samples that produced two effect sizes, we calculated an aggregated mean effect size, weighting effect sizes by their respective sample sizes. All dependent variables—IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores—were coded such that positive scores indicated outgroup favoritism and negative scores indicated ingroup favoritism from the perspective of members of the disadvantaged groups.[14]

We further calculated effect size estimates for conservatism using the same procedures detailed above. For samples that produced two effect sizes (i.e., corresponding to 6- and 7-point

---

[14] Note that the number of datasets does not equal the number of effect sizes within the present study. Instead, effect size estimates were calculated at the sample level and some datasets contributed multiple independent effect sizes from multiple independent samples. These were cases in which a dataset was comprised of studies that assessed evaluations toward different target categories. For example, the Religion IAT dataset included different independent studies, assessing evaluations towards Christianity vs. Judaism, Christianity vs. Islam, and Judaism vs. Islam. This dataset contributed three independent samples, because Jewish and Muslim participants participated in all three studies.

scales), we calculated an aggregated mean effect size, weighting effect sizes by their respective sample sizes. Effect size estimates for conservatism were coded such that positive scores indicated more conservative self-placement and negative scores indicated more liberal self-placement. Finally, we used the same approach to calculate effect size estimates for the three measures of stigma: IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores, using the data of all participants of each study who did not self-categorize as belonging to the disadvantaged target group. Effect size estimates for stigma were coded such that positive scores indicated a preference for advantaged groups relative to disadvantaged groups and negative scores indicated a preference for disadvantaged groups relative to advantaged groups.

Analyses were conducted using a meta-analytic framework. We employed a random-effects model to allow for the assumption that different studies have different underlying true effects without assuming that there is only one true effect underlying the observed study results (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2010; Cheung, 2015). As such, the studies included in our meta-analyses are assumed to be a random sample from a population of studies which, in principle, allows for the meta-analytic results to be generalized beyond the included studies (Cheung, 2015). For fitting a random-effects model, we weighted effects by their inverse variance to estimate an average population effect size. We calculated the inverse variance w of d_z for IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores following the formula provided by Lipsey (2001, p. 72):

$$w = \frac{1}{SE^2}$$

where

$$SE^2 = \sqrt{\frac{n+n}{n*n} + \frac{d_z}{2(n+n)}}$$

.

First, we estimated the heterogeneity of effects within each measure type, and then conducted follow-up moderator analyses with a series of mixed-effects meta-regressions. In these meta-regressions, political ideology and stigma were included as continuous moderators, and social group was dummy-coded as a categorical moderator. All analyses were done using R.[15] Analyses scripts are accessible at https://osf.io/cxp9z/.[16]

---

[15] R (Version 3.6.2; R Core Team, 2017) and the R-packages *apaTables* (Version 2.0.5; Stanley, 2018), *bookdown* (Version 0.17; Xie, 2016), *cowplot* (Version 1.0.0; Wilke, 2017), *data.table* (Version 1.12.8; Dowle & Srinivasan, 2017), *here* (Version 0.1; Müller, 2017), *knitr* (Version 1.28; Xie, 2015), *metafor* (Version 2.1.0; Viechtbauer, 2010), *papaja* (Version 0.1.0.9,942; Aust & Barth, 2018), *png* (Version 0.1.7; Urbanek, 2013), *tidyverse* (Version 1.3.0; Wickham, 2017), and *xtable* (Version 1.8.4; Dahl, 2016)

[16] Additional analyses can be found in Appendix A, including: individual-level correlational analyses using other measures of system justifying beliefs among a sub-sample of studies; country-level analyses using cultural value dimensions; parallel analyses assessing the

Results

Testing the Predictions of SJT at the Individual Level

The predictions derived from SJT (Jost et al., 2004) are formulated at the individual level, as applying to members of disadvantaged groups. Consequently, we first report a set of analyses based on individual-level data.

Implicit versus explicit measures. We fitted three separate random-effects models, using the three measures of intergroup evaluation as dependent variables.

IAT D Scores. We observed a significant mean effect of $d_z = 0.43$, $z = 5.34$, $p < .001$, 95% *CI* [0.27; 0.59] on IAT *D* Scores. This medium-sized positive effect indicates that, on average, members of disadvantaged groups displayed outgroup favoritism on the IAT. The estimated amount of total heterogeneity was $\tau^2 = 0.46$, $Q(72) = 80{,}468.60$, $p < .001$, accounting for a large proportion of the total variability. The percentage of the heterogeneity not attributable to sampling error was $I^2 = 99.94\%$, indicating that a high percentage of the estimated heterogeneity was due to genuine between-sample variability (Higgins, Thompson, Deeks, & Altman, 2003).

One-item preference scores. We observed a mean effect of $d_z = 0.02$, $z = 0.22$, $p = .827$, 95% *CI* [-0.14; 0.18] on one-item preference scores. This null effect indicates that, on average,

---

relationship between ideology and intergroup evaluations in advantaged groups and group members; and additional figures.

members of disadvantaged groups displayed neither ingroup nor outgroup favoritism on the one-item preference measures. The estimated amount of total heterogeneity was τ^2 = 0.47, $Q(72)$ = 186,325.30, $p < .001$, accounting for a large proportion of the total variability, and I^2 = 99.95%, indicating high total heterogeneity due to genuine between-sample variability.

Feeling thermometer difference scores. We observed a mean effect of d_z = -0.20, z = -3.83, $p < .001$, 95% *CI* [-0.30; -0.10] on feeling thermometer difference scores. This small negative effect indicates that, on average, members of disadvantaged groups displayed ingroup favoritism on feeling thermometer difference scores. The estimated amount of total heterogeneity was τ^2 = 0.18, $Q(72)$ = 45,597.29, $p < .001$, accounting for a substantial proportion of the total variability, and I^2 = 99.83%, indicating high total heterogeneity due to genuine between-sample variability.

Conservatism. To test the relationship between disadvantaged group members' conservative beliefs and intergroup evaluation at the individual level, we calculated the correlation between conservatism and the three measures of intergroup evaluation within each sample and fitted three separate random-effects models, weighting each correlation coefficient by its corresponding sample size.

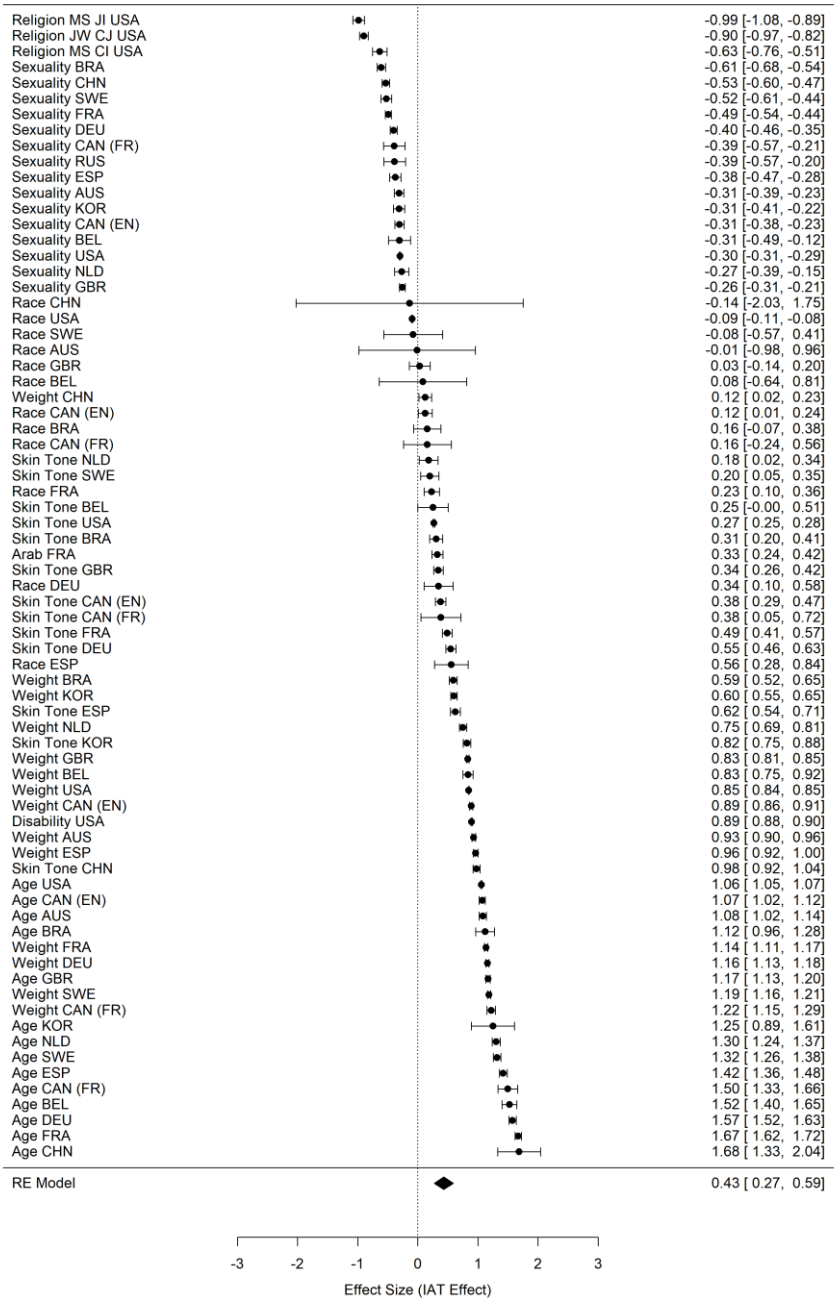

Figure 1. Caterpillar plot of random-effects meta-analysis of IAT effects (IAT *D* Scores) with study effects ordered by effect size. Positive scores indicate outgroup favoritism and negative scores indicate ingroup favoritism from the perspective of the disadvantaged groups. Error bars depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Dataset labels denote the intergroup domain for each study and the respective country. Abbreviations for religious groups: MS = Muslim participants, JW = Jewish participants, JI = Judaism vs. Islam, CJ = Christianity vs. Judaism, CI = Christianity vs. Islam. Country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

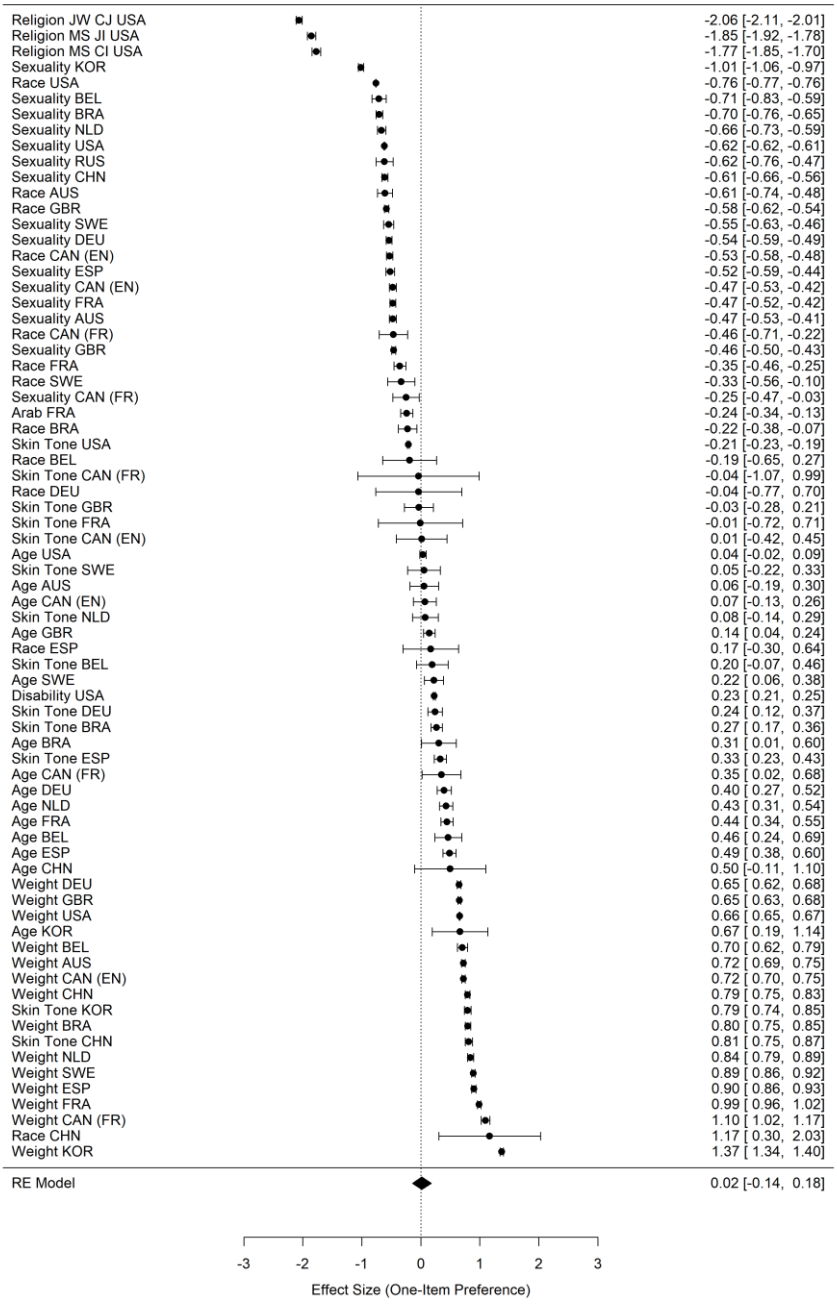

Figure 2. Caterpillar plot of random-effects meta-analysis of one-item preference scores with study effects ordered by effect size. Positive scores indicate outgroup favoritism and negative scores indicate ingroup favoritism from the perspective of the disadvantaged groups. Error bars depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Dataset labels denote the intergroup domain for each study and the respective country. Abbreviations for religious groups: MS = Muslim participants, JW = Jewish participants, JI = Judaism vs. Islam, CJ = Christianity vs. Judaism, CI = Christianity vs. Islam. Country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

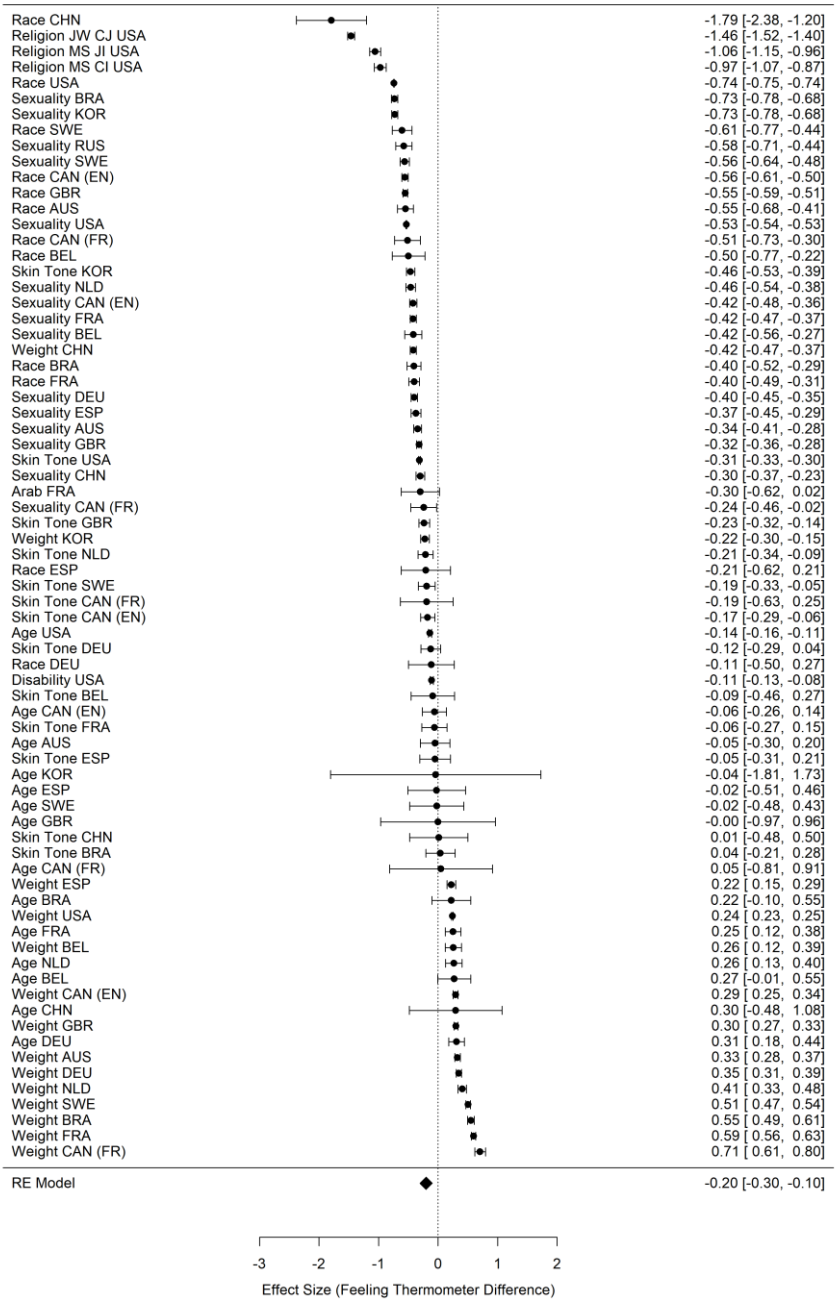

Figure 3. Caterpillar plot of random-effects meta-analysis of feeling thermometer (difference) scores with study effects ordered by effect size. Positive scores indicate outgroup favoritism and negative scores indicate ingroup favoritism from the perspective of the disadvantaged groups. Error bars depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Dataset labels denote the intergroup domain for each study and the respective country. Abbreviations for religious groups: MS = Muslim participants, JW = Jewish participants, JI = Judaism vs. Islam, CJ = Christianity vs. Judaism, CI = Christianity vs. Islam. Country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

IAT D Scores. We observed an average effect of r = .08, z = 10.80, p < .001, 95% *CI* [0.06; 0.09], indicating that the correlation between conservatism and IAT *D* Scores at the individual level was very small. The estimated amount of total heterogeneity was τ^2 = 0.0023, $Q(71)$ = 2,532.83, p < .001, accounting for a large proportion of the total variability, I^2 = 94.73%.

One-item preference scores. We observed an average effect of r = .07, z = 6.79, p < .001, 95% *CI* [0.05; 0.09], indicating that the correlation between conservatism and one-item preference scores at the individual level was very small. The estimated amount of total heterogeneity was τ^2 = 0.01, $Q(71)$ = 2,486.61, p < .001, accounting for a large proportion of the total variability, I^2 = 97.80%.

Feeling thermometer difference scores. We observed an average effect of r = .08, z = 7.12, p < .001, 95% *CI* [0.06; 0.10], indicating that the correlation between conservatism and feeling thermometer difference scores at the individual-level was very small. The estimated amount of total heterogeneity was τ^2 = 0.01, $Q(71)$ = 2,915.18, p < .001, accounting for a large proportion of the total variability, I^2 = 98.37%.

Taken together, individual-level analyses indicate that members of disadvantaged groups displayed outgroup favoritism on implicit measures, but no intergroup preference or ingroup favoritism on explicit measures. Furthermore, individuals who self-report being more conservative also displayed more favorable evaluations of the advantaged group relative to the disadvantaged group on all three measures. Additionally, and importantly, these average effects of intergroup evaluation were characterized by high levels of heterogeneity. Therefore, we conducted follow-up group-level analyses to examine whether this heterogeneity could be explained by a number of theoretically-derived moderators.

Testing and Extending the Predictions of SJT at the Social Group Level

The present research compares intergroup evaluations measured from a wide variety of disadvantaged groups. SJT does not make specific predictions about how relative levels of intergroup evaluation should vary across disadvantaged social groups *per se*. However, visual inspection of the caterpillar plots in Figure 2, 3, and 4 suggests that effects are clustered by social groups. For example, in all three figures, studies examining evaluations of religious groups and sexuality are clustered in the upper end of the distribution, indicating ingroup favoritism, whereas studies examining age- and weight-related evaluations are largely clustered in the bottom of the distribution, indicating outgroup favoritism. These data provide the opportunity to examine social group as a moderator in exploratory analyses, in order to assess the generalizability of the predictions of SJT across different intergroup domains. Additionally, we examined whether the relationships between conservatism and outgroup favoritism observed at the individual level also persists at the social group level. More specifically, we examined whether different disadvantaged groups display different levels of conservatism (i.e., are positioned differently on a conservatism-liberalism dimension) and whether group levels of conservatism are related to their average preferences for the ingroup relative to the outgroup. Finally, we examined the relationship between stigma and intergroup evaluation among disadvantaged groups. Whereas most previous research in this domain has relied on individuals' self-reported perceptions of stigma against their ingroup (e.g., Pinel, 1999), in the present research we operationalized stigma in terms of the intergroup biases of everybody else in the sample, which aligns more closely with classic conceptualizations of stigma (e.g., Link & Phelan, 2001) and treats stigma as an objectively measurable cultural phenomenon.

Social group. To explore the degree to which intergroup evaluation in disadvantaged groups was moderated by the intergroup domain, we treated the different IAT versions as proxies for the intergroup domain in a series of mixed-effects meta regressions.

IAT D Scores. We first fitted a mixed-effects model, treating IAT version as a categorical moderator, dummy-coding each level of the moderator (i.e., each IAT version) and using disadvantaged group members' IAT *D* Scores as the dependent variable. This analysis yielded a significant moderation effect, $Q_M(8) = 885.89$, $p < .001$. We observed negative IAT effects for the Religion IAT, $d_z = -0.84$, $SE = 0.13$, $p < .001$, 95% *CI* [-1.09; -0.59], and Sexuality IAT, $d_z = -0.39$, $SE = 0.06$, $p < .001$, 95% *CI* [-0.50; -0.27], indicating ingroup favoritism. In contrast, we observed positive IAT effects for the Skin Tone IAT, $d_z = 0.45$, $SE = 0.06$, $p < .001$, 95% *CI* [0.33; 0.57], the Weight IAT, $d_z = 0.86$, $SE = 0.06$, $p < .001$, 95% *CI* [0.75; 0.98], the Age IAT, $d_z = 1.33$, $SE = 0.06$, $p < .001$, 95% *CI* [1.21; 1.45], and the Disability IAT, $d_z = 0.89$, $SE = 0.21$, $p < .001$, 95% *CI* [0.47; 1.31], indicating outgroup favoritism. The remaining null IAT effects indicated neither ingroup nor outgroup favoritism on the Arab IAT, $d_z = 0.33$, $SE = 0.22$, $p = .135$, 95% *CI* [-0.10; 0.75], and the Race IAT, $d_z = 0.15$, $SE = 0.08$, $p = .060$, 95% *CI* [-0.01; 0.30].

One-item preference scores. We next fitted the same mixed-effects model, now using disadvantaged group members' one-item preference scores as the dependent variable. This analysis also yielded a significant moderation effect, $Q_M(8) = 612.32$, $p < .001$. We observed negative effects for one-item preference scores for the Black versus White comparisons, $d_z = -0.39$, $SE = 0.07$, $p < .001$, 95% *CI* [-0.53; -0.25], religious ingroup versus outgroup comparisons, $d_z = -1.90$, $SE = 0.13$, $p < .001$, 95% *CI* [-2.14; -1.65], and gay versus straight comparisons, $d_z = -$

0.58, *SE* = 0.06, $p < .001$, 95% *CI* [-0.69; -0.47], indicating ingroup favoritism. In contrast, we observed positive effects for dark-skinned versus light-skinned comparisons, $d_z = 0.25$, *SE* = 0.07, $p < .001$, 95% *CI* [0.11; 0.38], overweight versus normal weight comparisons, $d_z = 0.84$, *SE* = 0.06, $p < .001$, 95% *CI* [0.73; 0.95], and old versus young comparisons, $d_z = 0.31$, *SE* = 0.06, $p < .001$, 95% *CI* [0.18; 0.43], indicating outgroup favoritism. The remaining null effects indicated neither ingroup nor outgroup favoritism for disabled versus abled comparisons, $d_z = 0.23$, *SE* = 0.21, $p = .286$, 95% *CI* [-0.19; 0.65] and Arab versus White comparisons, $d_z = -0.24$, *SE* = 0.22, $p = .281$, 95% *CI* [-0.67; 0.19].

Feeling thermometer difference scores. We fitted a third mixed-effects model in a similar fashion, now using disadvantaged group members' feeling thermometer difference scores as the dependent variable. This analysis also yielded a significant moderation effect, $Q_M(8) = 277.92$, $p < .001$. We observed negative effects for Black versus White evaluations, $d_z = -0.54$, *SE* = 0.07, $p < .001$, 95% *CI* [-0.67; -0.41], religious ingroup versus outgroup evaluations, $d_z = -1.17$, *SE* = 0.12, $p < .001$, 95% *CI* [-1.40; -0.93], gay versus straight evaluations, $d_z = -0.46$, *SE* = 0.05, $p < .001$, 95% *CI* [-0.56; -0.35], and dark-skinned versus light-skinned evaluations, $d_z = -0.18$, *SE* = 0.06, $p = .005$, 95% *CI* [-0.30; -0.05], indicating ingroup favoritism. In contrast, we observed positive effects for overweight versus normal weight evaluations, $d_z = 0.29$, *SE* = 0.05, $p < .001$, 95% *CI* [0.19; 0.40], indicating outgroup favoritism. The remaining null effects indicated neither ingroup nor outgroup favoritism for old versus young evaluations, $d_z = 0.11$, *SE* = 0.07, $p = .120$, 95% *CI* [-0.03; 0.25], Arab versus White evaluations, $d_z = -0.30$, *SE* = 0.26, $p = .250$, 95% *CI* [-0.81; 0.21], and disabled versus abled evaluations, $d_z = -0.11$, *SE* = 0.20, $p = .594$, 95% *CI* [-0.50; 0.29].

Conservatism. To test the relationship between disadvantaged groups' conservatism and intergroup evaluation at the level of the social group, we fitted three separate mixed-effects meta-regression models with self-reported conservatism aggregated at the sample level as continuous moderators (see Table 2, column 4, for sample level aggregates of conservatism), and using disadvantaged groups' sample-aggregated IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores as dependent variables. Figure 5 depicts the relationship between sample averages of conservatism (y-axes) and sample averages of intergroup evaluations (x-axes) among disadvantaged groups, with each panel corresponding to a different measure of intergroup evaluation.

IAT D Scores. First, we fitted a mixed-effects model with sample aggregates of self-reported conservatism as a continuous moderator, using disadvantaged groups' IAT *D* Scores as dependent variable. We observed a significant moderating effect, $Q_M(1) = 16.53$, $p < .001$, which accounted for 20.18% of the heterogeneity. This indicates that samples with higher averages of self-reported conservatism were more likely to demonstrate implicit outgroup favoritism (see Figure 5, Panel A).

One-item preference scores. Again, we fitted a mixed-effects model with sample aggregates of self-reported conservatism as a continuous moderator, now using disadvantaged groups' one-item preference scores as dependent variable. We observed a significant moderating effect, $Q_M(1) = 14.83$, $p < .001$, which accounted for 15.98% of the heterogeneity. This indicates that samples with higher averages of self-reported conservatism were more likely to demonstrate outgroup favoritism on one-item preference scores (see Figure 5, Panel B).

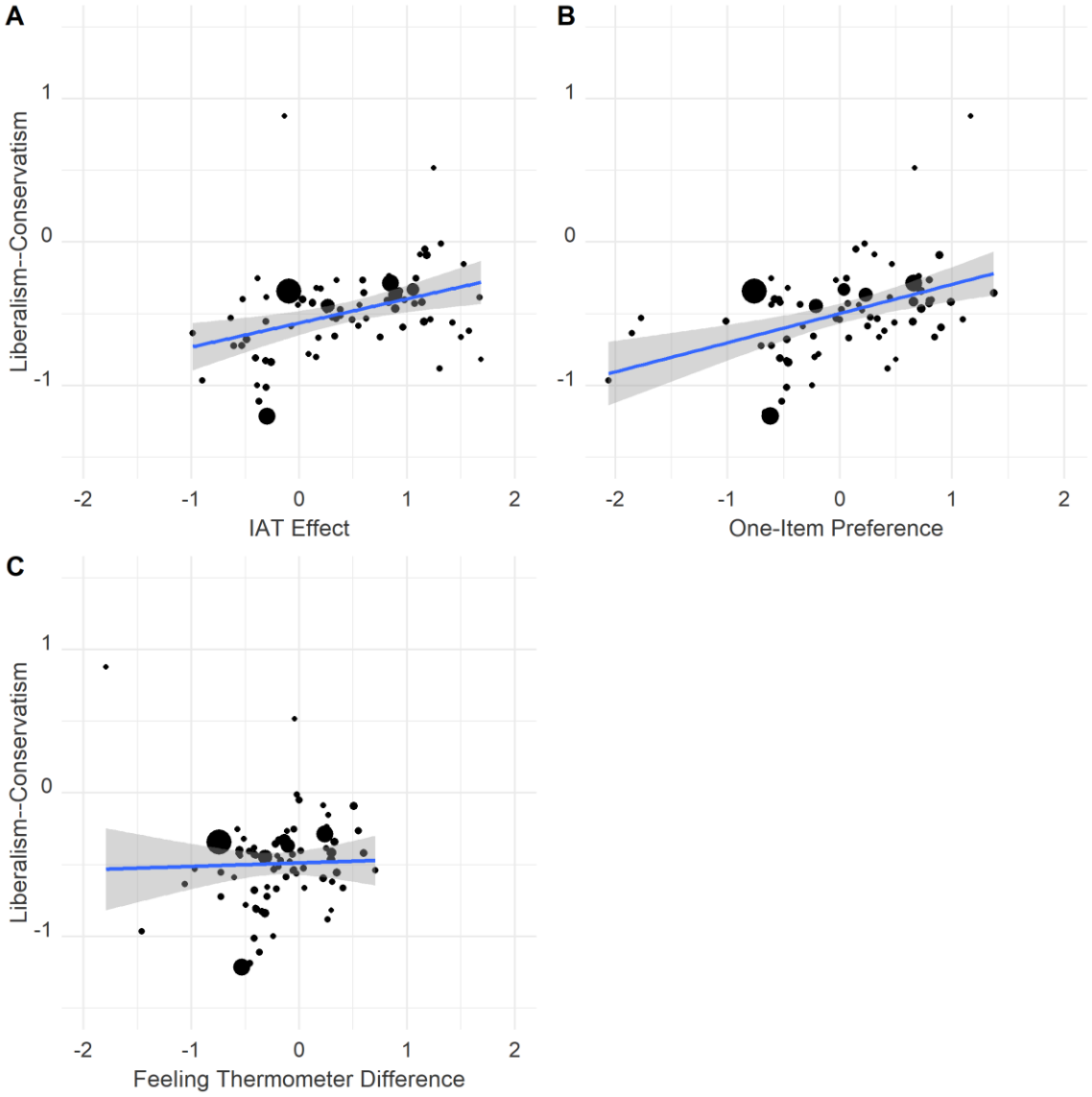

Figure 4. Scatterplot depicting the relationship between conservatism and intergroup evaluations among disadvantaged groups. Y-axes reflect disadvantaged groups' sample-level mean Cohen's d_z for conservatism, measured on a one-item 6- and/or 7-point scale. X-axes reflect mean Cohen's d_z for IAT *D* Scores (Panel A), one-item preference scores (Panel B), and feeling thermometer difference scores (Panel C). Each circle corresponds to a different social group, with circle size reflecting sample size. Positive values on the y-axes indicate more conservative attitudes. Positive values on the x-axes indicate more favorable evaluations of the advantaged group relative to the disadvantaged group, which reflects outgroup favoritism for these samples of disadvantaged groups.

Feeling thermometer difference scores. Lastly, we fitted the same mixed-effects model with feeling thermometer difference scores as the dependent variable and sample aggregates of self-reported conservatism as a continuous moderator. Here, we did not observe a significant

moderating effect, $Q_M(1) = 0.78$, $p = .376$, with conservatism accounting for only 1.18% of the heterogeneity in feeling thermometer difference scores. This indicates that sample averages of self-reported conservatism were unrelated to sample averages of feeling thermometer difference scores (see Figure 5, Panel C).

Stigma. We operationalized stigma in terms of how a disadvantaged group is evaluated by people who are not members of that social group (i.e., "the rest of society"). Specifically, stigma estimators were manifest in the present analyses using three measures: IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores collected from all participants who indicated not belonging to the disadvantaged group in each study. We used these stigma measures as continuous moderators in a series of mixed-effects meta-regression models, with disadvantaged groups' IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores as dependent variables.

Figure 6 depicts the relationship between the three measures of stigma and the three measures of intergroup evaluation among disadvantaged groups, with positive values on both the x- and y-axes indicating more favorable evaluations of the advantaged group relative to the disadvantaged group. Consequently, positive values on the y-axes reflect outgroup favoritism among disadvantaged groups, and positive values on the x-axes reflect higher levels of stigma against disadvantaged groups.

IAT D Scores. We first fitted three separate mixed-effects meta-regression models with the three stigma measures as continuous moderators, using disadvantaged groups' IAT *D* Scores as dependent variable. We observed a significant moderation effect of IAT *D* Score stigma on disadvantaged groups' IAT *D* Scores, $Q_M(1) = 170.55$, $p < .001$, accounting for 71.31% of

heterogeneity; a non-significant moderation effect of one-item preference score stigma on disadvantaged groups' IAT *D* Scores, $Q_M(1) = 3.57$, $p = .059$, accounting for 3.53% of heterogeneity; and a significant moderation effect of feeling thermometer difference score stigma on disadvantaged groups' IAT *D* Scores, $Q_M(1) = 7.28$, $p = .007$, accounting for 8.08% of heterogeneity.[17] Taken together, these results indicate that disadvantaged groups displayed more implicit outgroup favoritism to the extent that their social group was stigmatized by others in terms of the implicit measure, an effect which accounted for substantial amounts of the heterogeneity. The same descriptive but non-significant trend was observed in terms of explicit, one-item preference score stigma. However, the opposite trend was observed in terms of explicit, feeling thermometer difference score stigma (see Figure 6, Panels A, B, C).

One-item preference scores. We fitted the same three mixed-effects meta-regression models with the three stigma measures as continuous moderators, this time using disadvantaged groups' one-item preference scores as dependent variable. We observed a significant moderation effects of IAT *D* Score stigma on disadvantaged groups' one-item preference scores, $Q_M(1) = 23.28$, $p < .001$, accounting for 23.81% of heterogeneity; a significant moderation effect of one-item preference score stigma on disadvantaged groups' one-item preference scores, $Q_M(1) = 25.05$, $p < .001$, accounting for 25.44% of heterogeneity; and a non-significant moderation effect

[17] Visual inspection of the Figure 6 (Panel C) suggests that this moderation effect was in the opposite direction than expected. Disadvantaged groups' IAT *D* Scores were higher (indicating more outgroup favoritism) the *less* they were stigmatized by others on feeling thermometers.

of feeling thermometer difference score stigma on disadvantaged groups' one-item preference scores, $Q_M(1) = 1.38$, $p = .239$, accounting for 0.36% of heterogeneity. Taken together, these results indicate that disadvantaged groups displayed more explicit outgroup favoritism to the extent that their social group was stigmatized by others in terms of the implicit measure and the explicit one-item preference measure. However, feeling thermometer difference score stigma was unrelated to disadvantaged groups' one-item preference scores (see Figure 6, Panel D, E, F).

Feeling thermometer difference scores. Finally, we fitted the same three mixed-effects meta-regression models with the three stigma measures as continuous moderators, this time using disadvantaged groups' feeling thermometer difference scores as dependent variable. We observed a significant moderation effects of IAT *D* Score stigma on disadvantaged groups' feeling thermometer difference scores, $Q_M(1) = 27.62$, $p < .001$, accounting for 30.00% of heterogeneity; a significant moderation effect of one-item preference score stigma on disadvantaged groups' feeling thermometer difference scores, $Q_M(1) = 12.68$, $p < .001$, accounting for 17.29% of heterogeneity; and a non-significant moderation effect of feeling thermometer difference score stigma on disadvantaged groups' feeling thermometer difference scores, $Q_M(1) = 3.36$, $p = .067$, accounting for 4.62% of heterogeneity. These results indicate that disadvantaged groups displayed less explicit ingroup favoritism on feeling thermometer difference scores to the extent that their social group was stigmatized by others on the implicit measure and (descriptively) both explicit measures (see Figure 6, Panel G, H, I).

Conservatism versus stigma. According to SJT, both conservatism and stigma should moderate intergroup bias among disadvantaged groups, but the theory makes no predictions about the relative influences of each of these constructs. The present research provides an opportunity

to examine whether conservatism and stigma account for heterogeneity above and beyond the heterogeneity accounted for by the other. We examined this in a meta-analytic framework, by fitting hierarchical multivariate meta-regression models with intergroup evaluation as the dependent variable and sample averages of conservatism and the three stigma measures as moderators (see Harrer, Cuijpers, Furukawa, & Ebert, 2019).[18] For each measure of intergroup evaluation, we first fitted a reduced model with disadvantaged groups' sample aggregates of conservatism as a continuous moderator. Next, we fitted the full model, adding the three measures of stigma. We then compared the model fit of both models by using a likelihood ratio test (LRT) and by comparing *Akaike's Information Criterion* (AIC) values for both models. Lastly, in order to control for Type I error, we tested the robustness of the full model by performing a permutation test with 1,000 iterations (see Higgins & Thompson, 2004; Viechtbauer, Lopez-Lopez, Sanchez-Meca, & Marin-Martinez, 2015).

---

[18] Correlational analyses of the three measures of stigma indicate that they are not highly correlated, thus making multicollinearity unlikely: IAT *D* Score stigma and one-item preference score stigma, $r(71) = .15$, $p = 0.20$, 95% *CI* [-0.08; 0.37]; IAT *D* Score stigma and feeling thermometer difference score stigma, $r(71) = -.21$, $p = 0.08$, 95% *CI* [-0.42; 0.02]; one-item preference score stigma and feeling thermometer difference score stigma, $r(71) = .22$, $p = 0.06$, 95% *CI* [-0.01; 0.43].

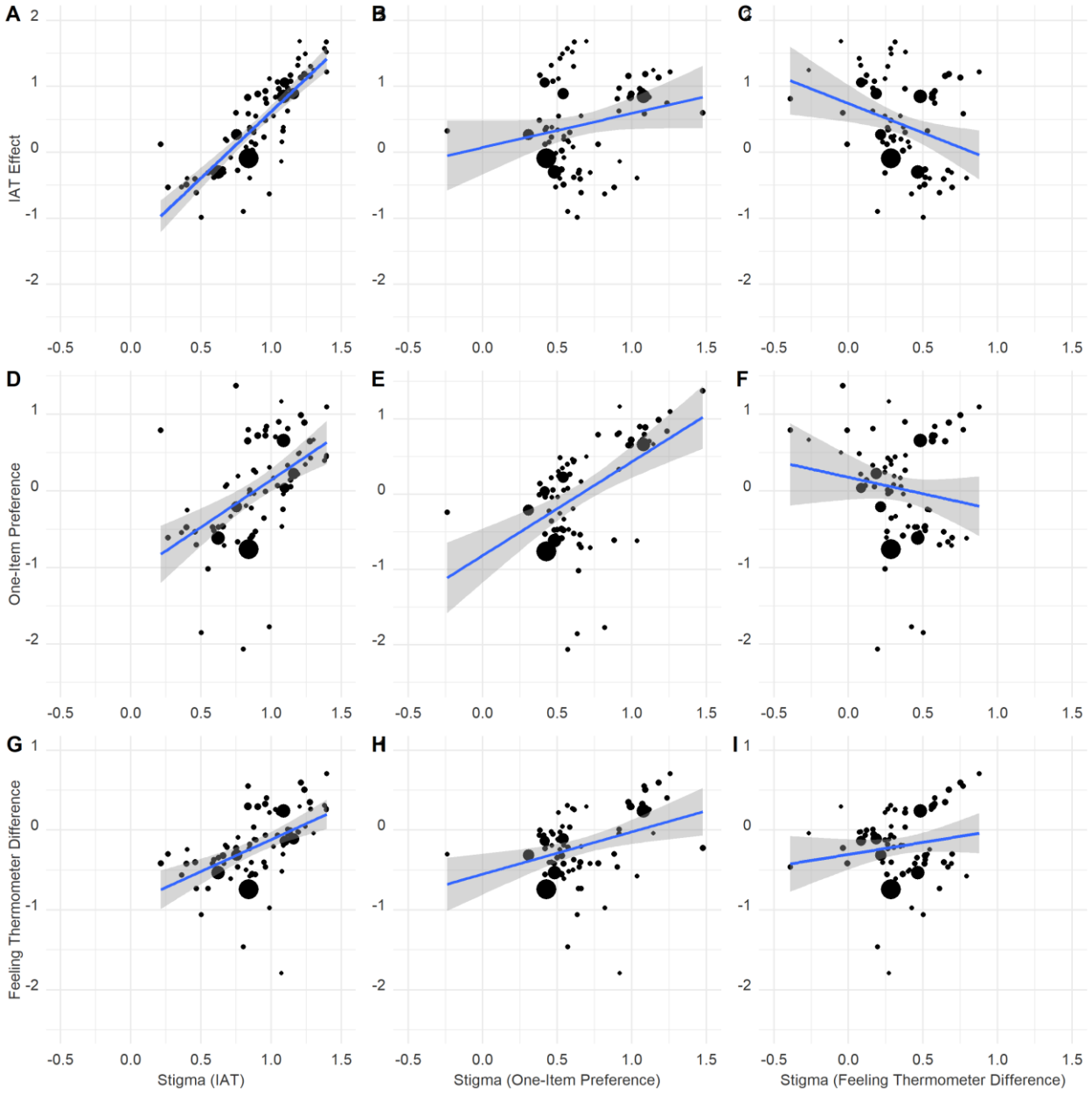

Figure 5. Scatterplots depicting the relationship between stigma and intergroup evaluation among disadvantaged groups. Y-axes reflect mean Cohen's d_z for disadvantaged groups' IAT *D* Scores (Panels A-C), one-item preference scores (Panel D-F), and feeling thermometer difference scores (Panel G-I). X-axes reflect stigma, operationalized as mean Cohen's d_z for non-disadvantaged group members' IAT *D* Scores (left column), one-item preference scores (middle column), and feeling thermometer difference scores (right column). Positive values on both the x- and y-axes indicate more favorable evaluations of the advantaged group relative to the disadvantaged group. Consequently, positive values on the y-axes reflect outgroup favoritism among disadvantaged groups, and positive values on the x-axes reflect stigma against disadvantaged groups.

IAT D Scores. We first fitted a mixed-effects meta-regression model with sample averages of conservatism as a continuous moderator, using disadvantaged groups' sample

averages of IAT *D* Scores as dependent variable. We observed a significant overall moderating effect, $F_{1,71} = 16.36$, $p < .001$, which accounted for 21.29% of the heterogeneity. Next, we fitted the full mixed-effects meta-regression model, adding the three measures of stigma. Again, we observed a significant overall moderating effect, $F_{4,68} = 52.30$, $p < .001$. Crucially, the moderating effect of conservatism became non-significant in the full model, $\beta = 0.17$, $p = .325$, 95% *CI* [-0.17; 0.51], whereas IAT *D* Score stigma, $\beta = 1.85$, $p < .001$, 95% *CI* [1.52; 2.17], one-item preference score stigma, $\beta = 0.33$, $p = .040$, 95% *CI* [0.02; 0.64], and feeling thermometer difference score stigma, $\beta = -0.49$, $p = .014$, 95% *CI* [-0.87; -0.10], remained significant predictors of disadvantaged groups' IAT *D* Scores.[19] The full model accounted for 76.68% of the heterogeneity. The likelihood ratio test indicated that the full model indeed had a better fit than the reduced model, $\chi^2 = 85.24$, $p < .001$, and AIC values were lower (indicating better fit) for the full model, $AIC_c = 64.69$, than for the reduced model, $AIC_c = 143.01$. Lastly, we performed a permutation test, providing evidence for the robustness of the full model, $F_{4,68} = 52.30$, $p = .001$. This indicates that the effect of disadvantaged groups' level of conservatism on implicit outgroup favoritism was fully accounted for by stigma. Moreover, stigma accounted for heterogeneity above and beyond the heterogeneity accounted for by conservatism. In other words, at the social group level, stigma explained more variance in implicit intergroup evaluation than conservatism, and conservatism had no independent effect on IAT *D* Scores.

---

[19] Note however that other participants' feeling thermometer difference scores were negatively correlated with disadvantaged groups' IAT *D* Scores.

One-item preference scores. Again, we fitted a mixed-effects meta-regression model with sample averages of conservatism as a continuous moderator, now using disadvantaged groups' sample averages of one-item preference scores as dependent variable. We observed a significant overall moderating effect, $F_{1,71} = 15.22$, $p < .001$, which accounted for 17.06% of the heterogeneity. Next, we fitted the full mixed-effects meta-regression model, adding the three measures of stigma. Again, we observed a significant overall moderating effect, $F_{4,68} = 16.16$, $p < .001$. Crucially, the moderating effect of conservatism became non-significant in the full model, $\beta = 0.27$, $p = .238$, 95% *CI* [-0.18; 0.73], whereas IAT *D* Score stigma, $\beta = 0.87$, $p < .001$, 95% *CI* [0.40; 1.34] and one-item preference score stigma, $\beta = 1.13$, $p < .001$, 95% *CI* [0.67; 1.59], remained significant predictors of disadvantaged groups' one-item preference scores. The effect of feeling thermometer difference score stigma on disadvantaged groups' one-item preference scores was non-significant, $\beta = -0.42$, $p = .139$, 95% *CI* [-0.98; 0.14]. The full model accounted for 48.38% of the heterogeneity. The likelihood ratio test indicated that the full model indeed had a better fit than the reduced model, $\chi^2 = 33.64$, $p < .001$, and AIC values were lower for the full model, $AIC_c = 118.11$, than for the reduced model, $AIC_c = 144.82$. Lastly, we performed a permutation test, providing evidence for the robustness of the full model, $F_{4,68} = 16.16$, $p = .001$. This indicates that the effect of disadvantaged groups' level of conservatism on explicit intergroup evaluation was fully accounted for by stigma. Moreover, stigma accounted for heterogeneity above and beyond the heterogeneity accounted for by conservatism. In other words, at the social group level, stigma explained more variance in explicit intergroup evaluation than conservatism, and conservatism had no independent effect on one-item preference scores.

Feeling thermometer difference scores.

Lastly, we fitted a mixed-effects meta-regression model with sample averages of conservatism as a continuous moderator, now using disadvantaged groups' sample averages of feeling thermometer difference scores as dependent variable. The overall moderating effect was non-significant, $F_{1,71} = 0.81$, $p = .371$, and accounted for 2.60% of the heterogeneity. Next, we fitted the full mixed-effects meta-regression model, adding the three measures of stigma. Here, we observed a significant overall moderating effect, $F_{4,68} = 13.25$, $p < .001$. Crucially, the moderating effect of conservatism remained non-significant in the full model, $\beta = -0.14$, $p = .368$, 95% *CI* [-0.46; 0.17], whereas IAT *D* Score stigma, $\beta = 0.87$, $p < .001$, 95% *CI* [0.55; 1.18] and one-item preference score stigma, $\beta = 0.44$, $p = .007$, 95% *CI* [0.12; 0.75], were significant predictors of disadvantaged groups' feeling thermometer difference scores, and the effect of feeling thermometer difference score stigma was non-significant, $\beta = 0.38$, $p = .054$, 95% *CI* [-0.01; 0.77]. The full model accounted for 48.36% of the heterogeneity. The likelihood ratio test indicated that the full model indeed had a better fit than the reduced model, $\chi^2 = 40.14$, $p < .001$, and AIC values were lower for the full model, $AIC_c = 61.13$, than for the reduced model, $AIC_c = 94.34$. Lastly, we performed a permutation test, providing evidence for the robustness of the full model, $F_{4,68} = 13.25$, $p = .001$. This indicates that the effect of disadvantaged groups' level of conservatism on explicit intergroup evaluation was fully accounted for by stigma. Moreover, stigma accounted for heterogeneity above and beyond the heterogeneity accounted for by conservatism. In other words, at the social group level, stigma explained more variance in explicit intergroup evaluation than conservatism, and conservatism had no independent effect on feeling thermometer difference scores.

Taken together, group-level meta-analyses extend SJT by testing its predictions at a new unit of analyses. Exploratory analyses using social group as a moderator revealed that intergroup evaluations in disadvantaged groups were moderated by the intergroup domain: Whereas some disadvantaged groups consistently displayed outgroup favoritism, others consistently displayed ingroup favoritism, and yet others displayed diverging patterns on implicit and explicit measures. Additionally, group-level conservatism consistently moderated disadvantaged groups' implicit intergroup evaluations, but inconsistently moderated their explicit intergroup evaluations. Similarly, stigma operationalized in terms of others' implicit intergroup evaluations consistently moderated disadvantaged groups' intergroup evaluations, whereas stigma operationalized in terms of others' explicit intergroup evaluations inconsistently moderated disadvantaged groups' intergroup evaluations. Finally, when both conservatism and stigma were entered into the same models, only stigma was consistently related to intergroup evaluations of disadvantaged groups.

General Discussion

The present research used large datasets from 73 samples of online participants collected in 14 countries to investigate moderators of intergroup evaluation in a wide variety of disadvantaged groups. We tested SJT's predictions, whenever possible, at two levels of analysis: the individual level and the social group level. At the individual level, members of disadvantaged groups on average displayed a medium-sized effect of outgroup favoritism on the IAT, but either a small effect of ingroup favoritism or no intergroup preference on two explicit measures. These findings are consistent with SJT's predicted dissociation between implicit and explicit measures (Jost et al., 2004). In follow-up, exploratory analyses that treated social group as a moderator, intergroup evaluation among disadvantaged groups was moderated by the intergroup domain:

Whereas some disadvantaged groups consistently displayed outgroup favoritism, others consistently displayed ingroup favoritism, and yet others displayed diverging patterns on implicit and explicit measures. Additionally, and supporting the predictions of SJT, implicit and explicit intergroup evaluations were moderated by self-reported conservatism. Importantly, the magnitude of effect sizes depended on the level of analysis, indicating small effects at the individual level and medium-sized effects at the social group level. Lastly, at the social group level, disadvantaged groups displayed higher levels of outgroup favoritism the more negatively their own social group was evaluated relative to an advantaged outgroup in their societal context, which supports the hypothesis that stigma is related to outgroup favoritism in disadvantaged groups (Jost & Banaji, 1994; Jost et al., 2004).

SJT at the Level of the Individual Versus the Social Group

The present research underscores the utility of conducting analyses at both the individual and social group level. By testing the predictions of SJT at two levels of analysis, our findings provide insight into qualitatively distinct psychological processes: the individual-level analyses reflect individual differences, whereas the group-level analyses reflect group processes.

Conservatism. We observed marked differences in the magnitude of the relationships between conservatism and intergroup evaluations between units of analysis: Whereas moderator analyses conducted at the social group level (i.e., between samples) revealed that conservatism accounted for considerable proportions of the variance of intergroup evaluations between disadvantaged groups, analyses conducted at the individual level (i.e., within-samples) revealed that conservatism accounted for little variance in intergroup evaluations within disadvantaged groups. These large differences between between-sample and within-sample analyses are

striking—but have been observed in other research on implicit bias as well (sf. Payne, Vuletich, & Lundberg, 2017). One explanation for these large differences might be that within-sample correlations were attenuated by the relative unreliability of the measures (e.g., Hunter & Schmidt, 2007), whereas aggregation at the social group level in our between-sample analyses likely reduced measurement error (e.g., Rivers, Rees, Calanchini, & Sherman, 2017; Rushton et al., 1983). However, the differences appear too pronounced to gloss over as statistical artifact. We therefore offer and discuss speculations about the underlying processes that may explain these differences between levels of analysis that we hope may inspire future research in this domain.

At the individual level, conservatism of individual members of disadvantaged groups is the unit of analysis. We assume that disadvantaged group members' level of conservatism reflects individual differences in the preference for inequality, acceptance of the status quo, or system justifying beliefs more generally (Jost et al., 2004; Jost et al., 2003), as well as other psychological correlates of conservatism (for a review, see Hodson & Dhont, 2015). In turn, individual differences in conservatism are related to the extent to which some group members display more outgroup favoritism than other group members. The small average effect size of these correlations suggests that the differences between individual members' level of conservatism are relatively weakly related to their individual tendency to display ingroup or outgroup favoritism.

At the social group level, conservatism of the disadvantaged group as a whole is the unit of analysis. Disadvantaged groups inhabit different positions on the conservatism-liberalism spectrum, with some groups on average leaning less liberal (e.g., overweight participants) than

others (e.g., gay and Lesbian participants).[20] We propose that disadvantaged groups' position along the conservatism-liberalism spectrum reflect group-level processes, such as group histories, social norms, or cultural traditions. In the present research, we identify stigma as a group-level process that moderates (and, in fact, fully accounts for) the relationship between group-level conservatism and intergroup evaluations among disadvantaged groups. There are at least two possible explanations for this finding.

One possibility is that the measures of conservatism and stigma may constitute different operationalizations of the same latent construct, with stigma being the superior measure of the latent variable. However, we deem this explanation less likely, given the conceptual differences between stigma—operationalized here as the negative group evaluations by others—and ideological self-placement (cf. Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016). Another possibility is that stigma might influence conservatism in disadvantaged groups. From this perspective, we propose that the relative frequency of experiences of stigmatization (which can include individuals personally experiencing stigma, as well as hearing reports of stigmatization from fellow group members) may lead disadvantaged groups as a whole to lean more or less conservative over time.

Consistent with the possibility that stigma increases conservatism among disadvantaged groups, members of disadvantaged groups who internalize negative stereotypes and evaluations

---

[20] Note that samples leaned fairly liberal on average, as indexed by their negative sign in Table 2, fourth column.

of their group tend to assimilate to the dominant culture (David, Schroeder, & Fernandez, 2019). Furthermore, some members of disadvantaged groups cope with stigma by gravitating towards ideological beliefs that provide a sense of safety, such as authoritarianism (Henry, 2011). Thus, the relationships between stigma and ideological beliefs in disadvantaged groups might reflect a self-regulatory strategy. Similarly, SJT proposed that system justifying beliefs would serve a "palliative function" (Jost & Hunyady, 2002), helping members of disadvantaged groups who are stigmatized to cope with negative emotions. Taken together, these ideas about relationships between stigma and ideological beliefs were formulated at the individual level, but could provide a framework for future theorizing at the social group level. Based on this framework, conservatism as a group-level construct may not primarily (or necessarily) reflect individual political ideology, but rather a group-based cultural adaptation process, by which members of disadvantaged groups adhere to a more or less conservative group norm in response to the level of stigma faced by their group. That said, such causal claims remain speculative, given the correlational nature of the present research. Future research—ideally longitudinal—is necessary to investigate a causal effect of stigma on group conservatism, as well as potential cultural or social processes that may mediate and/or moderate such effect(s).

In addition to stigma, future research might investigate other group-level processes to explain disparate relationship between conservatism and intergroup evaluation at the individual versus social group levels. One such group-level process might be *group consciousness* (e.g., Duncan, 1999; Gurin, Miller, & Gurin, 1980; see Ashmore, Deaux, & McLaughlin-Volpe, 2004, for a related concept), the tendency to reflect on the ingroup's relative position in society. Group consciousness can vary between people but also between groups (Gurin et al., 1980). Our finding that conservatism is more strongly related to intergroup evaluations at the social group versus

individual level might suggest that disadvantaged groups differ in levels of group consciousness more so than do disadvantaged individuals. Other group-level processes that relate to intergroup evaluations, such as entitativity (Effron & Knowles, 2015) or the permeability of group boundaries (Bettencourt, Charlton, Dorr, & Hume, 2001), might also help to explain the diverging findings observed here. By continuing to investigate the relationship between group status and intergroup evaluations at multiple levels of analysis, future research may build on the present research to more fully develop SJT as a group-level theory.

Stigma. In the present research, conservatism explained substantial variance in intergroup evaluation between social groups, but this variance was fully accounted for by stigma. Importantly, this relationship was not moderated by measurement type: at the group level, both implicit and explicit intergroup evaluations in disadvantaged groups were related to stigma. One possible interpretation of these results is that stigma influences intergroup evaluations—and with this speculation go the usual caveats about causal claims and correlational data. While future, experimental work is necessary to support this claim, this pattern of results is nevertheless consistent with SJT's proposed "internalization of inferiority" (Jost et al., 2004; Jost & van der Toorn, 2012), as well as with the more general notion that intergroup evaluation depends on how social groups are evaluated by society (Allport, 1954; Dasgupta, 2004; Lane et al., 2005; Livingston, 2002).

The present research not only supports the existing literature on intergroup evaluation among disadvantaged groups, but also extends it with novel findings. For example, prior individual-level research has usually operationalized stigma subjectively, in terms of disadvantaged group members' perceptions of how their social group is evaluated by others

(Livingston, 2002; Rudman et al., 2002). In contrast, the present research treats stigma as an objective cultural phenomenon, operationalized as the rest of society's measured evaluations of the disadvantaged group. Consequently, our work offers a novel perspective on why disadvantaged groups sometimes display outgroup favoritism: Disadvantaged groups' evaluations of their own groups appear to align with everyone else's evaluations.

Our finding that stigma moderated disadvantaged groups' intergroup evaluation would seem to be consistent with the Bias of Crowds model (BoC; Payne et al., 2017). Based on the principle of concept accessibility, the BoC model suggests that implicit bias does not merely reflect personal attitudes but, instead, reflects reflects context-related attitude accessibility. Consequently, implicit bias should be stronger in contexts where people are more frequently exposed to direct or indirect expressions of intergroup bias. According to the BoC model, implicit bias is best understood to reflect concepts that are activated by contextual cues and briefly pass through peoples' minds. From this perspective, implicit bias is better conceptualized as a stable property of places and situations rather than a stable property of people. Thus, the BoC model would seem to suggest that the strong relationship between disadvantaged groups' intergroup evaluations and the rest of society's intergroup evaluations (i.e., stigma) observed in the present research reflects a common context-related cause, such as structural inequality. Moreover, our findings that stigma moderates intergroup evaluations across both implicit and explicit measures may reflect an extension of BoC which, to date, is only articulated in terms of implicit bias—which, in turn, may suggest that the group as unit of analysis is more relevant to BoC than is the measurement approach.

That said, the BoC perspective does not perfectly explain the pattern of results reported here. For example, disadvantaged religious and sexual minority groups consistently demonstrated ingroup favoritism across all measures of intergroup bias. However, these groups are objectively stigmatized by the rest of society, in that non-disadvantaged groups' evaluations reflect preferences for the advantaged over disadvantaged groups. Thus, in at least some cases, the intergroup biases of disadvantaged groups do not perfectly correspond to the intergroup biases of the rest of society, which suggests either that certain biases do not reflect a common (e.g., structural) source, or that the influence of this common source is moderated by other processes (e.g., other individual differences or group processes; Branscombe, Schmitt, & Harvey, 1999; Cadinu & Rothbart, 1996; Jost et al., 2004; Tajfel & Turner, 1979). Future research should continue to examine this.

Open Questions and Future Directions

Conservatism and system justifying beliefs. Our analyses provide support for the prediction that as political conservatism increases, outgroup favoritism becomes more likely among disadvantaged groups (Jost et al., 2004). Nevertheless, the present research points to the need for more theorizing about *how* conservatism translates into intergroup evaluation in disadvantaged groups because the moderating effects of conservatism were inconsistent across attitude measures: the relationships between conservatism and intergroup evaluations observed on the IAT and one-item preference measure did not persist for the feeling thermometer. Research has so far primarily tried to explain links between conservatism and outgroup attitudes in dominant groups (see Hodson & Dhont, 2015, for a review), but has not yet articulated the mechanism by which conservative ideology might shape intergroup evaluation among

disadvantaged groups. Future work on the underlying psychological processes and mechanisms of this relationship might benefit from considering the following three perspectives.

A first important step towards a more process-oriented understanding would be to investigate whether higher conservatism among disadvantaged groups is related to more positive evaluations of advantaged outgroups, more negative evaluations of the disadvantaged ingroup, or both. A variety of theoretical perspectives propose that ingroup favoritism primarily reflects positive ingroup evaluations rather than negative outgroup evaluations (e.g., Allport, 1954; Brewer, 1999; Brewer & Campbell, 1976; Greenwald & Pettigrew, 2014; Mummendey & Otten, 1998). However, to date, no perspective makes clear predictions about the relative contributions of positive and negative evaluations to outgroup favoritism in general or among disadvantaged groups specifically, nor are there clear predictions about how conservatism might moderate these evaluations. Developing more refined theories about the relationship between conservatism and outgroup favoritism among disadvantaged groups will help to advance research in this domain.

Second, further research is needed to determine which aspects of conservatism are related to intergroup evaluation among disadvantaged groups. For example, Jost and colleagues (2003) proposed that conservatism is based on two core ideologies: the opposition to change and preference for inequality. In the present research, we relied on a single-item political orientation measure (ranging from conservative to liberal), so we were not able to disentangle the contributions of these two core ideologies. Future research might employ scales assessing sub-components of conservatism, such as acceptance of inequality (e.g., Ho et al., 2015) or opposition to change (e.g., White, Kinney, Danek, Smith, & Harben, 2020), in order to better understand how conservatism moderates intergroup evaluation among disadvantaged groups.

Third, SJT predicts that system justifying tendencies are related to intergroup evaluations (Jost et al., 2004), but conservatism is only one possible manifestation of system justifying beliefs (see Jost & Hunyady, 2005). For example, system justifying beliefs have also been operationalized in terms of economic system justification (Jost & Thompson, 2000), general levels of system justification (Kay & Jost, 2003), and social dominance orientation (SDO; Pratto et al., 1994). Supporting the predictions of SJT (Jost et al., 2004), economic system justification was positively related to outgroup favoritism in South Italians (a disadvantaged group, see Jost et al., 2002). Yet, other studies have found no relationship between economic system justification and outgroup favoritism (Jost & Thompson, 2000, Study 4), or between SDO and intergroup evaluation (Ashburn-Nardo et al., 2003), among Black participants. Complementing previous findings, we conducted auxiliary analyses on a small subset of Project Implicit datasets, which suggest that relationships between system justifying beliefs and outgroup favoritism may depend on the specific measure of system justifying beliefs. These analyses (and their limitations) are described in greater detail in Appendix A. Still, more research and theorizing are needed to clarify which ideologies and belief systems are related to intergroup evaluations among members of disadvantaged groups.

Lastly, and more broadly, the motivational processes underlying intergroup evaluation among members of disadvantaged groups will be better understood to the extent that each hypothesized motivation is measured directly. Specifically, SJT posits that members of disadvantaged groups' motives to see themselves and their ingroup positively are often in conflict with their motives to justify and defend the social systems in which they live (Jost & Banaji, 1994; Jost et al., 2004). The present research relied on a measure of system justification motives that does not distinguish among these three motivations (i.e., political conservatism), so our

findings do not provide clear insights regarding this motivational conflict. We thus strongly encourage researchers to directly investigate these proposed motivational structures underlying intergroup evaluations by separately measuring system, social group, and ego motives (cf. Kay & Jost, 2014). Future research will benefit from using measures that provide sufficient granularity to differentiate between motives, and perhaps provide insight into whether motivational conflicts underlie variations in intergroup evaluations among members of disadvantaged groups.

Intergroup domain. Exploratory analyses using social group as a moderator indicated that intergroup evaluation in disadvantaged groups was characterized by a high degree of variability. Some disadvantaged groups always displayed outgroup favoritism on both implicit and explicit measures (e.g., overweight or older participants), whereas other disadvantaged groups always displayed ingroup favoritism on both types of measures (e.g., religious or gay and Lesbian participants), and others displayed no preference on the implicit measure but ingroup favoritism on the explicit measure (e.g., Black participants). This pattern of results is not easily explained by SJT as it is currently articulated (Jost & Banaji, 1994; Jost et al., 2004), and thus, seems to suggest a boundary condition. We speculate here about why intergroup evaluation might vary across disadvantaged groups.

One recently-raised idea focuses on the role of societal discourses in shaping intergroup biases. Charlesworth and Banaji (2019) proposed that the extent to which society prioritizes issues might account for patterns of intergroup biases. From this perspective, our finding that disadvantaged groups' intergroup evaluations systematically aligned with the rest of society's intergroup evaluations (i.e., stigma) seems to suggest that both group's evaluations reflect a common influence of social priorities.

Another explanation for why intergroup evaluation might vary across disadvantaged groups focuses on social norms and other meta-evaluations. For example, age-related bias is not prohibited by strong egalitarian norms in Western culture: Bias against older people and in favor of younger people is among the largest and most consensual of biases against a social group in the United States (Levy & Banaji, 2002; Nosek et al., 2002, 2007). Similarly, social norms about the suppression of weight-related prejudice have been shown to be weaker compared to other forms of prejudice (e.g., Crandall & Eshleman, 2003; Crandall, Eshleman, & O'Brien, 2002; Degner & Wentura, 2009). Previous research has also highlighted weight-related bias as a pervasive and unique form of prejudice in that body weight is often regarded as controllable (e.g., Crandall, 1994). Consequently, overweight individuals are frequently seen as responsible (e.g., Tiggemann & Rothblum, 1997) and blamed for their weight and associated stigma—with overweight individuals often sharing these assumptions (e.g., Crandall, 1994; see Crandall, Merman, & Hebl, 2009, for a review). Given that norms and other meta-evaluations vary across groups, future research should investigate the extent to which they moderate intergroup evaluation in disadvantaged groups.

Effects of intergroup domain versus country. Building on the present research's focus on intergroup evaluations as they vary across social groups, future research might also investigate the extent to which intergroup evaluations vary across countries. The distinction between social groups and countries as unit of analysis is important because, on the one hand, some disadvantaged groups might generally be stigmatized more than other disadvantaged groups. However, on the other hand, some countries might be characterized by higher levels of overall stigma than others (e.g., Marini et al., 2013). To test the latter, we report in Appendix A a series of analyses in which country is included as a categorical moderator. We observed only

inconsistent country-level moderation effects on intergroup evaluations in disadvantaged groups. Specifically, country moderated intergroup evaluations on implicit measures, but only inconsistently moderated intergroup evaluations on explicit measures. Furthermore, auxiliary moderator analyses using country-level indices of cultural value dimensions did not reveal consistent relationships between cultural values and intergroup evaluations in disadvantaged groups (see Appendix A). Thus, country-level differences do not seem to be consistently related to intergroup evaluations in the context of the present research.

That said, these country-level analyses are complicated by two issues. First, given the rather low power of these analyses (n = 14 countries), they are not strong tests of the relationship between country-level factors and disadvantaged groups' intergroup evaluations. Second, different intergroup domains were studied in different countries. Many datasets were available for some countries (e.g., United States, United Kingdom), but relatively fewer datasets were available for other countries (e.g., Korea, Russia). Consequently, an analysis that treats country as a moderator is inherently biased because the moderator variable "country" is not independent of the moderator variable "intergroup domain". Thus, a moderation effect by country might reflect the fact that specific intergroup domains were examined in some countries but not others (i.e., biased selection). Taken together, our auxiliary analyses do not provide consistent evidence for country-level effects on intergroup evaluations in disadvantaged groups. Nevertheless, this issue should be addressed in future theory-driven research.

In contrast to the possibility that some countries might be characterized by higher levels of overall stigma than others, we consistently observed that some disadvantaged groups always displayed ingroup favoritism and others always displayed outgroup favoritism, regardless of their

country of origin. One potential interpretation of these moderation effects by intergroup domain is that, regardless of the societal context, certain disadvantaged groups are generally stigmatized more, whereas other disadvantaged groups are generally stigmatized less. In other words, differences in stigma associated with different social groups might (at least partly) explain the effects of intergroup domain on intergroup evaluations. As an illustration, we have restructured the caterpillar plot of IAT *D* Score effect sizes according to both country and effect size (Figure A5; Appendix). Additionally, in this restructured figure we have plotted the stigma estimates for each intergroup domain for each country. This reconfigured figure illustrates two main takeaways. First, a series of thumbnail copies emerge depicting a consistent pattern of effects across intergroup domains: Within each country's cluster, gay and Lesbian participants always demonstrate the highest degree of ingroup favoritism, and overweight and older participants always demonstrate the highest degree of outgroup favoritism. Second, the pattern of intergroup evaluations among disadvantaged groups closely aligns with stigma: Sexual identities are always associated with the lowest levels of stigma, and weight and age identities with the highest levels of stigma, leaving stigma levels related to ethnic identities in the middle of the distribution. Taken together, across countries, similar hierarchies emerge for both intergroup evaluations and stigma. Future research might build upon these observations to investigate the extent to which intergroup evaluations among disadvantaged groups are moderated by level of analysis (i.e., social group, country).

Internalization of inferiority. When taken at face value, the observed strong relations between stigma and intergroup evaluations in disadvantaged groups appear consistent with SJT's proposed "internalization of inferiority" (Jost et al., 2004, p. 881). Based on the concept of false consciousness, SJT proposes that disadvantaged groups internalize negative evaluations of and

stereotypes about the ingroup (Jost & Banaji, 1994). "[I]nterpreting outgroup favoritism as an indicator of internalization" (Jost et al., 2004, p. 894) is thus a straightforward deduction from this idea. This conclusion, however, would rely on the assumption that measures of intergroup evaluations are (direct or indirect) indicators of internalized attitudes. This assumption is tentative because SJT does not articulate a clear conceptualization of internalization.

The associative-propositional evaluation model (APE; Gawronski & Bodenhausen, 2006) offers two possible conceptualizations of internalization. On the one hand, internalization may refer to any process of associative learning that results in the formation of an internal representation (e.g., Gawronski & Bodenhausen, 2018). The APE model postulates that associative learning reflects spatio-temporal contingencies: the more often members of disadvantaged groups experience information about their group paired with negative evaluation, the more likely they are to form negative associations in long-term memory. Consequently, members of disadvantaged groups will have negative group-relevant experiences in proportion to the extent that their group is stigmatized. From this perspective, associative learning offers a parsimonious explanation of the relationship between stigma and intergroup evaluations observed in the present research—and especially the finding that stigma correlations are strongest for the implicit measure may be interpreted as supporting this rationale.

That said, we are not convinced that this associative learning account captures the gist of the internalization processes presumed to underlie the formation of "false consciousness" in system justification theory. Instead, "consciousness" suggests a degree of introspective awareness and/or deliberation on the subjective truth-value of information. This characterization corresponds to propositional learning, as articulated by the APE model. From this perspective, we

would conceptualize a negative group evaluation to be internalized only when the internal representation is based on some degree of awareness and/or subjective acceptance of validity of the negative evaluation of one's own social group.

To the extent that the internalization of inferiority can be conceptualized as an internal representation of associations between one's own group and negative evaluations, a number of further questions emerge. For example, internalized group evaluations might be understood as a structure of relatively stable associations in memory, formed either through associative or propositional learning processes in the course of socialization (Gawronski & Bodenhausen, 2018). In this case, the strength of the association determines the chronic accessibility of intergroup evaluations within the individual which, in turn, is reflected in responses on implicit measures such as the IAT. However, recent theoretical developments offer an alternative interpretation. For example, the BoC model conceptualizes implicit bias as a "social phenomenon that passes through individual minds" (Payne et al., 2017, p. 236) that does not require the presumption of internalized stable evaluations. In that sense, the effects of outgroup favoritism observed in the present research can be interpreted to reflect cultural contexts where stigmatizing associations are more readily activated than others.

The present research cannot address whether intergroup evaluation in disadvantaged groups reflects internalized stable evaluations, propositions, or situationally-dependent concepts. Instead, the questions we raise here about how internalization of inferiority is conceptualized by SJT (Jost et al., 2004; Jost & van der Toorn, 2012) illustrate that the theory might benefit from more clearly articulating assumptions about the psychological processes underlying intergroup evaluations in disadvantaged groups.

The use of feeling thermometers. In the present research, we observed consistent moderating effects of both conservatism and stigma on IATs and one-item preference measures, but inconsistent effects on feeling thermometer difference scores. Feeling thermometers have been labeled "notoriously unreliable" (Broockman, Kalla, & Aronow, 2015, p. 3), and we think that measurement error could have played a role in these inconsistent findings. Moreover, the calculation of difference scores from feeling thermometers likely increased measurement error, further reducing statistical power in the moderator analyses (Edwards, 1995; Hunter & Schmidt, 2007; Overall & Woodward, 1975). Additionally, IATs and one-item preference measures are structurally similar to one another, in that both target groups are evaluated relative to one another, whereas responses on feeling thermometers are made in the context of one target group at a time. We cannot discern the extent to which measurement error, structural differences, or other conceptual differences among measures contribute to the observed discrepancies. However, these discrepancies point to the need for further research into measurement properties and the psychological concepts underlying feeling thermometers.

Limitations of the Present Research

One limitation of the present research is that our analyses relied on only one implicit measure: the IAT. As such, the present research is not poised to answer questions regarding implicit measures in general. Moreover, previous research suggests that different implicit measures often do not correlate strongly and, thus, might assess different constructs (e.g., Degner & Wentura, 2009, 2010). Thus, it is unclear whether and to what extent the magnitude and direction of intergroup evaluations among members of disadvantaged groups might depend on the types of implicit measures. Consequently, this limitation highlights the need for a meta-

analysis synthesizing research on intergroup evaluations among disadvantaged groups across different implicit measures, which would allow for generalizations beyond one operationalization of the construct (see Essien & Degner, 2020).

Another limitation of the present research is that it relies solely on data from Project Implicit. Participants visit the demonstration website voluntarily and, consequently, are neither random nor representative samples of the general population, or any specific population at all. That said, this limitation is, in part, offset by the size and diversity of the Project Implicit samples, relative to what could reasonably be expected from samples of university undergraduates typically employed in psychological research (Henrich, Heine, & Norenzayan, 2010; Sears, 1986). Consequently, the Project Implicit datasets (Xu et al., 2017, 2018) used in the present research help to advance our understanding of SJT, and intergroup relations more generally, by providing very good statistical power and internal validity to examine a wide variety of social identities, some of which may be rare or otherwise hard to sample.

Concluding Remarks

Using large samples of online participants, the present research examined hypotheses proposed by SJT (Jost et al., 2004) regarding intergroup evaluation in disadvantaged groups. The present research advances the literature by testing these hypotheses at both the individual and social group levels. Across nations and social identities, we found that disadvantaged groups generally displayed outgroup favoritism on the implicit measure, but ingroup favoritism or no intergroup preference on explicit measures. Exploratory analyses revealed that intergroup evaluation in disadvantaged groups was moderated by the intergroup domain. Furthermore, implicit and explicit intergroup evaluations were moderated by political ideology and

stigmatization. Taken together, these findings generally support SJT, but at the same time highlight the need for more theory-driven research into the boundary conditions of SJT.

CHAPTER 3: GROUP EVALUATIONS AS SELF-GROUP DISTANCING: INGROUP TYPICALITY MODERATES EVALUATIVE INTERGROUP BIAS IN STIGMATIZED GROUPS[21]

Abstract

Outgroup favoritism among members of stigmatized can be seen as a form of self-group distancing. We examined how intergroup evaluations in stigmatized groups vary as a function of ingroup typicality. In Study 1 and 2, Black participants (N = 125,915; N = 766) more strongly preferred light-skinned or White relative to dark-skinned or Black individuals the lighter their own skin tone. In Study 3, overweight participants (N = 147,540) more strongly preferred normal-weight relative to overweight individuals the lower their own body weight. In Study 4, participants with disabilities (N = 35,058) more strongly preferred non-disabled relative to disabled individuals the less visible they judged their own disability. Relationships between ingroup typicality and intergroup evaluations were at least partially mediated by ingroup identification (Study 2 and 3). A meta-analysis across studies yielded an average effect size of r = .12. Furthermore, higher ingroup typicality was related to both ingroup and outgroup evaluations. We discuss ingroup typicality as an individual constraint to self-group distancing among stigmatized group members and its relation to intergroup evaluations.

---

[21] This chapter is based on Essien, I., Otten, S., & Degner, J. (in press). Group evaluations as self-group distancing: Ingroup typicality moderates evaluative intergroup bias in stigmatized groups. *European Journal of Social Psychology*.

Introduction

Picture Tyree and Jamal, who both view themselves as Black Americans: Whereas Jamal is very dark-skinned, Tyree's complexion is very light. This difference is also reflected in their daily life experiences: Most people agree that Jamal is Black; fewer people are that confident when judging Tyree. May these different perceptions and experiences influence how these two individuals generally think and feel about Black Americans relative to White Americans? The present research explores how categorizing oneself as a member of a stigmatized group, yet appearing more or less typical for or similar to that group, may shape how we feel about our ingroup. More specifically, we argue that ingroup and outgroup evaluations can reflect a tendency to distance the self from a stigmatized identity, and that this tendency is constrained by the extent to which an individual appears more or less typical for the ingroup.

Self-Group Distancing among Members of Stigmatized Groups

In many societies, members of stigmatized groups—social groups that are ascribed comparatively less prestige than others (e.g., Brown-Iannuzzi, Lundberg, Kay, & Payne, 2015; Tajfel & Turner, 1979)—frequently experience discrimination and threats to their social identity (e.g., Schmitt, Branscombe, Postmes, & Garcia, 2014). However, individuals differ in how they cope with their ingroup's disadvantage. Experiences of disadvantage prompt some individuals to pursue group-level strategies aimed at improving the status of the ingroup (e.g., collective action; Tajfel & Turner, 1979); yet others pursue individual-level strategies aimed at improving their personal situation (see de Lemus & Stroebe, 2015). Individual-level and group-level strategies to cope with group disadvantage are often incompatible with one another. *Self-group distancing*

represents one such individual-level strategy that comes at the expense of group-level outcomes. Broadly, self-group distancing describes strategic behaviors displayed by members of stigmatized groups, who sacrifice group goals in order to pursue individual goals (e.g., Derks, van Laar, & Ellemers, 2016). For example, individuals might distance themselves by perceiving or emphasizing their dissimilarities with the ingroup (e.g., Weiss & Lang, 2012), endorsing negative stereotypes about the ingroup (e.g., Derks, Ellemers, van Laar, & Groot, 2011), describing themselves more in terms of (positive) outgroup characteristics (e.g., Derks, Ellemers et al., 2011; Derks, van Laar, Ellemers, & Raghoe, 2015), psychologically distancing themselves from the ingroup (e.g., Derks, van Laar, Ellemers, & Groot, 2011), or evaluating the stigmatized ingroup negatively (e.g., Guimond, Dif, & Aupy, 2002). The current research focuses on relative preferences for an outgroup over the ingroup, which we argue also represents a way of distancing the self from a stigmatized social identity.

Scholars attribute self-group distancing among members of stigmatized groups to the biased and identity threatening social contexts they have to navigate in their daily lives (see Derks et al., 2016). According to this view, self-group distancing is a consequence of stigmatized group members' effort to assimilate to outgroup contexts, in which they are frequently exposed to threats to their social identity. These outgroup contexts are shaped by dominant groups, and stigmatized group members thus likely encounter negative ingroup stereotypes as well as positive outgroup stereotypes, decreasing their willingness to be categorized according to their (stigmatized) group membership (Branscombe, Ellemers, Spears, & Doosje, 1999). Consequently, distancing the self from negative ingroup stereotypes and/or associating the self with positive outgroup stereotypes seems to provide a way for stigmatized group members to cope with social identity threats, and to personally thrive in social contexts which are biased

against them. Lastly, not all members of stigmatized groups react to identity threatening contexts in a similar fashion. Instead, self-group distancing seems to be more likely among individuals who are less identified with their stigmatized ingroup (Derks et al., 2016). Thus, the extent to which individuals identify with their stigmatized ingroup seems to provide an important antecedent to self-group distancing.

So far, most research has investigated self-group distancing in organizational and work settings, and specifically among women in leadership roles. For example, low gender identified senior policewomen described themselves in more masculine terms when they were reminded of experiences in which they were stereotyped (Derks, van Laar et al., 2011). Similarly, low gender identified senior women leaders in another study who reported having experienced more gender discrimination characterized themselves with more masculine traits (Derks, Ellemers et al., 2011). Lastly, self-group distancing has also been observed in other social identity domains. For example, older adults were more likely to distance themselves from their age group when they were exposed to negative age stereotypes (Weiss & Freund, 2012). Other research suggests that gay men might distance themselves from negative ingroup stereotypes by displaying stereotypically male behaviors (Bishop, Kiss, Morrison, Rushe, & Specht, 2014; Clarkson, 2006; Eguchi, 2009).

Taken together, self-group distancing represents an individual-level strategy aimed at individual mobility, often at the expense of group interests; is thought to reflect stigmatized group members' responses to social identity threats; and is more likely observed among individuals who are less identified with their ingroup. Based on these insights, we argue that ingroup and/or outgroup evaluations among members of stigmatized groups may reflect self-group distancing.

Group Evaluations as a Manifestation of Self-Group Distancing

Ingroup favoritism is a robust phenomenon, influencing feelings, beliefs, and behaviors (e.g., Greenwald & Pettigrew, 2014; Hewstone, Rubin, & Willis, 2002; Mullen, Brown, & Smith, 1992). Theorizing suggests that identifying with and favoring one's ingroup affects well-being (e.g., self-esteem; Tajfel & Turner, 1979), and might also provide a buffer against social rejection (Branscombe, Schmitt, & Harvey, 1999). The occurrence of ingroup favoritism, however, also depends on a groups' position within status and power hierarchies in society. Specifically, members of stigmatized groups do not always display ingroup favoritism.

Whereas stigmatized group members often self-report similar degrees of ingroup liking as non-stigmatized group members (e.g., Mullen et al., 1992), they sometimes display evaluative outgroup favoritism on implicit measures (e.g., Nosek, Banaji, & Greenwald, 2002).[22] In our view, outgroup favoritism on evaluative measures reflects, at least to some degree, self-group distancing (see also Derks et al., 2016). Self-group distancing among members of stigmatized groups is often a response to stereotyping in biased contexts (e.g., Derks, Ellemers et al., 2011; Derks, van Laar et al., 2011; Weiss & Freund, 2012). Because the valence of stereotypes is causally related to the valence of group evaluations (e.g., Phills, Hahn, & Gawronski, 2020),

---

[22] This paper uses the term "implicit" to refer to indirect measurement procedures and their outcomes. However, this "implicit-as-indirect" conceptualization (Corneille & Hütter, 2020, p. 1) does not equate implicit and explicit measurement outcomes with different mental representations or features of automaticity.

stigmatized group members' exposure to stereotypes is likely to have downstream consequences for group evaluations. Consequently, outgroup favoritism may directly follow from the activation or endorsement of negative ingroup stereotypes and/or positive outgroup stereotypes. This rationale is further supported by findings that outgroup favoritism is more pronounced in groups to the extent that they are stigmatized by others (i.e., negatively evaluated; Essien, Calanchini, & Degner, in press; Rudman, Feinberg, & Fairchild, 2002). Taken together, we view outgroup favoritism among members of stigmatized groups as one manifestation of self-group distancing, because such group evaluations might at least in part reflect the activation or endorsement of negative ingroup and/or positive outgroup stereotypes.

Findings regarding stigmatized group members' intergroup evaluations on implicit measures vary greatly, and different studies have documented ingroup favoritism, outgroup favoritism, or no group preferences (Axt, Ebersole, & Nosek, 2014; Degner, Essien, & Reichardt, 2016; Essien et al., 2020; Rae, Newheiser, & Olson, 2015; Rudman et al., 2002). Thus far, these variations in intergroup evaluations have not been exhaustively explained. Here, we suggest that one moderator of group evaluations may be the degree of similarity individuals perceive between themselves, their ingroup, and a non-stigmatized outgroup. More specifically, we assume that group members who *appear* as less phenotypically prototypical for the stigmatized ingroup and thus more similar to a non-stigmatized outgroup may be more likely to display outgroup favoritism as a form of self-group distancing.

Phenotypic Prototypicality as (Dis)Similarity with Group Prototypes

Phenotypic prototypicality refers to the degree to which individuals' appearances are perceived as similar to a group prototype (Davies, Hutchinson, Osborne, & Eberhardt, 2016). For example, Blacks with darker skin tone or with more Afrocentric facial features (e.g., broader nose, fuller lips) are more readily perceived as prototypically Black. Such phenotypic racial prototypicality has been linked to many real-life outcomes (Maddox, 2004). For example, Black individuals with darker skin tone were evaluated more negatively than those with lighter skin tone on both, implicit and explicit measures (e.g., Hagiwara, Kashy, & Cesario, 2012). Furthermore, Black individuals who were perceived as more prototypical were more likely rejected by non-Black outgroup members (Hebl, Williams, Sundermann, Kell, & Davies, 2012), more strongly associated with attributes stereotypically linked with Blacks (e.g., Blair, Judd, Sadler, & Jenkins, 2002), and more likely perceived as threatening than those who were perceived as less prototypical (e.g., Dixon & Maddox, 2005; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Kahn & Davies, 2011; Ma & Correll, 2011). Such effects influence outcomes in other important areas of everyday life, such as the educational system: Compared with individuals with lighter skin tone, dark-skinned individuals were judged as less competent, and these judgments were associated with lower educational expectations (Meeus, Mayor, González, Brown, & Manzi, 2017). Together, these studies suggest that higher phenotypic prototypicality among members of stigmatized groups is related to more negative experiences in outgroup contexts.

These findings regarding prototypicality also suggest that phenotypic appearance, in addition to group membership per se, shapes interaction experiences. We propose that,

eventually, these experiences may relate to stigmatized group members' own perceptions and attitudes about both ingroup and outgroups. In other words, we suggest that to the extent that ingroup typicality reflects phenotypic appearance, it might influence stigmatized group members' ability to (psychologically) distance themselves from or connect with their ingroup and outgroups. However, most research has investigated how stigmatized group members are perceived, judged, and treated by *others*, and only few studies have actually looked at how phenotypic prototypicality may relate to their *own* perceptions and group evaluations.

Perceived ingroup prototypicality may be associated with a number of psychological processes that construe the self as proximal to or distant from the ingroup and, in turn, may relate to differences in group evaluations. For example, less prototypical Black individuals displayed less ingroup identification (Harvey, LaBeach, Pridgen, & Gocial, 2005). Similarly, Black participants with lighter skin tone reported less closeness to Black people than those with darker skin tone (Brown, Ward, Lightbourn, & Jackson, 1999). Lastly, less prototypical Black and Latino individuals were less identified with their racial ingroups than more prototypical individuals (Wilkins, Kaiser, & Rieck, 2010). In sum, to the extent that members of stigmatized groups are perceived as more prototypical they seem to be more identified with their ingroup. Crucially, these relationships with identification were not only observed for self-reported prototypicality (Harvey et al., 2005), but also for other-rated prototypicality (Brown et al., 1999; Wilkins et al., 2010). Thus, relationships between ingroup prototypicality and identification may at least in part reflect actual phenotypic differences in appearance between members of stigmatized groups. In our view, this suggests that ingroup prototypicality may constrain stigmatized group members' ability to identify with the ingroup. Because ingroup identification is

related to ingroup liking and satisfaction with the ingroup (Leach et al., 2008), low ingroup prototypicality may also be related to lower levels of ingroup favoritism.

Furthermore, perceived ingroup typicality may be associated with stigmatized group members' perception of group boundaries: Stigmatized group members who appear less phenotypically prototypical may perceive group boundaries as more permeable, thus perceiving a higher liberty to distance the self from the ingroup. Social Identity Theory (Tajfel & Turner, 1979) assumes that the perception of permeability of group boundaries influences whether group members who are unsatisfied with the lower status of their ingroup attempt to leave the group, a strategy termed *individual mobility*. According to social identity theory, individual mobility might not only manifest itself in the sense of physically leaving a group, but also in the sense of psychologically distancing the self from the group, that is by disidentifying from the former ingroup (Tajfel & Turner, 1979). Hence, we argue that ingroup typicality might pose an individual constraint to the permeability of group boundaries, affecting stigmatized group members' tendency to distance themselves from the ingroup.

Lastly, weak group boundaries may decrease ingroup identification (cf. Reimer et al., 2016), and increase perspective-taking with outgroup members (e.g., Todd & Burgmer, 2013), thereby decreasing positive ingroup evaluations and enhancing positive outgroup evaluations (e.g., Rae et al., 2015). Taken together, we suggest that phenotypic appearance should be related to the extent to which members of stigmatized groups perceive themselves as typical for the ingroup. We further suggest that ingroup typicality and the resulting sense of (dis)similarity and (dis)connection with the ingroup may be related to ingroup and outgroup favoritism in stigmatized group members' intergroup evaluations on implicit and explicit measures.

The Present Research

We present four studies in which we investigated effects of ingroup typicality on evaluative ingroup favoritism on implicit and explicit measures. We analyzed data from two sources: Project Implicit and the American National Election Studies (ANES). Project Implicit is a website where visitors can complete various Implicit Association Tests (IAT; Greenwald, McGhee, & Schwartz, 1998) as well as explicit measures. For many target domains, data have been collected for more than a decade, yielding massive, openly available datasets (https://osf.io/y9hiq/; see Xu et al., 2017). ANES provides survey data from representative probability samples of adult U.S. citizens, which focus on political behavior and also include measures of intergroup attitudes (see http://www.electionstudies.org). We reviewed these sources for datasets, in which (a) participants self-categorized as members of a stigmatized social identity and that (b) included at least one variable differentiating between different levels of phenotypic prototypicality as a proxy for ingroup typicality. Three Project Implicit datasets and two ANES waves (2012 and 2016) fulfilled these criteria, investigating intergroup evaluations regarding skin tone and racial attitudes (Study 1 and 2), weight status (Study 3), and disability (Study 4). Different variables were available as indirect indicators of ingroup typicality. In Studies 1 and 2, we assumed that Black participants with lighter skin tone have lower perceived/phenotypic typicality for Blacks (or higher similarity to Whites). We investigated effects of self-reported skin tone (Study 1) and other-observed skin tone (Study 2). In Study 3, we assumed that overweight participants have lower perceived/phenotypic typicality for the overweight category (or higher similarity to the normal-weight category) the lower their self-reported weight status and the lower they believed others judge their weight status. Finally, in Study 4, we assumed that disabled participants have lower perceived/phenotypic typicality for the category disabled and higher

similarity to non-disabled people the more they were able to hide their disability; the less they felt affected by their disability in daily life; and the less severe they judged their disability. In all four studies, we examined how these variables were related to group evaluations. In addition, we explored in Study 2 and 3 whether ingroup typicality was related to ingroup identification. Lastly, we conducted a series of meta-analyses across studies, which examined (a) the overall magnitude of the relationship between ingroup typicality and intergroup evaluations and (b) whether ingroup typicality was related to ingroup and/or outgroup evaluations. All analyses were done using R.[23] Analyses scripts are accessible via Open Science Framework (https://osf.io/z4xwx/).

Study 1

In Study 1, we explored how Black participants' evaluations of light-skinned and dark-skinned people varied depending on their own skin tone perception.

Method

Participants.

[23] R (Version 3.6.2; R Core Team, 2017) and the R-packages corx (Version 1.0.2; Conigrave, 2019), data.table (Version 1.12.8; Dowle & Srinivasan, 2017), here (Version 0.1; Müller, 2017), irr (Version 0.84.1; Gamer, Lemon, Fellows, & Singh, 2019), jmv (Version 1.0.8; Selker, Love, & Dropmann, 2018), knitr (Version 1.28; Xie, 2015), MBESS (Version 4.6.0; Kelley, 2018), metafor (Version 2.1.0; Viechtbauer, 2010), papaja (Version 0.1.0.9942; Aust & Barth, 2018), and tidyverse (Version 1.3.0; Wickham, 2017).

Participants were visitors of the Project Implicit demonstration website (https://implicit.harvard.edu) between 2004 and 2015 who self-categorized as Black or African American (N = 125,915). Given such large sample size, achieved power for small effects (ρ = .1) was $1 - \beta$ = 1.0 (two-tailed).

Measures.

Self-reported skin tone. Black participants' self-reported skin tone was measured on scale ranging from 1 (*very dark*) to (*very light*)—see Table 2 for all response options and descriptive statistics.

Skin tone IAT. In the skin tone IAT, attribute stimuli were positive and negative words that had to be categorized as *good* versus *bad*. The target stimuli were dark-skinned and light-skinned male and female faces—see https://osf.io/y9hiq/, for a complete overview of stimuli and procedures.

Self-reported preference. Participants indicated their relative preference for dark-skinned over light-skinned people on a scale from 1 (*strong preference for dark-skinned people*) to 7 (*strong preference for light-skinned people*), with the midpoint indicating no preference.

Feeling thermometers. Participants indicated their feelings regarding light-skinned and dark-skinned people using two scales from 1 (*very cold*) to 11 (*very warm*).

Table 1

Mean IAT D Scores, self-reported preference scores, feeling thermometer difference scores, and their standard deviations as a function of Black participants' self-reported skin tone in Study 1.

	IAT			Preference			Thermometer		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
All Black participants	0.09	0.45	125915	3.79	1.08	115738	-0.37	1.86	121597
I consider my skin to be									
(1) very dark	0.05	0.46	3834	3.46	1.33	3536	-1.13	2.58	3749
(2) dark	0.04	0.45	24534	3.57	1.13	22853	-0.83	1.94	24085
(3) somewhat dark	0.06	0.45	21910	3.71	1.06	20497	-0.59	1.84	21530
(4) medium	0.09	0.45	42636	3.86	1.02	39623	-0.27	1.71	41838
(5) somewhat light	0.14	0.45	14785	4.04	1.01	13953	0.09	1.68	14564
(6) light	0.19	0.46	11191	4.11	1.07	10523	0.26	1.72	11003
(7) very light	0.21	0.48	1843	3.98	1.17	1742	0.14	2.13	1811
White participants	0.40	0.40	465925	4.27	0.95	434900	0.54	1.72	450702

Note. IAT = IAT *D* Score; Preference = self-reported preference score; Thermometer = feeling thermometer difference score. Higher means on IATs and self-report measures indicate a stronger preference for light-skinned over dark-skinned individuals. Variable sample sizes are based on missing values in the different dependent variables. As means of comparison, we also report average scores for the non-stigmatized group (i.e., White participants).

Results

We used IAT *D* Scores and self-reported preference scores as reported in the dataset. In addition, we computed feeling thermometer difference scores by subtracting participants' evaluations of dark-skinned people from evaluations regarding light-skinned people. In all measures, more positive scores indicate a relative preference for light-skinned people over dark-

skinned people. Table 1 presents descriptive statistics for measures of intergroup evaluations for the different levels of self-reported skin tone. As means of comparison, we also report average scores for the non-stigmatized group (e.g., White participants).

Overall, Black participants had a positive IAT *D* Score, which significantly differed from zero $t(125{,}914) = 69.69$, $p < .001$, $d_z = 0.20$, 95% *CI* [0.19; 0.20], indicating an overall small preference of light-skinned relative to dark-skinned people on the IAT. We also tested participants' mean self-reported preference scores against the scale midpoint, $t(115{,}737) = -66.77$, $p < .001$, $d_z = -0.20$, 95% *CI* [-0.20; -0.19], and their thermometer difference scores against zero, $t(121{,}596) = -69.65$, $p < .001$, $d_z = -0.20$, 95% *CI* [-0.21; -0.19], indicating a preference for dark-skinned relative to light-skinned people on both self-report measures.

Table 2 reports correlations between self-reported skin tone and group evaluations. Importantly, we observed a small positive correlations between participants' self-reported skin tone and their IAT *D* Scores, $r(120731) = .09$, $p < .001$, 95% *CI* [0.09; 0.10], their self-reported preference scores, $r(112725) = .16$, $p < .001$, 95% *CI* [.16; .17], and their thermometer difference scores, $r(118578) = .20$, $p < .001$, 95% *CI* [.19; .20].

Table 2

Zero-order correlations of Black participants' self-reported skin tone, IAT D Scores, self-reported preference scores, and individual feeling thermometers in Study 1.

	1	2	3	4	5	*M*	*SD*
1. Typicality	-					3.67	1.34
2. IAT	.09***	-				0.09	0.45
3. Preference	.16***	.11***	-			3.79	1.08
4. Thermometer	.20***	.10***	.56***	-		-0.37	1.86
5. Ingroup	.13***	.08***	.25***	.41***	-	2.50	2.29
6. Outgroup	-.03***	.00	-.20***	-.41***	.67***	2.87	2.29

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; Typicality = measure of ingroup typicality (i.e., self-reported skin tone); IAT = IAT *D* Score; Preference = self-reported preference score; Ingroup = feeling thermometer regarding dark-skinned people; Outgroup = feeling thermometer regarding dark-skinned people; Thermometer = feeling thermometer difference score. Higher means on IATs and self-report measures indicate a stronger preference for light-skinned over dark-skinned individuals.

Discussion

Black participants showed stronger preferences for light-skinned relative to dark-skinned individuals on the IAT and two self-report measures to the extent that they self-reported lighter skin tone. This provides initial support for the assumption that differences in ingroup typicality explain variance in intergroup evaluations. However, a reversed interpretation remains plausible:

Participants may perceive their skin tone to be lighter *because* of their stronger relative preference for light-skinned relative to dark-skinned people and a relative disconnectedness with the ingroup. According to this reasoning, effects might be due to participants "adjusting" perceptions of their skin tone as a consequence of their attitudes, due to lower identification, or generally as a strategy to achieve cognitive consistency between their attitudes and their self-observations (cf. Gawronski, Brochu, Sritharan, & Strack, 2012). From this viewpoint, two people with the same skin tone may report very different levels of subjective skin tone, based on different levels of ingroup identification. Study 2 addresses this possible alternative interpretation.

Study 2

Study 2—a pre-registered analysis of ANES 2012 and 2016 data—aimed at replicating and extending findings of the previous study. Again, we used skin tone as proxy for perceived ingroup typicality in Black participants. Different from the previous study, the ANES datasets provided skin tone categorizations also as other-observations: Interviewers who conducted face-to-face interviews also reported participants' perceived skin tone. However, only self-report measures of intergroup evaluations were assessed (i.e., feeling thermometer scales).

Based on the results of Study 1, we expected Black participants to display higher levels of ingroup favoritism the darker their interviewer-assessed and self-reported skin tone. The pre-registration, materials, and analysis script are accessible via https://osf.io/kn7qv/.

In addition to the pre-registered analyses, we explored whether ingroup typicality was related to group identification among members of stigmatized groups. The results of these

exploratory analyses may reveal whether lower ingroup typicality involves psychological distancing from a stigmatized identity (i.e., disidentification). Together, these analyses may point towards involved psychological processes underlying the correlation between skin tone measures and intergroup evaluations.

Method

Participants. Participants were 766 respondents (291 male, 472 female, 3 unknown; MD_{age} = 43, *SD* = 15.99) from the ANES 2012 (*n* = 413) and ANES 2016 (*n* = 353) Time Series, who self-identified as African American or Black. All ANES 2012 respondents participated in face-to-face interviews. For ANES 2016, 119 respondents participated in face-to-face interviews and 234 respondents participated online. Power analysis was based on the lower boundary of the confidence interval for the correlation between self-reported skin tone and self-reported preference scores observed in Study 1. Given α = .05, and $1 - \beta$ = .95, a sample size of at least *N* = 425 was needed to detect an effect of ρ = .158 (one-tailed; see pre-registration at https://osf.io/95q4v/).

Measures.

Skin tone assessment. Skin tone was assessed using a skin color scale originally designed by Massey and Martin (2003), a 10-point graphical scale depicting a human hand in ten different shades (1 = very light; 10 = very dark). We recoded values in parallel to Study 1 such that higher values indicate lower ingroup typicality. Respondents' skin tone was recorded during two interviews; once at the end of the pre-election interview and once at the end of the post-election interview, respectively. Based on satisfying intra-class correlations of r_{ICC} = .85, 95% *CI* [0.82,

0.87], between interviewers' skin tone observations, we averaged the two skin tone observations. In ANES 2016, respondents additionally self-reported their skin tone using the same scale.[24]

Self-reported preference. In both studies, feeling thermometers were administered as part of the post-election data collection via computer-aided self-interviews. Evaluations of Blacks and Whites were assessed separately, using scales from 0 (*unfavorable/cold*) to 100 (*favorable/warm*). We computed a feeling thermometer difference score in parallel to Study 1 by subtracting evaluations regarding Blacks from evaluations regarding Whites. Positive values indicate more positive evaluations of Whites relative to Blacks.

Ingroup identification. Participants were asked how important being Black was to their identity on a scale ranging from 1 (*not at all important*) to 5 (*extremely important*).

Results

Exploratory analyses of the main effect of ingroup favoritism revealed that Black participants had a negative feeling thermometer difference score (M = -16.26, SD = 24.02), significantly different from zero $t(707) = -18.01$, $p < .001$, $d_z = -0.68$, 95% CI [-0.76; -0.59], thus replicating the effect of self-reported ingroup favoritism observed in Study 1.

---

[24] Note that in the ANES 2016 study, face-to-face respondents self-reported their skin tone in addition to interviewer-assessed skin tone, whereas online skin tone was only assessed via self-report.

Table 3

Zero-order correlations of Black participants' interviewer-assessed skin tone, self-reported skin tone, feeling thermometer difference scores, and individual feeling thermometers in Study 2.

	1	2	3	4	5	*M*	*SD*
1. Skin Tone (Other)	-					5.52	2.24
2. Skin Tone (Self)	.40***	-				5.48	1.96
3. Identification	.12**	.12*	-			4.29	1.07
4. Thermometer	.11*	-.06	.19***	-		16.26	24.02
5. Ingroup	-.01	-.10	.30***	.42***	-	85.32	19.12
6. Outgroup	-.12*	-.02	.05	-.68***	.38***	69.00	23.49

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; Skin Tone (Other) = interviewer-assessed skin tone; Skin Tone (Self) = self-reported skin tone; Identification = ingroup identification; Thermometer = feeling thermometer difference score; Ingroup = ingroup feeling thermometer; Outgroup = outgroup feeling thermometer. Higher means for skin tone measures indicate darker (interviewer-assessed and self-reported) skin tone; a higher feeling thermometer difference score indicates a stronger preference for Blacks relative to Whites; higher means for individual feeling thermometers indicate more favorable/warm evaluations of the respective target group.

Table 3 reports correlations between skin tone ratings and group evaluations. As predicted, we observed a small positive correlation between face-to-face respondents' feeling thermometer difference scores and their interviewer-assessed skin tone, $r(474) = .11$, $p = .006$, 95% *CI* [.04; 1.00]: Black participants showed higher levels of ingroup favoritism on feeling

thermometer difference scores the darker their observed skin tone. Surprisingly however, participants' self-reported skin tone was not positively correlated with feeling thermometer difference scores, $r(329) = -.06$, $p = .854$, 95% *CI* [-.15; 1.00]. In order to explore why the predicted effect did not replicate using respondents' self-reported skin tone ratings, we looked at the agreement between respondents' self-reported skin tone ratings and average interviewer-assessed skin tone ratings. Agreement was surprisingly low, $r_{ICC} = .57$, 95% *CI* [.36, .71].

Exploratory analyses. Next, we explored relationships between skin tone ratings and ingroup identification. We observed small positive correlations between participants' interviewer-observed skin tone and their levels of ingroup identification, $r(487) = .12$, $p = .010$, 95% *CI* [.03; 0.20]. This indicates that participants with darker observed skin tone displayed higher levels of ingroup identification than did participants with lighter skin tone.

We then conducted a mediation analysis using the "medmod" package[25] with 1,000 bootstrap resamples. This analysis used participants' feeling thermometer difference scores as dependent variable, interviewer-observed skin tone as a predictor, and ingroup identification as a mediator. Indeed, the indirect effect was significant, $b = -0.25$, $SE = 0.11$, p .023, 95% *CI* [-0.50, -0.07]. This indicates that the effect of skin tone on ingroup favoritism was mediated by participants' level of ingroup identification.

[25] The medmod package uses the "lavaan" package for computations.

Regarding participants' self-reported skin tone, we observed a small positive correlation with their levels of ingroup identification, $r(329) = .12$, $p = .034$, 95% *CI* [.01; 0.22]. This indicates that participants with darker skin tone reported higher levels of identification than those with lighter skin tone.

Discussion

Study 2 indicates that Black participants displayed higher levels of outgroup favoritism on a self-report measure the lighter their skin tone. The replication of effects from Study 1 with the observer skin tone ratings in Study 2 strengthens our interpretation that lower ingroup typicality leads to less ingroup favoritism. It is less likely that skin tone perceptions were systematically biased by participants' intergroup attitudes—skin tone was not self-reported, but recorded by the interviewers. That said, it is also possible that skin tone observations were influenced by participants responses during the interview, because skin tone observations were made at the end of the interview. On the other hand, it is also important to note that feeling thermometers and other sensitive information were assessed through computer-aided self-interviews without the interviewers' participation. Thus, it remains an open question whether or to what extent skin tone ratings might have been influenced by participants' behavior.

Contrary to expectations, the correlation of self-reported skin tone with ingroup favoritism observed in Study 1 did not replicate in Study 2. Explanations for this null finding might center on differences in measurement procedures between observed and self-reported skin tone. For example, observer skin tone ratings were assessed in face-to-face interviews whereas self-reported skin tone ratings were assessed during a self-administered online survey. Furthermore, observer skin tone ratings were assessed twice whereas self-reported skin tone was only assessed

once. Consequently, measurement error of self-reported skin tone measures may have been higher compared to observed skin tone measures, which would have differentially attenuated correlations between skin tone measures and ingroup favoritism (e.g., Hunter & Schmidt, 2007).

Lastly, exploratory analyses indicate that the correlations between interviewer-assessed skin tone on ingroup favoritism were partially mediated by participants' level of ingroup identification. Findings of Study 2 thus provide preliminary evidence that lower ingroup typicality might be associated with psychological distancing from a stigmatized identity. Together, results of Study 1 and 2 suggest that how Black Americans evaluate the ingroup and outgroup is related to the degree to which they appear typical of their ingroup in terms of skin tone (i.e., ingroup typicality). In Studies 3 and 4 we explored whether these effects (a) extent to other social categories and (b) are observed when using different operationalizations of ingroup typicality. A replication of these effects would suggest the operation of comparable basic processes mediating the relationship between ingroup typicality and intergroup evaluations.

Study 3

Study 3 investigated the link between ingroup typicality and intergroup bias regarding a different social categorization: Weight status. Weight is an important factor in interpersonal perception, with overweight individuals being frequently negatively stigmatized (e.g., Crandall, 1994). However, although negative evaluations of overweight individuals are widely shared within many Western societies, own body weight has been shown to be related to weight-related automatic prejudice, with overweight individuals displaying less anti-fat bias and even ingroup favoritism the higher their actual body weight (e.g., Degner & Wentura, 2009; Schwartz,

Vartanian, Nosek, & Brownell, 2006). Hence, Study 3 examined whether weight status categorization is related to intergroup evaluations on both the IAT and self-report measures. Furthermore, to further our understanding of the relationship between ingroup typicality and self-group distancing, we again explored relationships between ingroup typicality, group identification, and ingroup favoritism.

Method

Participants. Participants were visitors of the Project Implicit demonstration website between 2004 and 2015, who self-categorized as overweight (N = 147,540). Given such large sample size, achieved power for small effects (ρ = .1) was $1 - \beta$ = 1.0 (two-tailed).

Measures.

Weight IAT. The weight IAT followed the same procedure as the skin tone IAT in Study 1 with the exception that target stimuli were images of normal weight and overweight individuals (e.g., faces or body shapes)—see https://osf.io/y9hiq/, for a complete overview of experimental stimuli and procedures.

Self-reported preference. Participants indicated their relative preference for overweight over normal weight individuals on scales from 1 (*strong preference for overweight individuals*) to 7 (*strong preference for normal weight individuals*), with the midpoint indicating no preference. Participants indicated their feelings regarding overweight and normal weight individuals on scales from 1 (*very cold*) to 11 (*very warm*).

Perceived weight status.

Participants' reported their perceived weight status and reported how they thought others would judge their weight on a scale ranging from 1 (*very underweight*) to 7 (*very overweight*). For comparability with the previous studies, weight status was recoded such that higher values indicate lower weight. We used both variables as separate proxies for participants' ingroup typicality—see Table 4 for all response options and descriptive statistics.[26]

Ingroup identification. Participants were asked how much they identified with overweight people using a scale from 1 (*Not at all*) to 5 (*Strongly*). Exact wording was „How much do you identify with people who are fat?".[27]

[26] Note that we excluded overweight participants who reported that others would judge them as slightly, moderately, or very underweight (n = 2433).

[27] In our view, the phrasing "fat" does not appear to be a neutral way of addressing people who are heavyweight. However, we do not have clear hypotheses about how this wording might have influenced responses toward this item.

Table 4

Mean IAT D Scores, self-reported preference scores, feeling thermometer difference scores, and their standard deviations as a function of participants' self-reported weight status and their report of how others would judge their weight status in Study 3.

	IAT			Preference			Thermometer		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Overweight participants	0.37	0.43	147540	4.73	1.08	143116	0.47	2.32	146555
Currently, I am									
(1) very overweight	0.24	0.45	19276	4.40	1.19	18656	-0.16	2.56	19139
(2) moderately overweight	0.33	0.44	38576	4.56	1.08	37417	0.16	2.27	38342
(3) slightly overweight	0.41	0.42	89688	4.87	1.03	87043	0.74	2.24	89074
Other people would say that I am									
(1) very overweight	0.21	0.45	10498	4.30	1.20	10148	-0.30	2.57	10415
(2) moderately overweight	0.29	0.44	22766	4.49	1.08	21997	0.08	2.24	22645
(3) slightly overweight	0.38	0.43	44546	4.72	1.02	43146	0.53	2.20	44254
(4) normal weight	0.44	0.41	36219	4.99	1.03	35157	0.99	2.29	36022
Normal weight participants	0.46	0.40	170182	5.15	1.04	164764	1.36	2.37	168887

Note. IAT = IAT *D* Score; Preference = self-reported preference score; Thermometer = feeling thermometer difference score. Higher means on IATs and self-report measures indicate a stronger preference for normal weight over overweight individuals. Variable sample sizes are based on missing values in the different dependent variables. We excluded overweight participants who reported that others would judge them as slightly, moderately, or very underweight (n = 2,433). As means of comparison, we also report average scores for the non-stigmatized group (i.e., normal weight participants).

Results

Overall, overweight participants had a positive IAT *D* Score, which significantly differed from zero $t(147{,}539) = 328.59$, $p < .001$, $d_z = 0.86$, 95% *CI* [0.85; 0.86], indicating outgroup favoritism on the IAT (see Table 5). We tested participants' mean self-reported preference score against the scale midpoint, $t(143{,}115) = 254.99$, $p < .001$, $d_z = 0.67$, 95% *CI* [0.67; 0.68], and their thermometer difference score against zero, $t(146{,}554) = 77.50$, $p < .001$, $d_z = 0.20$, 95% *CI* [0.20; 0.21], indicating outgroup favoritism for both self-report measures.

Table 5 reports correlations between weight status, group evaluations, and ingroup identification. Crucially, we observed small correlations between IAT *D* Scores and participants' self-reported weight status, $r(147538) = .14$, $p < .001$, 95% *CI* [0.14; 0.15], and with their reports of how others would judge their weight status, $r(114027) = .16$, $p < .001$, 95% *CI* [0.16; 0.17]. We also observed small and small-to-medium correlations between self-reported preference scores and participants' self-reported weight status, $r(143114) = .17$, $p < .001$, 95% *CI* [.16; .17], and with their reports of how others would judge their weight status, $r(110446) = .21$, $p < .001$, 95% *CI* [.20; .21]. Lastly, we observed small correlations between thermometer difference scores and participants' self-reported weight status, $r(146553) = .15$, $p < .001$, 95% *CI* [.14; .15], and their reports of how others would judge their weight status, $r(113334) = .18$, $p < .001$, 95% *CI* [.17; .18].

Table 5

Zero-order correlations of overweight participants' self-reported weight status, reports of how others would judge their weight status, IAT D Scores, self-reported preference scores, feeling thermometer difference scores, and individual feeling thermometers in Study 3.

	1	2	3	4	5	6	7	*M*	*SD*
1. Weight (Self)	-							2.48	0.71
2. Weight (Others)	.75***	-						2.93	0.94
3. Identification	.38***	.38***	-					3.21	1.06
4. IAT	.14***	.16***	.14***	-				0.37	0.43
5. Preference	.17***	.21***	.32***	.22***	-			4.73	1.08
6. Thermometer	.15***	.18***	.31***	.18***	.59***	-		0.47	2.32
7. Ingroup	.10***	.12***	.30***	.16***	.41***	.62***	-	5.06	2.08
8. Outgroup	-.07***	-.09***	-.04***	-.05***	-.27***	-.53***	.34***	4.59	1.93

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; I am = self-reported weight status; Weight (Others) = reports of how others would judge participants' weight status; IAT = IAT *D* Score; Preference = self-reported preference score; Thermometer = feeling thermometer difference score; Ingroup = ingroup feeling thermometer; Outgroup = outgroup feeling thermometer. Higher means on IATs and self-report measures indicate a stronger preference for overweight relative to normal weight individuals.

Exploratory analyses.

Next, we explored relationships between the two indicators of ingroup typicality and ingroup identification. Participants self-reported weight status correlated positively with ingroup identification, $r(102575) = .38$, $p < .001$, 95% *CI* [.38; .39], indicating that participants identified more strongly with overweight people the higher their self-reported weight status. Furthermore, participants' ratings of how others would judge their weight status correlated positively with their level of ingroup identification, $r(87373) = .38$, $p < .001$, 95% *CI* [.37; .38], indicating that participants identified more strongly with overweight people the higher they rated that others would judge their weight status.

We then explored in a series of mediation analyses whether the correlation between typicality and outgroup favoritism was mediated by participants' level of ingroup identification. Both measures of typicality were highly correlated, $r(114027) = .75$, $p < .001$, 95% *CI* [.75; .76], so we calculated an ingroup typicality index by averaging the two items. First, we conducted a mediation analysis with participants' IAT *D* Scores as dependent variable, ingroup typicality as a predictor, and ingroup identification as a mediator. Indeed, the indirect effect was significant, $b = -0.02$, $SE = 0.00$, $p < .001$, 95% *CI* [-0.02, -0.02]. This indicates that the correlation between ingroup typicality and outgroup favoritism was partially mediated by participants' level of ingroup identification. Again, we conducted a mediation analysis, this time using participants' self-reported preference scores as dependent variable, ingroup typicality as a predictor, and ingroup identification as a mediator. The indirect effect was significant, $b = -0.16$, $SE = 0.00$, $p < .001$, 95% *CI* [-0.17, -0.16]. This indicates that the correlation between ingroup typicality and outgroup favoritism on self-report measures was partially mediated by participants' level of ingroup identification. Lastly, we conducted a mediation analysis, this time using participants' feeling thermometer difference scores as dependent variable, ingroup typicality as a predictor,

and ingroup identification as a mediator. The indirect effect was significant, $b = 0.34$, $SE = 0.01$, $p < .001$, 95% *CI* [0.33, 0.35]. This indicates that the correlation between ingroup typicality and outgroup favoritism on feeling thermometer was partially mediated by participants' level of ingroup identification.

Discussion

Results of Study 3 indicate that overweight participants displayed a stronger preference for normal weight relative to overweight individuals the lower their reported weight status and the lower their reports of how they believed others would judge their weight status—thus the less typical they appeared for the overweight category. Importantly, this effect was observed for intergroup evaluations on both the IAT and self-report measures. We also observed that both indicators of ingroup typicality were correlated with overweight participants' level of ingroup identification. Lastly, we observed that higher levels of outgroup favoritism were in part due to the fact that overweight participants who reported being less typical for their group were are also less likely to identify with that group.

Study 4

Study 4 focuses on yet another domain of intergroup perception—disability. While people might self-categorize as either disabled or abled, people who self-categorize as disabled might still perceive themselves as more or less conforming to a prototypical image of a disabled person, which might in turn influence ingroup and outgroup evaluations.

Method

Participants.

Participants were visitors of the Project Implicit demonstration website, who indicated that they had a disability (N = 35,058). Given such large sample size, achieved power for small effects (ρ = .1) was $1-\beta$ = 1.0 (two-tailed).

Measures.

Disability IAT. The Disability IAT followed the same procedures as IATs in Study 1 and 3 except that target stimuli were symbols indicating disability (e.g., crutches, guide dog, wheelchair) and ability (e.g., persons who walk, run, or ski)—see https://osf.io/y9hiq/, for a complete overview of experimental stimuli and procedures.

Self-reported preference. Participants indicated their relative preference for disabled over abled people on scales from 1 (*strong preference for disabled people*) to 7 (*strong preference for abled people*), with the midpoint indicating no preference.

Perceived ingroup typicality. Participants completed several measures that can be interpreted as proxies of perceived ingroup typicality or similarity to non-disabled people: the ability to hide their disability on a scale from 1 (*impossible to hide*) to 4 (*very able to hide*), how much their disability affected things they do in life on a scale from 1 (*nothing I do*) to 6 (*everything I do*), and the perceived severity of their disability on a scale from 1 (*very slight*) to 5 (*very severe*)—see Table 6 for all response options and descriptive statistics.

Table 6

Mean IAT D Scores, self-reported preference scores, feeling thermometer difference scores, and their standard deviations as a function of three measures of ingroup typicality in Study 4.

	IAT			Preference			Thermometer		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Disabled participants	0.37	0.43	147540	4.73	1.08	143116	0.47	2.32	146555
My disability is									
(1) impossible to hide	0.34	0.49	3687	4.20	1.21	3479	-0.24	2.48	3637
(2) slightly able to hide	0.40	0.48	5609	4.19	1.02	5299	-0.28	1.97	5529
(3) moderately able to hide	0.43	0.48	8783	4.20	0.99	8353	-0.25	1.88	8665
(4) very able to hide	0.43	0.46	13310	4.30	0.95	12726	-0.03	1.86	13174
My disability affects									
(1) everything I do	0.35	0.49	3307	4.07	1.25	3156	-0.52	2.72	3264
(2) very many things I do	0.35	0.49	3918	4.14	1.03	3716	-0.39	2.00	3864
(3) many things I do	0.41	0.47	9972	4.22	0.96	9491	-0.21	1.82	9865
(4) few things I do	0.44	0.46	8660	4.29	0.92	8240	-0.01	1.80	8548
(5) very few things I do	0.46	0.46	4725	4.36	0.99	4493	0.08	1.80	4661
(6) nothing I do	0.43	0.50	735	4.37	1.29	693	0.08	2.44	722
My disability is									
(1) very severe	0.35	0.52	1078	4.01	1.49	1022	-0.41	3.48	1060
(2) severe	0.38	0.48	5859	4.12	1.05	5565	-0.42	2.06	5773
(3) moderate	0.41	0.48	15110	4.22	0.96	14380	-0.19	1.85	14937
(4) slight	0.45	0.46	6598	4.35	0.96	6280	0.04	1.79	6522
(5) very slight	0.45	0.45	2658	4.40	0.99	2535	0.18	1.94	2624
Non-disabled participants	0.49	0.43	234676	4.42	0.97	227100	0.45	1.98	232983

Note. IAT = IAT *D* Score; Preference = self-reported preference score; Thermometer = feeling thermometer difference score. Higher means on IATs and self-report measures indicate a stronger preference for non-disabled over disabled individuals. Variable sample sizes are based on missing values in the different dependent variables. As means of comparison, we also report average scores for the non-stigmatized group (i.e., non-disabled participants).

Results

Overall, participants with disabilities had a positive IAT *D* Score, which significantly differed from zero $t(35{,}057) = 162.38$, $p < .001$, $d_z = 0.87$, 95% *CI* [0.85; 0.88], indicating outgroup favoritism on the IAT. Testing participants' mean self-reported preference score against the scale midpoint, $t(33{,}412) = 23.43$, $p < .001$, $d_z = 0.13$, 95% *CI* [0.12; 0.14], and their thermometer difference score against zero, $t(34{,}580) = -13.24$, $p < .001$, $d_z = -0.07$, 95% *CI* [-0.08; -0.06], indicated small effects of self-reported outgroup and ingroup favoritism, respectively.

Table 7 reports correlations between measures of ingroup typicality and group evaluations. Crucially, we observed correlations between IAT *D* Scores and participants' self-reported ability to hide their disability, $r(31387) = .06$, $p < .001$, 95% *CI* [0.05; 0.07], their judgments of how much the disability affects their lives, $r(31315) = .07$, $p < .001$, 95% *CI* [0.06; 0.08], and the perceived severity of their disability, $r(31301) = .05$, $p < .001$, 95% *CI* [0.04; 0.06], indicating very small correlations between ingroup typicality and outgroup favoritism on the IAT (see Table 6). Next, we correlated participants' self-reported preference and feeling thermometer difference scores with the different proxy variables. We observed similar correlations with the ability to hide the disability, $r(29855) = .04$, $p < .001$, 95% *CI* [.03; .05] and $r(31003) = .05$, $p < .001$, 95% *CI* [.04; .06], with the degree to which the disability affects their lives, $r(29787) = .09$, $p < .001$, 95% *CI* [.08; .10] and $r(30922) = .10$, $p < .001$, 95% *CI* [.08; .11], and with perceived severity, $r(29780) = .09$, $p < .001$, 95% *CI* [.08; .11] and $r(30914) = .09$, $p < .001$, 95% *CI* [.08; .10], respectively. This indicates that the correlations between measures of ingroup typicality and intergroup evaluations on self-report measures were very small.

Table 7

Zero-order correlations of disabled participants' self-reported measures of ingroup typicality (i.e., the ability to hide their disability; how much their disability affected things they do in life; and the perceived severity of their disability), IAT D Scores, self-reported preference scores, feeling thermometer difference scores, and individual feeling thermometers in Study 4.

	1	2	3	4	5	6	7	*M*	*SD*
1. Typicality 1	-							3.01	1.04
2. Typicality 2	.23***	-						3.31	1.24
3. Typicality 3	.30***	.58***	-					3.12	0.93
4. IAT	.06***	.07***	.05***	-				0.41	0.47
5. Preference	.04***	.09***	.09***	.15***	-			4.13	1.03
6. Thermometer	.05***	.10***	.09***	.12***	.46***	-		-0.14	1.98
7. Ingroup	.04***	.03***	.06***	.08***	.27***	.48***	-	3.62	2.21
8. Outgroup	.00	-.06***	-.02***	-.04***	-.15***	-.43***	.58***	3.76	2.15

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; Typicality 1 = ability to hide disability ; Typicality 2 = extent to which affected by disability; Typicality 3 = perceived severity of disability; IAT = IAT *D* Score; Preference = self-reported preference score; Thermometer = feeling thermometer difference score; Ingroup = ingroup feeling thermometer; Outgroup = outgroup feeling thermometer. Higher means on IATs and self-report measures indicate a stronger preference for disabled individuals relative to non-disabled individuals.

Discussion

Study 4 provided evidence that disabled participants display higher levels of outgroup favoritism the more they were able to hide their disability, the less they felt affected by their disability and the less severe they judged their disability. Although these effects were substantially smaller than in the previous studies, it is worth highlighting that the pattern of results was consistent across both IAT and self-report measures.

Meta-Analysis

Study 1 to 4 indicate that ingroup typicality is associated with ingroup and outgroup favoritism on IATs and self-report measures of intergroup evaluations. In order to compute a meta-analytic average effect size across studies, we first obtained one mean effect by averaging across *all* effects of each study that were weighted by their respective sample sizes. Next, we conducted a random effects meta-analysis of correlation coefficients across studies, using the metafor package (Viechtbauer, 2010), with effect sizes being weighted by their inverse sampling variance. This analysis yielded an average effect size of $r = .12$, $z = 4.18$, $p < .001$, 95% *CI* [0.06; 0.17], indicating an overall small effect of ingroup typicality on intergroup evaluations (see Figure 1).[28]

[28] We also observed a substantial amount of heterogeneity, $\tau^2 = 0.00$, $Q(3) = 228.16$, $p < .001$, accounting for a large proportion of the total variability, $I^2 = 99.36\%$.

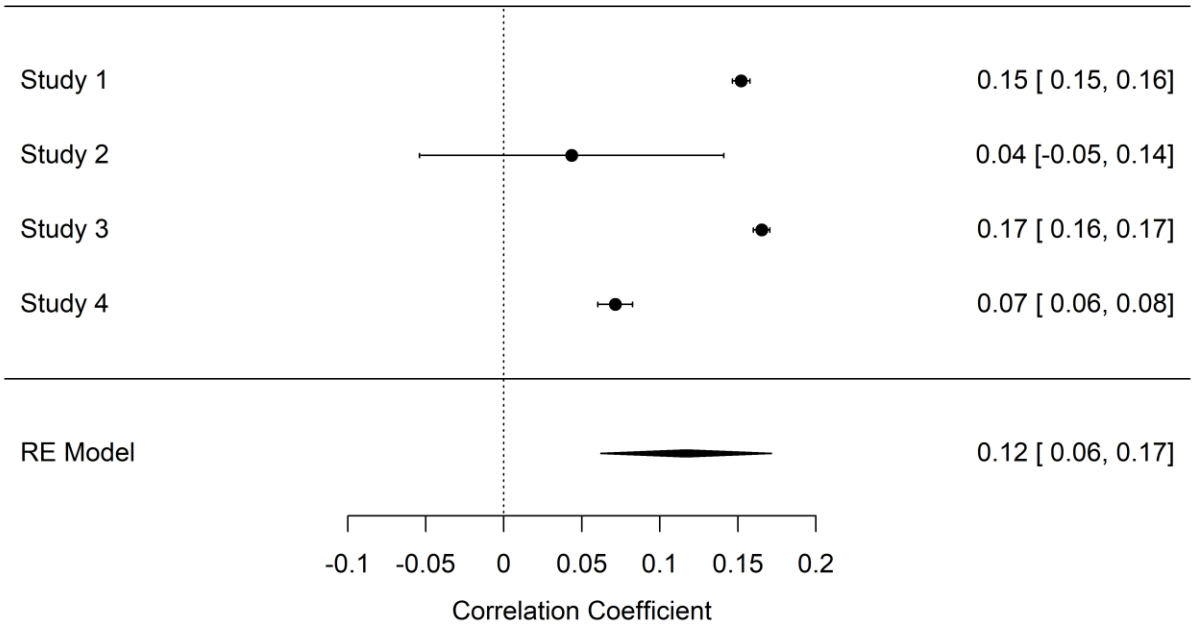

Figure 1. Forest plot of random-effects meta-analysis of results from Study 1 to 4. Error bars depict 95% confidence intervals.

In addition to combining effects across relative preference measures (i.e., IATs, self-reported preference scores, and feeling thermometer difference scores), the present data also provide the opportunity to conduct meta-analyses for correlations with single group evaluations (i.e., individual feeling thermometers). Correlations with relative preference measures versus single group evaluations are both informative: Correlations with relative preference measures allow inferences whether ingroup typicality is related to the extent to which people prefer the ingroup relative to the outgroup. In addition, correlations with single group evaluations allow inferences whether ingroup typicality is related to ingroup and/or outgroup evaluations. First, we

calculated average effect sizes as described above. Next, we conducted two separate random effects meta-analyses of correlation coefficients for ingroup and outgroup evaluations across studies, with effect sizes being weighted by their inverse sampling variance.

The meta-analysis using ingroup feeling thermometers yielded an average effect size of r = .07, z = 2.11, p = .035, 95% *CI* [0.00; 0.14]. This indicates that the effect of ingroup typicality on ingroup feeling thermometers was significant but very small. The meta-analysis using outgroup feeling thermometers yielded an average effect size of r = -.05, z = -2.97, p = .003, 95% *CI* [-0.08; -0.02].[29] This indicates that the effect of ingroup typicality on outgroup feeling thermometers was significant but very small. In sum, meta-analyses of correlations with single group evaluations suggest that higher levels of ingroup typicality are related to more positive ingroup evaluations *and* more negative outgroup evaluations. Consequently, both ingroup and outgroup evaluations contributed to correlations between ingroup typicality and feeling thermometer difference scores.[30]

---

[29] We observed substantial amounts of between-study heterogeneity for both ingroup evaluations, τ^2 = 0.00, $Q(3)$ = 212.87, p < .001, I^2 = 99.57%, and outgroup evaluations, τ^2 = 0.00, $Q(3)$ = 176.07, p < .001, I^2 = 97.72%.

[30] In addition to examining whether ingroup typicality is related to ingroup and/or outgroup evaluations, an important question is whether the magnitude of effect sizes differs for ingroup and outgroup evaluations. Unfortunately, due to the relatively small number of studies, the

General Discussion

The present research investigated whether members of stigmatized groups who appear less typical for their ingroup (e.g., light-skinned Black individuals; individuals with a disability that is less visible) are more likely to distance themselves from their group by evaluating the stigmatized ingroup less positively relative to a non-stigmatized outgroup. Across three social categories, intergroup evaluations on IATs and self-report measures varied with stigmatized group members' relative typicality for the ingroup. In Study 1, Black participants displayed a stronger preference for light-skinned relative to dark-skinned individuals the lighter their self-reported skin tone. This effect was partially replicated in Study 2, with Black participants displaying a stronger preference for Whites relative to Blacks the lighter their other-observed skin tone, but independent of their self-reported skin tone. In Study 3, overweight participants displayed a stronger preference for normal-weight relative to overweight individuals the lower their self-reported weight and the lower they believed others would judge their weight. In Study 4, participants with disabilities displayed a stronger preference for non-disabled relative to disabled individuals the more they reported being able to hide their disability, the less severe they judged their disability, and the less their disability affected their lives. In addition, exploratory analyses revealed that correlations between ingroup typicality and intergroup evaluations were at least partially mediated by stigmatized group members' level of ingroup identification (Study 2 and 3).

---

present data do not provide sufficient power to detect differences between the two subgroups of studies (Harrer, Cuijpers, Furukawa, & Ebert, 2019; Higgins & Thompson, 2004).

Together, these results highlight the role of subjective representations of ingroup typicality as one potential explanation for why members of stigmatized groups may (psychologically) distance themselves from their ingroup, in turn shaping group evaluations.

It is important to highlight that ingroup typicality was related to both ingroup and outgroup evaluations. In other words, to the extent that stigmatized group members displayed lower levels of ingroup typicality they also displayed more negative ingroup evaluations *and* more positive outgroup evaluations. The observed effects of ingroup typicality on intergroup evaluations are consistent across social categories and across measures, but they are small. This might at least in part reflect the use of single-item measures with relatively few response categories (e.g., Loo, 2002; Lozano, García-Cueto, & Muñiz, 2008), and future research might use more reliable multi-item scales. Another possibility, of course, is that the investigated effect itself *is* small. Albeit small, the effect is far from negligible, because even statistically small effects may have large consequences on a societal level—if they apply to many people or if they apply repeatedly to the same individuals (Greenwald, Banaji, & Nosek, 2015). Furthermore, some scholars argue that small effects may have less explanatory power for single events, but are likely "consequential in the not-very-long run" (Funder & Ozer, 2019, p. 156). For example, while a person's phenotypic ingroup typicality may not affect *all* their daily social interactions, it may affect the relative frequency of positive or negative experiences with ingroup and outgroup members, thus having a cumulative effect over time. In sum, we observed small effects of ingroup typicality, which may still be consequential at a societal level.

The observed relationships between ingroup typicality and intergroup evaluations fit into the literature on self-group distancing for a number of reasons. In our view, outgroup favoritism

can be understood as a way of distancing the self from a stigmatized social identity. This notion of outgroup favoritism as self-group distancing is based on our finding that outgroup favoritism was more prevalent among stigmatized group members who reported lower levels of ingroup typicality. Similarly, previous research on self-group distancing has demonstrated that individuals distance themselves from a stigmatized social identity by perceiving or emphasizing dissimilarities with the ingroup or similarities with a non-stigmatized outgroup (e.g., Derks, Ellemers et al., 2011; Derks et al., 2015; Weiss & Lang, 2012). Second, the notion of outgroup favoritism as self-group distancing is further corroborated by our finding that outgroup favoritism was more likely among stigmatized group members who reported lower levels of ingroup identification. This negative relationship between outgroup favoritism and ingroup identification is also consistent with the self-group distancing literature, which suggests that self-group distancing is more likely among stigmatized group members who are less identified with their ingroup (Derks, Ellemers et al., 2011; Derks et al., 2016; Derks, van Laar et al., 2011). Third, the notion of outgroup favoritism as self-group distancing is consistent with research suggesting that group members who distance themselves from a stigmatized social identity may endorse negative ingroup stereotypes and display negative ingroup evaluations (e.g., Derks, Ellemers et al., 2011; Guimond et al., 2002), which are arguably related to outgroup favoritism (see Table 2, 3, 5, and 7 for correlations between group evaluations and ingroup favoritism). Taken together, we argue that outgroup favoritism is a form of self-group distancing because of its relations with ingroup typicality, ingroup identification, and group evaluations.

That said, there is one noteworthy conceptual difference between our interpretation of the present findings and our reading of the self-group distancing literature. Our reading of this literature is that it conceptualizes ingroup identification as a *moderator* of self-group distancing.

According to this view, self-group distancing should be more likely among low identified stigmatized group members and less likely among high identified stigmatized group members (e.g., Derks, Ellemers et al., 2011; Derks, van Laar et al., 2011). Because the self-group distancing literature views ingroup identification as a moderator, such research seems less focused on explaining why stigmatized group members display varying levels of ingroup identification in the first place. The present findings extend the self-group distancing literature by suggesting that (lower) identification with the ingroup may itself be one possible consequence of (lower) ingroup typicality. Thus, to the extent that ingroup typicality is based on phenotypic appearance (e.g., variations in skin tone or facial features; weight status; the visibility of a disability), it may influence stigmatized group members' tendency to identify with their ingroup (e.g., Brown et al., 1999; Wilkins et al., 2010), with potential downstream consequences for group evaluations. According to this alternative view, ingroup identification is conceptualized as a *mediator* of the relationship between ingroup typicality and group evaluations. This view is consistent with our finding that ingroup identification partially mediated the effects of ingroup typicality on intergroup evaluations (Study 2 and 3).

Think back to our opening example of Tyree and Jamal, who might not be equally categorized by others' as Black. We argue that these differences in (perceived) ingroup typicality might feed back to stigmatized group members' tendency to self-categorize as an ingroup member and to identify with the ingroup. Hence, group members might experience more or less flexibility to identify themselves with or distance themselves from the ingroup, eventually affecting their evaluations of the ingroup and outgroups. In sum, we argue that phenotypic appearance places a boundary on stigmatized group members' ingroup typicality, affecting the tendency to psychologically connect with the ingroup, and eventually affecting group

evaluations. Future longitudinal research would seem best positioned to establish these hypothesized causal relationships between ingroup typicality, ingroup identification, and group evaluations.

In addition to ingroup identification, our explanation of the relations between ingroup typicality and group evaluations centers on the perceived permeability of group boundaries and individual mobility (Tajfel & Turner, 1979). In our view, ingroup typicality might introduce an *individual constraint* with implications for the perceived permeability of group boundaries for members of stigmatized groups. For example, varying perceived ingroup typicality might influence the likelihood for a given member of a stigmatized group to be categorized, and thus to self-categorize, as a group member. In other words, to the extent that ingroup typicality reflects phenotypic appearance, it might affect a persons' likelihood of being perceived and to perceive themselves as a group member. These speculations are consistent with social identity theory (Tajfel & Turner, 1979), which highlights that individual mobility necessarily implies that stigmatized group members dis-identify with the ingroup. In fact, previous research has documented that individual mobility is related to ingroup identification, and that those who anticipate upward mobility also tend to have more negative attitudes regarding the ingroup (e.g., Chipeaux, Kulich, Iacoviello, & Lorenzi-Cioldi, 2017).

Lastly, low perceived ingroup typicality might also limit stigmatized group members' ability to form social bonds with fellow ingroup members: Group members who are perceived, or who perceive themselves, as less typical, may have more difficulty forming connections with other ingroup members. For example, previous research suggests that differences in skin tone also play an important role within Black communities (e.g., Harvey, Tennial, & Hudson Banks,

2017), affecting group members' feelings of acceptance. One study found that Black university students with darker skin tone also felt more accepted by their Black peers compared to lighter skinned students (Harvey et al., 2005). Hence, in addition to constraining the permeability of group boundaries, ingroup typicality might also constrain stigmatized group members' ability to form attachment with fellow ingroup members or with the ingroup as a whole (i.e., ingroup identification).

However, we recognize that we cannot draw firm conclusions regarding the causal nature of the relationship between ingroup typicality, ingroup identification, and group evaluations. Thus, our data cannot confirm the implied causal model, where ingroup typicality provides an individual constraint for ingroup identification and the permeability of group boundaries. Furthermore, prominent theorizing suggests that constructs with similarities to ingroup typicality are in fact sub-components of identification. In particular, Leach et al. (2008) suggest that the extent to which individuals see themselves as similar to the ingroup (i.e., self-stereotyping) is one facet of ingroup identification. Thus, this model would at least suggest bi-directional relationships between self-stereotyping and other forms of identification. Moreover, it is certainly possible that causality might work the other way around. For example, stigmatized group members who are less identified might perceive themselves as less typical for the ingroup; or stigmatized group members who evaluate the ingroup more negatively might be less identified *and* perceive themselves as less typical. These outlined alternative relations point to the need of future research. Ideally, future research needs to investigate causal effects, for example through experimentally varying perceptions of ingroup typicality (e.g., by changing characteristics of ingroup and/or outgroup prototypes) and directly measuring perceptions of group boundary

permeability, thus furthering our understanding of the psychological processes underlying the effects of ingroup typicality on intergroup evaluations.

Previous research has spent a lot of effort investigating how members of advantaged groups perceive, judge, and treat stigmatized group members; at the same time, research has largely overlooked how perceptions and attitudes within stigmatized groups might differ due to within-group differences, and how these may affect intergroup relations. The observed consistent pattern of effects for three different and unrelated social categories and across different measures indicates that research on ingroup typicality may be a promising approach to study why stigmatized group members sometimes do or do not display ingroup favoritism.

CHAPTER 4: DO MEMBERS OF DISADVANTAGED GROUPS DISPLAY INCREASED INGROUP FAVORITISM ON PERSONALIZED MEASURES OF GROUP EVALUATIONS?

Abstract

Members of disadvantaged groups do not always display ingroup favoritism, and sometimes even display outgroup favoritism. Some scholars explain such outgroup favoritism with internalized negativity; others point to procedural properties of measures. We investigated group evaluations among disadvantaged groups employing different implicit and explicit measures. In Study 1, Turkish-German participants completed either a standard or personalized version of an Implicit Association Test. In Study 2, Muslim participants completed a normative and personalized version of an Affect Misattribution Procedure (Study 2a) as well as a reverse correlation task as a measure of mental representations of the ingroup (Study 2b). Participants also completed self-report measures of group evaluations. The main findings of both studies were overall effects of ingroup favoritism across measures. Contrary to previous findings, group evaluations on implicit measures were not moderated by procedural variations, suggesting that "personalizing" measures did not significantly increase ingroup favoritism among members of disadvantaged groups.

Previous research suggests that members of disadvantaged groups do not always display ingroup favoritism, and sometimes even display a preference for advantaged outgroups over their own group. Furthermore, such outgroup favoritism seems to vary with the measurement and appears to be more pronounced on so called implicit measures than on explicit measures (for a recent meta-analysis see Essien, Calanchini, & Degner, in press).[31] Whereas some researchers argue that internalized negativity regarding the ingroup may account for outgroup favoritism (e.g., Jost, Banaji, & Nosek, 2004), others argue that procedural properties of implicit measures may account for varying levels of outgroup favoritism (e.g., Olson, Crawford et al., 2009). The present research examined whether procedural aspects of measures might account for differences in group evaluations among members of disadvantaged groups.

Positive Regard for the Self and the Ingroup

Intergroup research has provided numerous demonstrations of ingroup favoritism, the preference for the ingroup relative to outgroups both in and outside the lab [e.g., Greenwald and Pettigrew (2014); Mullen, Brown, and Smith (1992); Voigt et al. (2017). Social identity theory (SIT; Tajfel & Turner, 1979) provides one prominent explanation for ingroup favoritism. SIT is

---

[31] Our use of the term "implicit" refers to indirect measurement procedures and their outcomes (e.g., evaluations). Thus, our definition does not entail assumptions about underlying features of automaticity or mental representations that differ between implicit and explicit measures (e.g., Corneille & Hütter, 2020).

based on the premise that individuals generally strive to achieve (or maintain) a positive view of themselves. Because individuals also belong to social groups and integrate self-categorization and identification with these ingroups into their self-concepts (e.g., Turner & Reynolds, 2012), they are generally motivated to belong to groups that are positively distinctive from other groups and/or create such positive distinctiveness via social comparison processes. If, however, people belong to groups that are generally negatively evaluated and/or are faced with low societal status (i.e., disadvantaged groups), positive distinctiveness is hard or impossible to achieve. SIT hypothesizes that members of such social groups either abandon this social identity or resort to social creativity strategies—biased comparisons between the ingroup and relevant outgroups—to achieve positive ingroup distinctiveness and overcome their disadvantage.

Internalized Negativity as an Explanation for Outgroup Favoritism

However, members of disadvantaged groups do not always display ingroup favoritism. For example, whereas a meta-analysis suggests a general pattern of moderate ingroup favoritism in members of disadvantaged groups (Mullen et al., 1992), a re-analysis suggests that some of these groups in the same dataset displayed outgroup favoritism (Jost, 2001). Moreover, other studies suggest that members of disadvantaged groups display outgroup favoritism on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). For example, IAT scores of Black participants in the USA document relative preferences for White over Black faces, suggesting outgroup favoritism (e.g., Ashburn-Nardo, Knowles, & Monteith, 2003); in contrast, they often also openly express preferences for Blacks relative to Whites on self-report measures, suggesting ingroup favoritism(e.g., Nosek, Banaji, & Greenwald, 2002).

Such observations of divergent ingroup and outgroup favoritism on implicit and explicit measures inspired System Justification Theory (SJT; Jost et al., 2004). SJT proposes that outgroup favoritism on implicit measures reflects internalized societal attitudes formed due to social learning of repeated negative evaluations and stereotypes directed at their ingroup. Jost and colleagues termed this internalizing of attitudes that run counter to personal or group interests "false consciousness" (Jost & Banaji, 1994, p. 3). Jost and colleagues (2004) also argued that this social learning process is supported by a general motivation to perceive society and current status hierarchies as legitimate and thus justify the status quo even at the expense of own group interest.

However, Jost and colleagues further speculated that cognitive, emotional, and social factors cause members of disadvantaged groups to avoid open displays of outgroup favoritism. They argued that the awareness of negative ingroup evaluations elicits cognitive dissonance and negative emotions among members of disadvantaged groups (Jost et al., 2004; Jost, Pelham, & Carvallo, 2002). They also speculated that members of disadvantaged groups face intense pressures within their communities to display ingroup pride. Based on these ideas, SJT proposes that members of disadvantaged groups openly display ingroup favoritism on explicit measures. Jost and colleagues further presume that implicit measures circumvent effects of social desirability concerns or even tap into unconscious processes, and regard these measures as more sensitive to detect internalized outgroup favoritism. Consequently, Jost and colleagues hypothesized that "[m]embers of low-status groups will exhibit outgroup favoritism even on (a) open-ended, nonreactive, qualitative measures, and (b) implicit, nonconscious cognitive, affective, and behavioral measures." (Jost et al., 2004, p. 892). There are indeed a number of studies that have documented such dissociation between implicit and explicit measures of group evaluations in members of disadvantaged groups. For example, older adults displayed a strong

preference for younger people over older people on an IAT, suggesting outgroup favoritism, but no preference for either group on self-report measures (Jost et al., 2004). Black participants did not display a preference for Black people relative to White people on an IAT, while displaying ingroup favoritism on self-report measures. Lastly, whereas gay and lesbian participants displayed a small preference for gay people relative to straight people on an IAT, but a strong effect of ingroup favoritism on self-report measures. Together, these findings support the hypothesis of a dissociation of implicit and explicit measures of group evaluations in members of disadvantaged groups (see also Livingston, 2002; Nosek et al., 2002; Rudman, Feinberg, & Fairchild, 2002). Of note, all of these studies have used a standard variant of the IAT as implicit measure, in which a target categorization task (e.g., Black and White people) alternates with a bipolar evaluation task (e.g., good vs. bad).

Procedural Properties as an Explanation for Outgroup Favoritism

There are, however, a number of studies with diverging findings, documenting either no group preference (e.g., Payne, Cheng, Govorun, & Stewart, 2005) or ingroup favoritism among members of stigmatized groups (e.g., Axt, Ebersole, & Nosek, 2014; Degner & Wentura, 2009; Dunham, Srinivasan, Dotsch, & Barner, 2014; Olson, Crawford et al., 2009). It is striking that many of these studies have either used non-standard versions of the IAT (e.g., Axt et al., 2014; Olson, Crawford et al., 2009) or entirely different measures (e.g., Degner & Wentura, 2009; Dunham et al., 2014). This observation suggests the possibility that characteristics of the standard-IAT procedure affect the outcome of the measures and are (at least partly) responsible for the observation of outgroup versus ingroup favoritism among members of disadvantaged groups.

For example, when using a variant of the IAT that instructs participants to focus on either only positive or only negative valence, Axt and colleagues (Axt et al., 2014; Axt, Moran, & Bar-Anan, 2018) observed different IAT scores indicating different levels of ingroup favoritism or no group preference among members of different disadvantaged groups. For example, when a race IAT focused participants' responses on the positive evaluation dimension (e.g., Black faces and positive words), Black participants displayed ingroup favoritism; when it focused responses on the negative evaluation dimension (e.g., Black faces and negative words), Black participants displayed no group preference or even outgroup favoritism.

In another series of studies, Degner and Wentura (2009) used an evaluative priming procedure with masked (subliminal) prime presentation to investigate weight bias among normal- and heavyweight participants. Heavyweight participants demonstrated medium-to-large effects of ingroup favoritism on the priming measure but did not express ingroup favoritism nor ingroup positivity on explicit measures. These results are somewhat surprising, given that weight bias appears to be a very strong form of stigmatization (e.g., Charlesworth & Banaji, 2019; Marini et al., 2013) and outgroup favoritism on standard IATs seems to be most pronounced among heavyweight individuals compared to members of other disadvantaged groups (Essien et al., in press). Degner and Wentura (2009) argued at the time that characteristics of the measurement procedure may explain why results deviated from typical IAT findings: Typical IAT studies require participants to explicitly categorize stimuli according to social group membership; in contrast, the evaluative priming paradigm employed a very short presentation and masking of prime stimuli, such that Participants were subjectively unaware of the measurement purpose of the studies. Consequently, they speculated that these procedural differences might have reduced self-impression management concerns (i.e., fear of devaluation for expressing positive evaluation

for a negatively stigmatized group membership), and participants might have been less motivated to hide automatically activated positivity. While this argument is in line with SJT's assumptions that norm perception, social desirability concerns, and impression management may play a role in the investigation of social group evaluations in members of disadvantaged groups, it leads to opposite conclusions. Jost and colleagues (2004) presumed that social norm perception drives members of disadvantaged groups to openly express ingroup favoritism while covertly favoring the outgroup ("false consciousness"). Degner and Wentura (2009), however, presumed that social norm perception drives members of disadvantaged groups to openly express outgroup favoritism while personally favoring the ingroup.

Moreover, researchers have argued that implicit measures might reflect awareness of cultural stereotypes rather than personal prejudice (e.g., Arkes & Tetlock, 2004). According to this view, a higher score on an implicit measure does not imply that a person endorses a negative attitude about a group, but merely reflects the awareness of cultural stereotypes or societal evaluation of that group. Similarly, it has been repeatedly argued that (standard) IATs may be open to confounds with societal norms and evaluations (Karpinski & Hilton, 2001; Olson & Fazio, 2004). According to such views, IATs are sensitive to the structure of the environment, and reflect peoples' perception of societal norms rather than their personal beliefs (see Payne, Vuletich, & Lundberg, 2017).

A related argument was advanced by Olson, Crawford and colleagues (2009). Based on previous debates that the IAT may measure extra-personal associations (e.g., Olson & Fazio, 2004), they juxtaposed standard evaluative Black-White and gay-straight IATs with personalized variants of the same measures. In the personalized IATs, evaluative labels "good" and "bad"

were replaced by “I like” and “I dislike”, respectively. In two studies, they observed strong effects of ingroup favoritism among Black participants (Study 1) and gay men (Study 2) on the personalized IAT variant. Effects in a standard IAT, however, were mixed: Black Americans displayed small effects of outgroup favoritism and gay participants displayed moderate effects of ingroup favoritism. Olson and colleagues (2009) explained the difference between measures with task-specific aspects of the IAT. They presumed that by using the labels “good” and “bad” and by providing error feedback, the standard IAT was prone to assess “extrapersonal associations”, which shift (“contaminate”) the measurement outcome towards society’s (negative) evaluation of a social group (Olson & Fazio, 2004, p. 653; see also Han, Olson, & Fazio, 2006). Taken together, the work by Olsen and colleagues (2009) challenged the notion that outgroup favoritism on the IAT among members of disadvantaged groups is due to internalized negativity. Instead, they argue that outgroup favoritism on the IAT reflects extra-personal associations—mere knowledge or awareness of societal evaluations. However, the theoretical conceptualization of extra-personal associations has been questioned, and there is high uncertainty about how extrapersonal associations can be conceptualized (e.g., Gawronski, Peters, & LeBel, 2008).

Yet, it is conceivable that small task characteristics that make the measure more or less susceptible to perceived normative pressures may still alter whether and to what extent members of disadvantaged groups display ingroup favoritism on the IAT. Moreover, the same argument may apply to other measures as well, such as evaluative priming (Degner & Wentura, 2009; Wentura & Degner, 2010) or the affective misattribution procedure (Payne et al., 2005). So far, however, Olson et al.’s (Olson, Crawford et al., 2009) findings have not been replicated—as research with members of disadvantaged groups is generally scarce. Furthermore, previous

research has not yet tested whether changes in the normative structure of tasks (i.e., "personalization") have similar effects on other measures of the implicit family.

The Present Research

The present research examines group evaluations among members of disadvantaged groups who are stigmatized in their societal contexts. One goal was to investigate whether members of disadvantaged groups display outgroup favoritism on different indirect and direct measures. In Study 1, Turkish-German participants completed an IAT and a self-report measure, assessing evaluations of Turkish targets relative to German targets. In Study 2, Muslim and non-Muslim participants completed an Affect Misattribution Procedure (AMP; Payne et al., 2005), a reverse correlation task, and self-report measures, assessing evaluations of Arabs and Muslims relative to Whites. Another goal was to investigate aspects of the measurement procedure affect group evaluations. Study 1 was a close replication of Olson and colleagues' (2009) study, juxtaposing a standard IAT with a personalized variant of the IAT. In Study 2, Muslim and Non-Muslim participants completed a "normative" and a "personalized" version of the AMP. In a second part of Study 2, participants completed a reverse correlation task in which they created classification images of how they envisioned a "typical Muslim". These classification images were then analyzed for group differences regarding their perceived trustworthiness, likeability, intelligence, threat, and Arab stereotypicality.

Study 1

The first goal of Study 1 was to investigate whether members of a negatively stigmatized ethnic group would display ingroup favoritism or outgroup favoritism on the IAT. The second

goal was to examine whether participants displayed higher levels of ingroup favoritism on a personalized version of the IAT than on a standard IAT. A third goal was to investigate whether participants displayed higher levels of ingroup favoritism on self-report measures than on IATs.

Method

Participants.

Fifty-three persons who self-categorized as Turkish[32] (24 female) participated in this study. A majority of the sample (n = 48) had a Turkish migration background; however, most participants were also German nationals (n = 42). Participants were recruited on campus (n = 13), at community events at a local Alevi Muslim community center (n = 24), and at a Turkish women's club (n = 16). Participants mean age was 35.57 (SD = 13.12). The study was conducted in the summer/fall of 2014. The study was introduced as a *cross-cultural study*, intended to investigate whether social information, such as names, are processed differently, depending on one's cultural or ethnic background. Participants were reimbursed with 4 euros.

Materials.

IATs. Evaluations of Turks relative to Germans were assessed using a standard version and a personalized version. Both IATs used 20 Turkish names (e.g., 'Ahmet') and 20 German names (e.g., 'Frank') as target stimuli and 20 positive and 20 negative words as attribute stimuli. Both IATs contained five blocks, with two practice blocks at the beginning and one practice

---

[32] We coded participants as "Turkish", who reported speaking Turkish, who reported Turkey as their country of origin, or who reported Turkish as their ethnic group membership.

block between the two combined tasks. In the first practice block, participants categorized names as ‘German’ or ‘Turkish’, in the second practice block, they categorized words as “positive” or “negative”. In the third block, participants completed a combined task, in which each target category shared a response key with an attribute category (e.g., Turkish names and positive words versus German names and negative words). In the fourth block, participants completed a practice task, categorizing name stimuli with reversed key assignments. The fifth block was again comprised of a combined task, this time with each target category sharing a response key with the respective reversed attribute category (e.g., Turkish names and negative words versus German names and positive words). Each practice block was comprised of 20 trials and each combined block was comprised of 40 trials. We counterbalanced the order in which participants had to complete the two combined tasks.[33]

In the *standard IAT*, attribute categories were labeled as ‘positive’ and ‘negative’. Participants received error feedback (a red *X*), following incorrect categorizations, and were instructed to correct erroneous responses by pressing the correct key as quickly as possible. In the *personalized IAT*, adapted from Olson and Fazio (2004), attribute categories were labeled as “I like” and “I don’t like”. Participants did not receive error feedback, and were not instructed to correct any responses.

---

[33] There was a non-significant effect of block sequence, $t(46.88) = -1.97$, $p = .055$, $d_s = -0.54$, 95% *CI* [-1.09; 0.01]. We report analyses using task order as covariate in Appendix B.

We calculated IAT *D* Scores (Greenwald, Nosek, & Banaji, 2003; see also Lane, Banaji, Nosek, & Greenwald, 2007) for both IAT versions. Specifically, we selected only correct trials with latencies greater than 300ms and lower than 10,000ms and calculated the difference between the two combined blocks divided by their pooled standard deviation.[34] Higher (i.e., positive) IAT *D* Scores indicate a preference for Germans relative to Turks, whereas lower (i.e., negative) IAT *D* Scores indicate a preference for Turks relative to Germans.[35]

Feeling thermometers. After completing the IAT, participants provided evaluations of various social groups—Russians, Christians, Germans, Jews, Turks, Muslims, Kurds—using 101-point feeling thermometers, which ranged from -50 (*cold*) to +50 (*warm*). We were exclusively interested in how participants evaluated Turks relative to Germans; the other groups served as filler items.

Additional measures.

---

[34] No participant matched the exclusion criteria outlined for the scoring algorithm—i.e., participants with latencies below 300ms on more than 10% of trials (Greenwald, Nosek, & Banaji, 2003).

[35] Note that the calculated IAT *D* Score *did not* involve an 'error panelty', because the personalized IAT did not involve error feedback.

Participants completed the following additional measures in fixed order.[36]

Participants first completed pictorial measures of overlap of self, ingroup, and outgroup (OSIO), adapted from Schubert and Otten (2002), assessing identification with Turks and Germans. Participants then proceeded to a verbal 10-item OSIO scale, again assessing their identification with Turks first and Germans afterwards. Within each block, items were presented in random order and participants used a scale from 1 (*do not agree at all*) to 7 (*perfectly agree*).

Next, participants completed a 6-item measure of belief in a just world (Schmitt et al., 2008)[37], using a scale from 1 (*do not agree at all*) to 6 (*perfectly agree*).

Finally, participants completed a 12-item measure of collective self-esteem (CSE; Luhtanen & Crocker, 1992)—assessing one's evaluations of the ingroup (*private CSE*), the assumed societal prestige of the ingroup (*public CSE*), and the centrality or importance of one's identity (*identity* subscale)—using a scale from 1 (*do not agree at all*) to 7 (*perfectly agree*).

Design and procedure. Participants were greeted by a White, female experimenter, and tested individually or in groups of up to four. First, participants were informed that their participation was voluntary and that they could cancel their participation at any time, without

---

[36] Note that an additional acquaintance measure assessing ingroup and outgroup contact is not listed here, because participants' responses were not recorded due to a programming error.

[37] Cronbach's α was 0.80.

disadvantage. Participants first completed either the standard or the personalized IAT (counterbalanced between participants), followed by feeling thermometers. Next, they completed additional measures in the order described above. Participants were then thanked and reimbursed for their participation.

Results

All analyses were conducted using R.[38]

IAT *D* Scores. Overall, IAT *D* Scores (M = -0.20, SD = 0.35) were significantly below zero, $t(52) = -4.05$, $p < .001$, $d_z = -0.56$, 95% *CI* [-0.84; -0.26], indicating a preference for Turkish relative to German in this sample. Scores on the personalized IAT (M = -0.23, SD = 0.31, d_z = -0.77) did not significantly differ from scores on the standard IAT (M = -0.16, SD = 0.40, d_z = -0.40), $t(48.69) = -0.78$, $p = .438$, $d_s = -0.21$, 95% *CI* [-0.75; 0.33]. This indicates that ingroup favoritism was not significantly different between the two measures.

Feeling thermometers. Participants' feeling thermometer difference scores of Turks (M = 21.47, SD = 27.78) and Germans (M = 16.53, SD = 28.11) did not differ significantly, $t(52)$ =

---

[38] R (Version 3.6.2; R Core Team, 2017) and the R-packages *cowplot* (Version 1.0.0; Wilke, 2017), *data.table* (Version 1.12.8; Dowle & Srinivasan, 2017), *jmv* (Version 1.0.8; Selker, Love, & Dropmann, 2018), *knitr* (Version 1.28; Xie, 2015), *MBESS* (Version 4.6.0; Kelley, 2018), *papaja* (Version 0.1.0.9942; Aust & Barth, 2018), and *tidyverse* (Version 1.3.0; Wickham, 2017).

1.30, $p = .199$, $d_z = 0.18$, 95% *CI* [-0.09; 0.45], indicating that Turkish-German participants did not express a significant felt preference for Turkish relative to German.[39]

Discussion

This study investigated evaluations of Turks relative to Germans in a sample of Turkish-German participants. We observed that Turkish-German participants displayed an evaluative preference for Turks relative to Germans on the IAT with moderate-to-large effect sizes. Furthermore, we did not observe differences between the personalized and standard version of the IAT. Lastly, we did not observe a significant effect of ingroup favoritism on feeling thermometers. These findings contradict hypotheses by SJT (Jost et al., 2004), according to which one would expect (a) an expression of ingroup favoritism on the feeling thermometer and (b) an expression of outgroup favoritism in the IAT. We thus carefully conclude that SJT's assumptions may not generalize to members of *any* disadvantage group. Furthermore, although effect sizes for ingroup favoritism were smaller effects in the standard IAT than in the personalized IAT, these differences were not significant and thus did not fully replicate findings

[39] For the sake of completeness, we report descriptive statistics and zero-order correlations of all additional measures in Appendix B. Note, however, that their interpretability is severely limited given the small sample size (Schönbrodt & Perugini, 2013).

by Olson and colleagues (2009) who had reported a large effect for the difference between the two measures.[40]

Study 2

Study 2 used a variety of measures to capture intergroup evaluation in Muslim participants and in a control sample of Non-Muslim participants. The study was comprised of two parts. In the first part, participants completed two sequential priming measures and two self-report measures of group evaluations, a series of stimulus rating tasks as well as additional individual difference measures. In the second part, participants completed a reverse correlation task. In order to facilitate comprehension of methods and results, we report these study parts as Study 2A and 2B. We emphasize, however, that data of the "two" studies were acquired from the same sample of participants.

Study 2A

Study 2A had the same goals as Study 1. First, we aimed at investigating whether Muslim participants displayed ingroup or outgroup favoritism, depending on whether their evaluations were assessed using indirect or direct measures of intergroup evaluation. Second, we investigated

---

[40] One obvious limitation of this analysis, however, is the small sample size. A G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) post hoc power analysis of the effect of IAT version with d = 0.21 and α = 0.05 suggests that the power to detect an effect of this magnitude was 1 - β = 0.12.

whether intergroup evaluation differed depending on whether participants' responses on indirect measures were framed in personalized ("I like" vs. "I do not like") or normative ("positive" vs. "negative") terms.

Method

Participants. One hundred and ten participants (85 female; M_{age} = 24.53, SD_{age} = 5.42) were recruited for this study. Forty-six participants self-identified as Muslim (i.e., selected 'Islam' as their religion). The majority of Muslim participants (n = 40) were German nationals, but more than one third of these participants reported dual citizenship (n = 15). Furthermore, a majority of Muslim participants (n = 40) indicated speaking another language, and a substantial number (n = 25) reported "Turkish" as part of their ethnic background. Sixty-four participants were non-Muslim (i.e., selected either Christian [n = 28] or "none" as religious affiliation [n = 34]). The majority of non-Muslim participants (n = 58) were German nationals, two reported dual citizenship, and six reported a nationality other than German.

The study was advertised as a cross-cultural study on "symbolic speech comprehension". In an information leaflet, participants were informed that the study's goal was to investigate the ability of Muslims and non-Muslims to guess the meaning of unknown symbols and whether Muslim and non-Muslim participants would evaluate these symbols differently. Muslim participants were recruited through professional and social networks of the Muslim experimenter and were reimbursed with 6 Euros. Non-Muslim participants were recruited with the same coverstory through the university's recruitment portal and reimbursed with 6 Euros or partial course credit.

Materials.

AMP. We created four sets of portrait images to serve as primes in the Affect Misattribution Procedure (AMP; Payne et al., 2005), in an attempt to appropriately capture evaluations regarding Arab, Arab-Muslim, and White people. We selected 12 images of White, male individuals and 12 images of Moroccan, male individuals from the Radboud faces database (Langner et al., 2010). Images were cropped, such that eyes and noses were approximately at the center of each image. We used photo editing software to create two more sets of stimuli: A set of 12 Arab-Muslim primes was created by adding a *taqiyah*—a round Muslim skullcap—on top of the heads of Arab individuals; a set of 12 White primes, with non-Muslim headgear was created by adding a wool hat on top of the heads of White individuals. We used a grey square to serve as a neutral control prime. As target stimuli, we selected 60 Chinese characters from a set of 200 Chinese characters, provided by Keith Payne.[41]

Liking of exemplars. In order to create a direct measure of group evaluations, which was structurally comparable to the AMP, we asked participants to rate all prime images presented in AMP regarding their perceived likeability on a scale from 1 (*very unlikable*) to 9 (*very likable*). Prime exemplars were presented one at a time in individual random order. We calculated mean exemplar liking scores for each of the four categories, namely Arab and White exemplars with and without headgear, respectively.

Feeling thermometers.

---

[41] http://bkpayne.web.unc.edu/research-materials/

Participants provided evaluations of social groups (Buddhists, Christians, Hinduists, Muslims, Jews, Atheists, and Germans), using 101-point feeling thermometers, which ranged from -50 (*cold*) to +50 (*warm*). We were primarily interested in how participants evaluated Muslims relative to Germans.

Additional measures. We administered further ratings of primes as control variables and manipulation check. Participants rated all prime exemplars regarding their perceived attractiveness on a scale from 1 (*very unattractive*) to 9 (*very attractive*) as well as regarding their perceived Muslim stereotypicality on a scale from 1 (*not at all Muslim*) to 9 (*very Muslim*).[42] Participants provided these ratings in two separate blocks, with exemplars being presented one at a time and in random order. We used pictorial measures of overlap of self, ingroup, and outgroup (OSIO), adapted from Schubert and Otten (2002) to assess participants' identification with Muslims and Germans. Participants also completed a 6-item measure of belief in a just world (Schmitt et al., 2008), using a scale from 1 (*do not agree at all*) to 6 (*perfectly agree*) (Cronbach's α = .64).[43]

Design and procedure.

---

[42] Analyses confirmed that Arab portraits with headgear were rated as more stereotypically Muslim than other exemplars—see detailed results in Appendix B.

[43] We report descriptive statistics and zero-order correlations of all additional measures in Appendix B. Note, however, that their interpretability is severely limited given the small sample size (Schönbrodt & Perugini, 2013).

The experiment followed a 2 (AMP Version: personalized vs. normative) by 2 (Prime Ethnicity: Arab vs. White) by 2 (Prime Headwear: present vs. absent) by 2 (Participant Group: Muslim vs. non-Muslim) mixed design. The study was conducted in the spring/summer of 2014. Upon arrival in the lab, participants were greeted by one of two female experimenters: A female experimenter of North African descent wearing a hijab (i.e., Arab-Muslim experimenter); or a White female who did not wear a head covering (i.e., White non-Muslim experimenter).[44] Participants were informed that their participation was voluntary and that they could abort participation at any time without disadvantage. Participants were tested individually or in groups of up to four on individual PCs in the computer lab of the university social psychology department. All measures were administered using the experimental software Inquisit (4.0.6.0). Participants first completed two versions of the AMP—a personalized version and a normative version—in counterbalanced order.

In both AMP versions, a prime was presented for 75ms, followed by a blank screen for 125ms. Next, a Chinese character was presented for 200ms, followed by a black-and-white pattern mask. The pattern mask remained until participants pressed one of the two response keys. Participants were instructed to ignore the prime stimulus, but to evaluate the relative pleasantness

---

[44] Based on the social tunig literature (e.g, Huntsinger, Sinclair, Kenrick, & Ray, 2016; Sinclair, Huntsinger, Skorinko, & Hardin, 2005), we analyzed whether participants' group evaluations differed depending on the presence of an Arab-Muslim versus White non-Muslim experimenter (see below).

of the Chinese character instead. The next trial began after a post-trial pause of 500ms. Each AMP version started with 10 practice trials. In both AMP versions, each of the 48 prime stimuli was presented once; the neutral prime was presented in 12 trials. Both AMP versions thus contained 60 trials in an individual random sequence. In the *personalized AMP*, participants were instructed to evaluate the visual pleasantness of Chinese characters, by using a left-hand key ("I don't like") or a right-hand key ("I like"). In the *normative AMP*, participants were told that their task was to guess whether the Chinese characters had a positive or negative meaning, by using a left-hand key ("negative") or a right-hand key ("positive").

Next, participants rated AMP stimuli for attractiveness and likeability in two separate blocks, which were presented in random order. Participants then rated AMP stimuli regarding their perceived Muslim stereotypicality. Next, participants completed the feeling thermometers, pictorial measures of self and group overlap regarding Muslims and Germans, the measure of belief in a just world, and a series of demographic questions.

Results

All analyses were conducted using R.[45]

---

[45] R (Version 3.6.2; R Core Team, 2017) and the R-packages *cowplot* (Version 1.0.0; Wilke, 2017), *data.table* (Version 1.12.8; Dowle & Srinivasan, 2017), *here* (Version 0.1; Müller, 2017), *jmv* (Version 1.0.8; Selker et al., 2018), *knitr* (Version 1.28; Xie, 2015), *MBESS* (Version 4.6.0;

AMP. We calculated the number of times participants pressed the positive key after each prime category, yielding five priming scores for each AMP version: Positive responses after Arab primes without headgear, Arab primes with headgear, White primes without headgear, White primes without headgear as well as positive responses after neutral primes. We then subtracted the frequencies of positive key presses following neutral priming from each of the four priming scores, thus yielding priming scores of relative positivity or negativity compared to the neutral prime condition.

We submitted participants' AMP scores to a two (AMP Version: personalized vs. normative) by two (Prime Ethnicy: Arab vs. White) by two (Headwear: no vs. yes) by two (Participant Group: Muslim vs. non-Muslim) ANOVA with repeated measures on the first three factors.[46] We observed a main effect of Prime Ethnicity $F(1,102) = 8.09$, $p = .005$, $\eta_p^2 = .07$, 95% *CI* [0.01; 0.16], which was qualified by a two-way interaction of Prime Ethnicity and Participant Group, $F(1,102) = 5.92$, $p = .017$, $\eta_p^2 = .05$, 95% *CI* [0.01; 0.13]—see Figure 1. All other main

---

Kelley, 2018), *papaja* (Version 0.1.0.9942; Aust & Barth, 2018), and *tidyverse* (Version 1.3.0; Wickham, 2017)

[46] We additionally conducted a five-way ANOVA additionally inserting experimenter group membership (Muslim vs. Non-Muslim) into the analyses. We did in fact observe significant interactions. However, because these analyzes are limited by the extremely low numbers of participants per cell (ranging from 12 to 40), we are reluctant to interpret these findings.

effects and interactions were non-significant. Post hoc comparisons revealed that Muslim participants more often pressed the positive key in Arab prime trials (M = -0.02, SD = 2.58) than in White prime trials (M = -0.84 , SD = 2.93), $t(102) = 3.39$, $p = .006$, $d_z = 0.50$, 95% CI [0.20; 0.80], indicating ingroup favoritism. In contrast, non-Muslim participants did not press the positive key more often in White prime trials (M = -0.24, SD = 2.06) than in Arab prime trials (M = -0.23 , SD = 2.13), $t(102) = 0.33$, $p > .999$, $d_z = 0.04$, 95% CI [-0.20; 0.29].

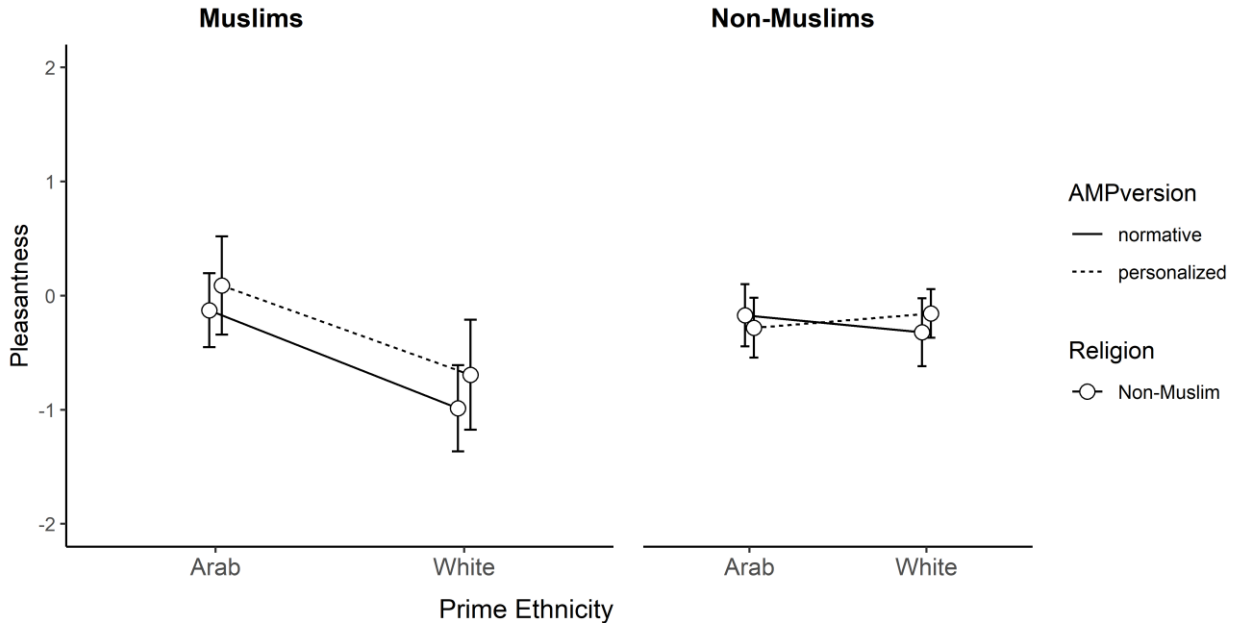

Figure 1. AMP priming scores of Muslim participants (left panel) and non-Muslim participants (right panel) in Study 2. Scores represent relative positivity or negativity compared to the neutral prime condition. Error bars depict standard errors.

Liking of exemplars. We submitted mean exemplar liking scores to a 2 (Prime Ethnicity: Arab vs. White) by 2 (Prime Headgear: yes vs. no) by 2 (Participant Group: Muslim vs. non-Muslim) ANOVA with repeated measures on the first two factors. This analysis yielded a main effect of Prime Ethnicity, $F(1,108) = 15.88$, $p < .001$, $\eta_p^2 = .13$, 95% CI [0.04; 0.23], which was qualified by a two-way interaction of Prime Ethnicity and Participant Group, $F(1,108) = 9.77$, $p =$

.002, η_p^2 = .08, 95% *CI* [0.02; 0.17]. Crucially, this two-way interaction was qualified by a three-way interaction of Prime Ethnicity, Prime Headgear, and Participant Group, $F(1,108) = 6.42$, p = .013, η_p^2 = .06, 95% *CI* [0.01; 0.14]—see Figure 2.[47]

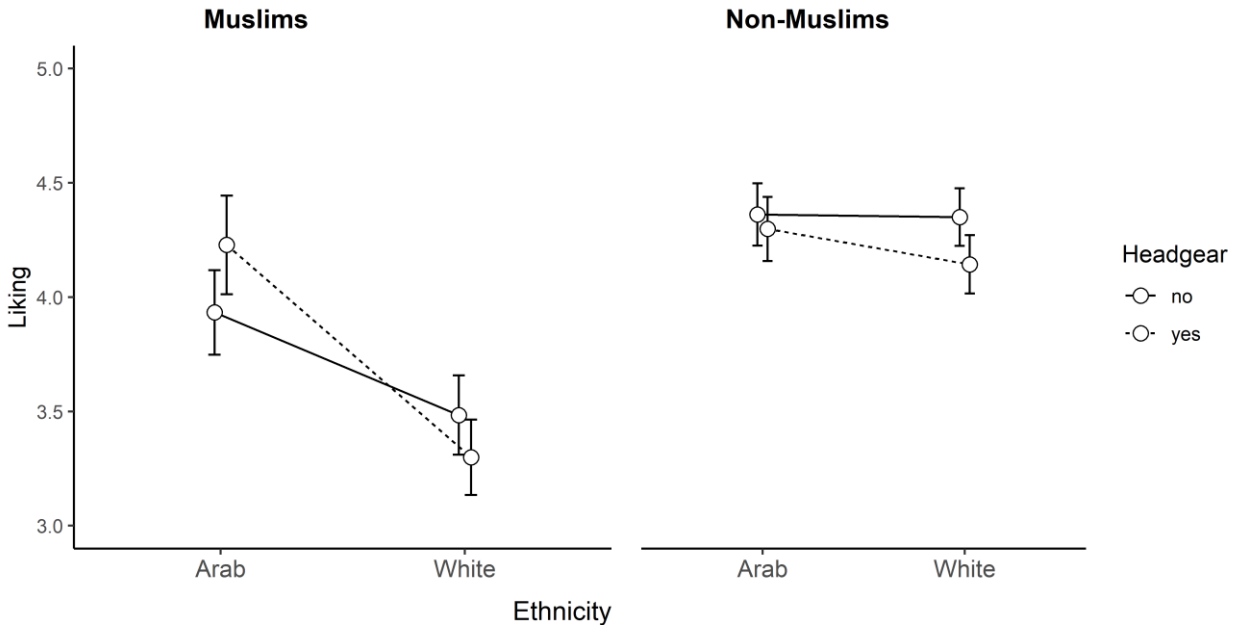

Figure 2. Exemplar liking scores of Muslim participants (left panel) and non-Muslim participants (right panel) in Study 2. Error bars depict standard errors.

[47] We also observed a main effect of Participant Group, $F(1,108) = 8.15$, p = .005, η_p^2 = .07, 95% *CI* [0.01; 0.16], a two-way interaction of Prime Headgear and Participant Group, $F(1,108) = 6.83$, p = .010, η_p^2 = .06, 95% *CI* [0.01; 0.14], which were, however, of less theoretical interest for the present research.

To decompose the three-way interaction. we conducted two follow-up ANOVAs, separately for Muslim and non-Muslim participants, and Bonferroni corrected post hoc t-tests. First, we submitted mean target liking scores of Muslim participants to a two (Prime Ethnicity: Arab vs. White) by two (Headgear: yes vs. no) ANOVA with repeated measures on both factors. This analysis yielded a main effect of Ethnicity, $F(1,45) = 17.75$, $p < .001$, $\eta_p^2 = .28$, 95% *CI* [0.00; 0.40], which was qualified by a two-way interaction of Prime Ethnicity and Prime Headgear, $F(1,45) = 15.54$, $p < .001$, $\eta_p^2 = .26$, 95% *CI* [0.04; 0.23]. Post hoc comparisons revealed that Muslim participants reported more liking for Arab targets with Muslim headgear than for Arab targets without Muslim headgear, $t(89.73) = -3.52$, $p = .004$, $d_z = -0.52$, 95% *CI* [-0.82; -0.22], more liking for Arab targets with Muslim headgear than for White targets without headgear, $t(56.00) = 4.29$, $p < .001$, $d_z = 0.63$, 95% *CI* [0.32; 0.94], and more liking for Arab targets with Muslim headgear than for White targets with headgear, $t(57.23) = 5.32$, $p < .001$, $d_z = 0.78$, 95% *CI* [0.46; 1.10]. Muslim participants also reported significantly more liking for Arab targets without Muslim headgear than for White targets with headgear, $t(56.00) = 3.65$, $p = .003$, $d_z = 0.54$, 95% *CI* [0.23; 0.84]. However, Muslim participants did not report significantly more liking for Arab targets without Muslim headgear, when compared with White targets without headgear, $t(57.23) = 2.57$, $p = .076$, $d_z = 0.38$, 95% *CI* [0.08; 0.68]—see Figure 2.

Next, we submitted mean target liking scores of non-Muslim participants the same two (Target Ethnicity: Arab vs. White) by two (Headgear: yes vs. no) ANOVA with repeated measures on both factors. This analysis yielded a main effect of Prime Headgear, $F(1,63) = 8.70$, $p = .004$, $\eta_p^2 = .12$, 95% *CI* [0.01; 0.16], which was qualified by a two-way interaction of Prime Ethnicity and Prime Headgear, $F(1,63) = 4.06$, $p = .048$, $\eta_p^2 = .06$, 95% *CI* [0.00; 0.11]. Post hoc

comparisons revealed that this two-way interaction was driven by non-Muslim participants' tendency to report more liking for White targets without headgear than for White targets with headgear, $t(118.59) = 3.57$, $p = .003$, $d_z = 0.45$, 95% *CI* [0.19; 0.70]. All other pair-wise comparisons were non-significant.

Feeling thermometers. We also compared Muslim participants' evaluations of Muslims and Germans on feeling thermometers, using a paired t-test, $t(45) = 5.51$, $p = < .001$, $d_z = 0.81$, 95% *CI* [0.48; 1.14], indicating that Muslim participants evaluated Muslims more favorably ($M = 34.80$, $SD = 20.10$) than Germans ($M = 14.80$, $SD = 21.91$). The same analysis of thermometer scores provided by non-Muslim participants revealed a significant difference in the opposite direction, $t(63) = -2.69$, $p = .009$, $d_z = -0.34$, 95% *CI* [-0.69; -0.10], indicating that non-Muslim participants evaluated Muslims significantly less favorably ($M = 6.86$, $SD = 19.21$) than Germans ($M = 16.00$, $SD = 23.14$).

Study 2B

In the second part of the study, participants completed a reverse correlation image classification task (Dotsch, Wigboldus, Langner, & van Knippenberg, 2008; Mangini & Biederman, 2004). Our goal was to explore Muslim participants' representations of a typical Muslim, and how they might differ from non-Muslim participants' representations of typical Muslims. The reverse correlation technique has been recently proposed as a data-driven indirect measure of mental representations of social groups (see also Degner, Mangels, & Zander, 2019).

The basic idea behind reverse correlation is that researchers create images of how participants' mentally represent social groups by averaging their responses on an image

classification task. First, the same image of a face—a so-called base face—is superimposed using different random noise patterns. The resulting images are differently appearing blurry faces, because each image is distorted differently by random noise. In a typical reverse correlation task, participants categorize faces regarding a social category of interest (e.g., gender; ethnicity) in a high number of trials (e.g., 300-800). A classification image is then created that averages the noise patterns of all images that participants choose and added to the base face. Thus, by creating images based on individual classification choices, reverse correlation allows visualization of participants' mental representations of social groups. These classification images can then be rated on (theoretically derived) evaluative dimensions. In principle, the task allows researchers to assess an infinite number of evaluative dimensions (cf. Brinkman, Todorov, & Dotsch, 2017). Reverse correlation has been used to study gendered (e.g., Degner et al., 2019) and ethnic (e.g., Dotsch et al., 2008) representations of social groups, and has also been applied to minimal groups contexts (Ratner, Dotsch, Wigboldus, van Knippenberg, & Amodio, 2014), making it a flexible tool for research on group evaluations.

Materials and Procedure

Part 1: Reverse correlation. We created a base image from 18 male, White faces and 18 male, Moroccan faces with neutral facial expression from the Radboud face database (Langner et al., 2010)—see Figure 3. The resulting morph was converted to grey-scale, blurred, and scaled to 512 x 512 pixels. We then generated 400 random noise patterns and for each noise pattern the mathematically opposite noise pattern (i.e., the negative). The noise patterns were constructed

using MATLAB code provided by Ron Dotsch (see Dotsch et al., 2008, for computational details).[48] This resulted in 800 stimuli in total.

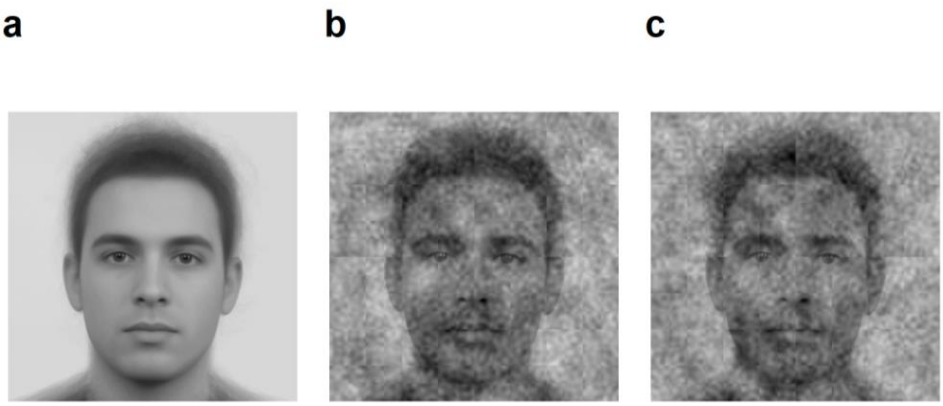

Figure 3. Base face (a) and exemplary individual classification images of Muslim participants (b) and non-Muslim participants (c). For illustrative purposes, we selected exemplar images, whose mean ratings of trustworthiness and stereotypicality were closest to the sample means of Muslim and non-Muslim participants, respectively.

Participants were told that this part of the study was designed to investigate whether religious group membership would be identifiable from a person's appearance. We told participants that they would see 400 image pairs of Muslim and non-Muslim individuals, which had been superimposed with random noise. On each trial, their task would be to identify which of the two people was Muslim. Every 100 trails, participants received mock feedback about their classification performance, as a means to increase motivation. After completing data collection,

[48] See also tutorial at http://www.rondotsch.nl/ rcicr/ for the more recent R package.

we averaged the noise patterns of all chosen stimuli per participant, thus creating 110 individual classification images (ICIs[49]).

Part 2: Image rating task. The 110 ICIs were rated by five independent samples of online participants, recruited via the online work platform *CrowdFlower* (now *Figure Eight*) and reimbursed with $0.30. Study participation was geographically restricted to respondents from the United Kingdom and the United States, and to more experienced workers with higher accuracy. The total sample size was $N = 287$ (136 female, 114 male, 35 unknown, 2 other) and participants' mean age was 37.83 years ($SD = 13.38$). A majority of the online sample (65.85%) self-identified as White or Caucasian (9% Hispanic, 6% Asian, 4% Native American or Pacific Islander, 4% African American, 2% other).[50]

The image rating task was advertised as an *online survey on person perception and first impressions* (see Appendix B for exact instructions). Online participants rated the ICIs regarding their perceived trustworthiness, likeability, intelligence, and threat on scales from 1 (*not at all*

---

[49] We avoid using the acronym "CI" originally introduced by Dotsch et al. (2008) for classification images, to stress that classification images were calculated individually and not as sample averages.

[50] Only respondents who passed an attention check (Oppenheimer, Meyvis, & Davidenko, 2009) were allowed to proceed with the rating task. Participants who did not pass the attention check ($n = 259$), were directly funneled to the end of the study.

[*trait*]) to 9 (*extremely* [*trait*]). These rating dimensions were chosen based on previous research using reverse correlation (Ratner et al., 2014), and because of research showing that stereotypes regarding Muslims in Western societies are related to threat (e.g., Spruyt & van der Noll, 2016). We also reasoned that four of these dimensions would loosely map onto two fundamental dimension of intergroup perception, namely warmth and competence (e.g., Fiske, Cuddy, Glick, & Xu, 2002): Specifically, we assumed that trustworthiness, likeability, and threat would most closely map onto the warmth dimension, whereas intelligence would map onto the competence dimension. We also included a fifth rating dimension, Arab stereotypicality. Online participants were asked to rate ICIs on a scale from -4 (*Caucasian / White*) to +4 (*Middle Eastern / Arab*). This rating dimension was chosen, because we wanted to explore whether Muslim and non-Muslim participants differed in the extend to which they associated stereotypically Arab or Middle Eastern features with Muslims.[51]

Each online judge rated all 110 ICIs on only one dimension. Stimuli were presented one at a time and in random order. Within each dimension, we aggregated ratings for each ICI. We calculated a mean rating score for each ICI on each dimension by aggregating across all participants (see Table 2 for descriptive statistics).

---

[51] We did not visually inspect ICIs prior to deciding which dimensions to include in the image-rating task, in order to avoid biased selection of dimensions (e.g., intentionally or unintentionally selecting dimensions based on visual information derived from ICIs).

Results

Correlational analyses. First, we performed exploratory correlational analyses of ICI rating dimensions (see Table 1). Note that correlations between ICI rating dimensions are actually between-subjects correlations, because each online participant rated ICIs only regarding one dimension. We found that those ICIs that were rated as more trustworthy, were also rated as more likeable, more intelligent, less threatening, and less stereotypically Arab looking. Furthermore, ICIs that were rated as more stereotypically Arab looking, were also rated as more threatening, less intelligent, and less likeable. ICIs that were rated as more likeable, were also rated as less threatening. Lastly, ICIs that were rated as more intelligent, were rated as less threatening, and more likeable.

Table 1

Zero-order correlations of ratings of individual classification images in Study 2B.

	1	2	3	4	*M*	*SD*
1. Trustworthiness	-				4.42	0.31
2. Threat	-.68***	-			4.54	0.35
3. Likeability	.67***	-.70***	-		4.28	0.27
4. Intelligence	.52***	-.40***	.54***	-	4.88	0.28
5. Stereotypicality	-.24*	.34***	-.26**	-.41***	5.40	0.50

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Primary analyses. Our main goal was to explore whether ICIs created by Muslim participants differed from ICIs created by non-Muslim participants in terms of perceived trustworthiness, likeability, intelligence, threat, and Arab stereotypicality. We submitted the five mean ICI ratings to a one-way multivariate analysis of variance (MANOVA) with Participant Group (Muslim vs. non-Muslim) as factor. We observed a significant multivariate effect of Participant Group on mean ICI ratings, Wilks' $\lambda = 0.86$, $F(5,103) = 3.49$, $p = .006$, and significant univariate effects for trust ratings and stereotypicality ratings (see Table 2 for descriptives and univariate test statistics). Muslim participants' ICIs were rated higher in trustworthiness than those of non-Muslim participants. Furthermore, Muslim participants' ICIs were rated less typical Arab looking than those of non-Muslim participants. All other univariate tests were non-significant.

Table 2:

Rating dimensions, means, SDs, sample sizes, and test statistics for ICI ratings in Study 2B.

	Muslims			Non-Muslims					
Dimension	*M*	*SD*	*N*	*M*	*SD*	*N*	*df*	*F*	*p*
Trustworthiness	4.52	1.95	57	4.36	1.95	57	1,107	7.26	0.008
Threat	4.47	1.92	54	4.58	1.93	55	1,107	2.38	0.126
Likeability	4.32	1.79	55	4.25	1.78	55	1,107	2.08	0.152
Intelligence	4.90	1.50	53	4.88	1.50	54	1,107	0.07	0.790

Stereotypicality	5.25	2.00	56	5.50	1.98	56	1,107	6.78	0.011

Discussion

Study 2 investigated group evaluations in a sample of German Muslim and non-Muslim participants. Replicating results of Study 1, Muslim participants displayed evaluative ingroup favoritism on the AMP. Again, we did not observe differences between a personalized and a normative version of the measure. Additionally, Muslim participants expressed ingroup favoritism in self-reported liking for Arab individuals relative to White individuals and also expressed more felt warmth regarding Muslims than regarding Germans on feeling thermometers. Finally, in a reverse correlation task, participants choices for faces that appeared more stereotypically Muslim resulted in more trustworthy-looking images compared to images created by non-Muslim participants, thus providing another indirect indicator of ingroup favoritism. However, ratings of classification images created by Muslim participants did not differ from those created by non-Muslim participants in terms of perceived likeability, intelligence, and threat.

Thus, results of Study 2A and B results point to consistent effects of ingroup favoritism in Muslim participants, which is consistent with other research on group evaluations in religious groups (e.g. Essien et al., in press).

General Discussion

The present research investigated intergroup evaluations among members of disadvantaged groups in German society employing different implicit and explicit measures. We

conducted two studies with Turkish-German (Study 1) and Muslim participants (Study 2)—two social identities that face immense social stigmatization in German society (e.g., Asbrock, 2010; Kaas & Manger, 2011; Kahraman & Knoblich, 2000; Schneider, Yemane, & Weinmann, 2014; Zick, Berghan, & Mokros, 2019; Zick, Küpper, & Krause, 2016). Our research further aimed at investigating potential differences between variants of implicit measures in their measurement of group evaluations. Therefore, participants completed either a standard or personalized version of an IAT (Study 1) or a normative or personalized version of the AMP (Study 2a). Additionally, we explored a reverse correlation task as an indirect measure of mental representations of the ingroup (Study 2b).

The main finding of both studies is consistent ingroup favoritism on implicit measures in both studies, whereas ingroup favoritism was less consistent on explicit measures. In Study 1, Turkish-German participants displayed a significant preference for Turkish relative to German targets on IATs, but no group preference when directly reporting their felt warmth regarding Turks and Germans. Ingroup favoritism on the IAT did not differ between the standard or personalized variants. In Study 2a, Muslim participants displayed a preference for Arab primes relative to White primes in an AMP. Again, ingroup favoritism in the AMP did not differ between personalized and normative variants of the measure. These findings contradict SJT (Jost et al., 2004), which postulates that members of disadvantaged groups display outgroup favoritism because of internalized negative evaluations regarding their ingroup.

Findings of our two studies also contrasts previous research (Olson, Crawford et al., 2009), because ingroup favoritism was unaffected by procedural variations of the measurements. Olsen and colleagues (2009) had argued that outgroup favoritism among members of

disadvantaged groups was less an effect of internalized attitudes but rather reproduction of an individual's awareness of cultural knowledge. They argued that reducing a measurement's susceptibility to such cultural knowledge about societal stigmatization of ones own ingroup may be reduced by "personalizing" the measures. In our studies, however, we observed effects of robust ingroup favoritism in both versions of the measures and there were no consistent differences between personalized measures and standard or normative measures. Thus, "personalizing" the measures to reduce normative demand did not correspond to more favorable evaluations regarding ingroup and outgroup among members of disadvantaged groups—perhaps because ingroup favoritism already occurred in the standard, normative variants of the measures. That said, it should be noted that sample sizes in our studies (as in the original studies by Olson, Crawford et al., 2009) were fairly small. Insufficient power to detect small differences between measures could explain why we did not observe moderation effects by measurement approach.

Together, our findings suggest that members of two disadvantaged groups, who are severely stigmatized in their societal context, did not display outgroup favoritism on implicit and explicit measures. Instead, they displayed either no group preferences or ingroup favoritism. These results arguably stand in contrast to SJT's (Jost et al., 2004) hypothesis that members of disadvantaged groups likely display outgroup favoritism on implicit measures as an indicator of internalized negativity regarding their ingroup identities.

Our finding that members of disadvantaged ethnic and religious groups displayed ingroup favoritism on implicit and (less consistently) explicit measures is consistent with other research documenting similar effects. In a recent meta-analysis of intergroup evaluations, Essien and colleagues observed high heterogeneity of group evaluations across multiple disadvantaged

groups (Essien et al., in press). Ingroup favoritism on implicit measures among members of disadvantaged ethnic and racial groups was not always mirrored by ingroup favoritism on explicit measures, which is consistent with findings of Study 1. Consistent with our finding in Study 2 that Muslim participants displayed ingroup favoritism across measures, their meta-analysis also observed strong effects of ingroup favoritism among members of disadvantaged religious groups on both implicit and explicit measures (Essien et al., in press). Similarly, children from a Muslim minority in another study displayed ingroup favoritism in terms of perceived warmth of their reverse correlation classification images depicting Muslims (Dunham et al., 2014). Such levels of ingroup favoritism might seem surprising, given that anti-Muslim sentiment is pervasive in Germany (Savelkoul, Scheepers, van der Veld, & Hagendoorn, 2012; Zick et al., 2019, 2016) and throughout the world (Hasan, 2019; Strabac & Listhaug, 2008), and based on research suggesting that stigma is related to outgroup favoritism among members of disadvantaged groups (e.g., Essien et al., in press). Dunham and colleagues (2014) speculated that aspects of religious identity might insulate members of religious groups against potential effects of stigma on ingroup devaluation. They argued that group affirming belief systems (e.g., moral codes) and practices (e.g., rituals, festivities)—which are common in many religions—might also foster positive attitudes regarding the religious ingroup. Furthermore, theorizing and research suggests that perceiving pervasive discrimination may cause members of disadvantaged groups to identify even stronger with their ingroups, as a buffer against societal rejection (Branscombe, Schmitt, & Harvey, 1999). Currently, we can only speculate about the mechanisms underlying such ingroup favoritism. But our findings extend previous research (Dunham et al., 2014; Essien et al., in press) by suggesting that *social identity* itself is an important moderator of group evaluations among disadvantaged groups.

Most importantly, our findings illustrate that there are many open questions with regard to how members of disadvantaged groups represent and evaluate their social identities, many of which have been overlooked in previous research, which heavily focused on investigating majority members' group evaluations. It also illustrates a dire gap in theorizing and research on the underlying psychological mechanism, antecedents, and consequences of intergroup evaluations and ingroup favoritism among members of disadvantaged groups.

REFERENCES

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

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or would jesse jackson 'fail' the implicit association test? *Psychological Inquiry*, *15*(4), 257–278. https://doi.org/10.1207/s15327965pli1504_01

Asbrock, F. (2010). Stereotypes of social groups in germany in terms of warmth and competence. *Social Psychology*, *41*(2), 76–81. https://doi.org/10.1027/1864-9335/a000011

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men* (pp. 221–236). Pittsburgh: Carnegie Press.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition*, *21*(1), 61–87. https://doi.org/10.1521/soco.21.1.61.21192

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin*, *130*(1), 80–114. https://doi.org/10.1037/0033-2909.130.1.80

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science*, *25*(9), 1804–1815. https://doi.org/10.1177/0956797614543801

Axt, J. R., Moran, T., & Bar-Anan, Y. (2018). Simultaneous ingroup and outgroup favoritism in implicit social cognition. *Journal of Experimental Social Psychology*, *79*, 275–289. https://doi.org/10.1016/j.jesp.2018.08.007

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Lawrence Erlbaum.

Bettencourt, B. A., Charlton, K., Dorr, N., & Hume, D. L. (2001). Status differences and in-group bias: A meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. *Psychological Bulletin*, *127*(4), 520–542. https://doi.org/10.1037/0033-2909.127.4.520

Bishop, C. J., Kiss, M., Morrison, T. G., Rushe, D. M., & Specht, J. (2014). The association between gay men's stereotypic beliefs about drag queens and their endorsement of hypermasculinity. *Journal of Homosexuality*, *61*(4), 554–567. https://doi.org/10.1080/00918369.2014.865464

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, *83*(1), 5–25. https://doi.org/10.1037//0022-3514.83.1.5

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity* (pp. 35–58). Oxford, England: Blackwell Science.

Branscombe, N. R., Schmitt, M. T., & Harvey, R. D. (1999). Perceiving pervasive discrimination among african americans: Implications for group identification and well-being. *Journal of Personality and Social Psychology*, *77*(1), 135–149. https://doi.org/10.1037/0022-3514.77.1.135

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B., & Campbell, D. T. (1976). *Ethnocentrism and intergroup attitudes: East african evidence*. Oxford, England: Sage.

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Broockman, D., Kalla, J., & Aronow, P. (2015). Irregularities in lacour (2014). *Working Paper*.

Brown, K. T., Ward, G. K., Lightbourn, T., & Jackson, J. S. (1999). Skin tone bias and racial identity among african americans: A theoretical and research framework. In R. L. Jones (Ed.), *Advances in african american psychology* (pp. 191–215). Hampton, VA: Cobb Publishers.

Brown-Iannuzzi, J. L., Lundberg, K. B., Kay, A. C., & Payne, B. K. (2015). Subjective status shapes political preferences. *Psychological Science*, *26*(1), 15–26. https://doi.org/10.1177/0956797614553947

Cadinu, M. R., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of Personality and Social Psychology*, *70*(4), 661–677. https://doi.org/10.1037/0022-3514.70.4.661

Carels, R. A., Wott, C. B., Young, K. M., Gumble, A., Koball, A., & Oehlhof, M. W. (2010). Implicit, explicit, and internalized weight bias and psychosocial maladjustment among treatment-seeking adults. *Eating Behaviors*, *11*(3), 180–185. https://doi.org/10.1016/j.eatbeh.2010.03.002

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Cheung, M. W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118957813

Chipeaux, M., Kulich, C., Iacoviello, V., & Lorenzi-Cioldi, F. (2017). I want, therefore i am - anticipated upward mobility reduces ingroup concern. *Frontiers in Psychology*, *8*, 1451. https://doi.org/10.3389/fpsyg.2017.01451

Clark, K. B., & Clark, M. P. (1950). Emotional factors in racial identification and preference in negro children. *The Journal of Negro Education*, *19*(3), 341. https://doi.org/10.2307/2966491

Clarkson, J. (2006). Everyday joe versus pissy, bitchy, queens: Gay masculinity on straightacting.com. *The Journal of Men's Studies*, *14*(2), 191–207. https://doi.org/10.3149/jms.1402.191

Coates, T.-N. (2015). The black family in the age of mass incarceration. *The Atlantic*. Retrieved from https://www.theatlantic.com/magazine/archive/2015/10/the-black-family-in-the-age-of-mass-incarceration/403246/

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*, 1–21. https://doi.org/10.1177/1088868320911325

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology*, *66*(5), 882–894. https://doi.org/10.1037/0022-3514.66.5.882

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, *129*(3), 414–446. https://doi.org/10.1037/0033-2909.129.3.414

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378. https://doi.org/10.1037/0022-3514.82.3.359

Crandall, C. S., Merman, A., & Hebl, M. (2009). Anti-fat prejudice. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 469–487). New York, NY: Psychology Press; US.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*(4), 608–630.

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, *30*(10), 1332–1346. https://doi.org/10.1177/0146167204264654

Czellar, S. (2006). Self-presentational effects in the implicit association test. *Journal of Consumer Psychology*, *16*(1), 92–100. https://doi.org/10.1207/s15327663jcp1601{\textunderscore }11

Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, *17*(2), 143–169. https://doi.org/10.1023/B:SORE.0000027407.70241.15

David, E. J. R., Schroeder, T. M., & Fernandez, J. (2019). Internalized racism: A systematic review of the psychological literature on racism's most insidious consequence. *Journal of Social Issues*, *75*(4), 1057–1086. https://doi.org/10.1111/josi.12350

Davies, P. G., Hutchinson, S., Osborne, D., & Eberhardt, J. L. (2016). Victims' race and sex leads to eyewitness misidentification of perpetrator's phenotypic stereotypicality. *Social Psychological and Personality Science*, *7*(6), 491–499. https://doi.org/10.1177/1948550616644655

Degner, J., Essien, I., & Reichardt, R. (2016). Effects of diversity versus segregation on automatic approach and avoidance behavior towards own and other ethnic groups. *European Journal of Social Psychology*, *46*(6), 783–791. https://doi.org/10.1002/ejsp.2234

Degner, J., Mangels, J., & Zander, L. (2019). Visualizing gendered representations of male and female teachers using a reverse correlation paradigm. *Social Psychology*, *50*(4), 233–251. https://doi.org/10.1027/1864-9335/a000382

Degner, J., & Wentura, D. (2009). Not everybody likes the thin and despises the fat: One's weight matters in the automatic activation of weight-related social evaluations. *Social Cognition*, *27*(2), 202–221.

Degner, J., & Wentura, D. (2010). Automatic prejudice in childhood and early adolescence. *Journal of Personality and Social Psychology*, *98*(3), 356–374. https://doi.org/10.1037/a0017993

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 176–193). New York: Guilford Press; US.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*(3), 347–368. https://doi.org/10.1037/a0014211

de Lemus, S., & Stroebe, K. (2015). Achieving social change: A matter of all for one? *Journal of Social Issues*, *71*(3), 441–452. https://doi.org/10.1111/josi.12122

Derks, B., Ellemers, N., van Laar, C., & Groot, K. de. (2011). Do sexist organizational cultures create the queen bee? *The British Journal of Social Psychology*, *50*(3), 519–535. https://doi.org/10.1348/014466610X525280

Derks, B., van Laar, C., & Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *The Leadership Quarterly*, *27*(3), 456–469. https://doi.org/10.1016/j.leaqua.2015.12.007

Derks, B., van Laar, C., Ellemers, N., & Groot, K. de. (2011). Gender-bias primes elicit queen-bee responses among senior policewomen. *Psychological Science*, *22*(10), 1243–1249. https://doi.org/10.1177/0956797611417258

Derks, B., van Laar, C., Ellemers, N., & Raghoe, G. (2015). Extending the queen bee effect: How hindustani workers cope with disadvantage by distancing the self from the group. *Journal of Social Issues*, *71*(3), 476–496. https://doi.org/10.1111/josi.12124

Devos, T., & Banaji, M. R. (2005). American = white? *Journal of Personality and Social Psychology*, *88*(3), 447–466. https://doi.org/10.1037/0022-3514.88.3.447

Dixon, T. L., & Maddox, K. B. (2005). Skin tone, crime news, and social reality judgments: Priming the stereotype of the dark and dangerous black criminal. *Journal of Applied Social Psychology*, *35*(8), 1555–1570. https://doi.org/10.1111/j.1559-1816.2005.tb02184.x

Doosje, B., Spears, R., & Ellemers, N. (2002). Social identity as both cause and effect: The development of group identification in response to anticipated and actual changes in the intergroup status hierarchy. *British Journal of Social Psychology*, *41*(1), 57–76. https://doi.org/10.1348/014466602165054

Dotsch, R., Wigboldus, D. H. J., 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

Du Bois, W. E. B. (1897). Strivings of the negro people. *The Atlantic Monthly*, *80*(478), 194–198.

Duncan, L. E. (1999). Motivation for collective action: Group consciousness as mediator of personality, life experiences, and women's rights activism. *Political Psychology*, *20*(3), 611–635. https://doi.org/10.1111/0162-895X.00159

Dunham, Y., Srinivasan, M., Dotsch, R., & Barner, D. (2014). Religion insulates ingroup evaluations: The development of intergroup attitudes in india. *Developmental Science*, *17*(2), 311–319. https://doi.org/10.1111/desc.12105

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy. *Psychological Science*, *17*(5), 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes*, *64*(3), 307–324. https://doi.org/10.1006/obhd.1995.1108

Effron, D. A., & Knowles, E. D. (2015). Entitativity and intergroup bias: How belonging to a cohesive group allows people to express their prejudices. *Journal of Personality and Social Psychology*, *108*(2), 234–253. https://doi.org/10.1037/pspa0000020

Eguchi, S. (2009). Negotiating hegemonic masculinity: The rhetorical strategy of "straight-acting" among gay men. *Journal of Intercultural Communication Research*, *38*(3), 193–209. https://doi.org/10.1080/17475759.2009.508892

Essien, I., Calanchini, J., & Degner, J. (in press). Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000302

Essien, I., & Degner, J. (2020). *Implicit intergroup evaluations in negatively stigmatized groups: A meta-analysis*. Unpublished manuscript.

Essien, I., Otten, S., & Degner, J. (in press). Group evaluations as self-group distancing: Ingroup typicality moderates evaluative intergroup bias in stigmatized groups. *European Journal of Social Psychology*.

Fanon, F. (1986). *Black skin, white mask*. London: Pluto Press. (Original work published 1952).

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

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The mode model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). New York: Academic Press. https://doi.org/10.1016/S0065-2601(08)60318-4

Fiedler, K., & Bluemke, M. (2005). Faking the iat: Aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology*, *27*(4), 307–316. https://doi.org/10.1207/s15324834basp2704{\textunderscore }3

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Flore, P. C., Mulder, J., & Wicherts, J. M. (2019). The influence of gender stereotype threat on mathematics test scores of dutch high school students: A registered report. *Comprehensive Results in Social Psychology*, 1–35. https://doi.org/10.1080/23743603.2018.1559647

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25–44. https://doi.org/10.1016/j.jsp.2014.10.002

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. Volume 44, pp. 59–127). Academic Press. https://doi.org/10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin*, *13*(3). https://doi.org/10.5964/spb.v13i3.28024

Gawronski, B., Brochu, P. M., Sritharan, R., & Strack, F. (2012). Cognitive consistency in prejudice-related belief systems: Integrating old-fashioned, modern, aversive, and implicit forms of prejudice. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 369–389). New York: Guilford Press; US.

Gawronski, B., & Hahn, A. (2019). Implicit measures: Procedures, use, and interpretation. In H. Blanton, J. M. LaCroix, & G. Webster (Eds.), *Measurement in social psychology* (pp. 29–55). New York, NY: Taylor & Francis. Retrieved from http://www.bertramgawronski.com/documents/GH2019Measurement.pdf

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are implicit attitudes unconscious? *Consciousness and Cognition*, *15*(3), 485–499. https://doi.org/10.1016/j.concog.2005.11.007

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*(3), 300–312. https://doi.org/10.1177/0146167216684131

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass*, *2*(2), 1002–1023. https://doi.org/10.1111/j.1751-9004.2008.00085.x

Greenaway, K. H., Haslam, S. A., Cruwys, T., Branscombe, N. R., Ysseldyk, R., & Heldreth, C. (2015). From we to me: Group identification enhances perceived personal control with consequences for health and well-being. *Journal of Personality and Social Psychology*, *109*(1), 53–74. https://doi.org/10.1037/pspi0000019

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*(7), 603–618. https://doi.org/10.1037/0003-066X.35.7.603

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, *72*(9), 861–871. https://doi.org/10.1037/amp0000238

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

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist*, *69*(7), 669–684. https://doi.org/10.1037/a0036056

Guimond, S., Dif, S., & Aupy, A. (2002). Social identity, relative group status and intergroup attitudes: When favorable outcomes change intergroup relations... For the worse. *European Journal of Social Psychology*, *32*(6), 739–760. https://doi.org/10.1002/ejsp.118

Gurin, P., Miller, A. H., & Gurin, G. (1980). Stratum identification and consciousness. *Social Psychology Quarterly*, *43*(1), 30. https://doi.org/10.2307/3033746

Hagiwara, N., Kashy, D. A., & Cesario, J. (2012). The independent effects of skin tone and facial features on whites' affective reactions to blacks. *Journal of Experimental Social Psychology*, *48*(4), 892–898. https://doi.org/10.1016/j.jesp.2012.02.001

Hahn, A., & Gawronski, B. (2014). Do implicit evaluations reflect unconscious attitudes? *Behavioral and Brain Sciences*, *37*(01), 28–29. https://doi.org/10.1017/S0140525X13000721

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2013). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*(3), 1369–1392. https://doi.org/10.1037/a0035028

Han, H., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally created extrapersonal associations on the implicit association test. *Journal of Experimental Social Psychology*, *42*(3), 259–272. https://doi.org/10.1016/j.jesp.2005.04.006

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. (2019). *Doing meta-analysis in r: A hands-on guide*. https://doi.org/10.5281/zenodo.2551803

Harvey, R. D., LaBeach, N., Pridgen, E., & Gocial, T. M. (2005). The intragroup stigmatization of skin tone among black americans. *Journal of Black Psychology*, *31*(3), 237–253. https://doi.org/10.1177/0095798405278192

Harvey, R. D., Tennial, R. E., & Hudson Banks, K. (2017). The development and validation of a colorism scale. *Journal of Black Psychology*, *43*(7), 740–764. https://doi.org/10.1177/0095798417690054

Hasan, M. (2019). Dear mark zuckerberg: Facebook is an engine of anti-muslim hate the world over. Don't you care? Retrieved from https://theintercept.com/2019/12/07/facebook-mark-zuckerberg-muslims-islamophobia/

Hebl, M. R., Williams, M. J., Sundermann, J. M., Kell, H. J., & Davies, P. G. (2012). Selectively friending: Racial stereotypicality and social rejection. *Journal of Experimental Social Psychology*, *48*(6), 1329–1335. https://doi.org/10.1016/j.jesp.2012.05.019

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Henry, P. J. (2011). The role of stigma in understanding ethnicity differences in authoritarianism. *Political Psychology*, *32*(3), 419–438. https://doi.org/10.1111/j.1467-9221.2010.00816.x

Herbert, B. (1993). In america; a sea change on crime. *The New York Times*. Retrieved from https://www.nytimes.com/1993/12/12/opinion/in-america-a-sea-change-on-crime.html

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Higgins, J. P. T., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, *23*(11), 1663–1682. https://doi.org/10.1002/sim.1752

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., … Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new sdo7 scale. *Journal of Personality and Social Psychology*, *109*(6), 1003–1028. https://doi.org/10.1037/pspi0000033

Hodson, G., & Dhont, K. (2015). The person-based nature of prejudice: Individual difference predictors of intergroup negativity. *European Review of Social Psychology*, *26*(1), 1–42. https://doi.org/10.1080/10463283.2015.1070018

Hoffarth, M. R., & Jost, J. T. (2017). When ideology contradicts self-interest: Conservative opposition to same-sex marriage among sexual minorities—a commentary on pinsof and haselton (2016). *Psychological Science*, *28*(10), 1521–1524. https://doi.org/10.1177/0956797617694866

Houwer, J. de, van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in Experimental Social Psychology*, *61*, 127–183. https://doi.org/10.1016/bs.aesp.2019.09.004

Hunter, J. E., & Schmidt, F. L. (2007). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, California: Sage Publications.

Huntsinger, J. R., Sinclair, S., Kenrick, A. C., & Ray, C. (2016). Affiliative social tuning reduces the activation of prejudice. *Group Processes & Intergroup Relations*, *19*(2), 217–235. https://doi.org/10.1177/1368430215583518

Jackson, L. A., Sullivan, L. A., Harnish, R., & Hodge, C. N. (1996). Achieving positive social identity: Social mobility, social creativity, and permeability of group boundaries. *Journal of Personality and Social Psychology*, *70*(2), 241–254. https://doi.org/10.1037/0022-3514.70.2.241

Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: In group preferences and related behaviors and beliefs among gay and straight men. *Personality and Social Psychology Bulletin*, *30*(5), 629–642. https://doi.org/10.1177/0146167203262076

Jost, J. T. (2001). Outgroup favoritism and the theory of system justification: A paradigm for investigating the effects of socioeconomic success on stereotype content. In G. Moskowitz (Ed.), *Cognitive social psychology* (pp. 89–102).

Jost, J. T. (2019). A quarter century of system justification theory: Questions, answers, criticisms, and societal applications. *British Journal of Social Psychology*, *58*(2), 263–314. https://doi.org/10.1111/bjso.12297

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Jost, J. T., Gaucher, D., & Stern, C. (2015). The world isn't fair": A system justification perspective on social stratification and inequality. In M. Mikulincer & P. R. Shaver (Eds.), *APA handbook of personality and social psychology* (Vol. 2, pp. 317–340). Washington DC: American Psychological Association.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Jost, J. T., & Hunyady, O. (2002). The psychology of system justification and the palliative function of ideology. *European Review of Social Psychology*, *13*(1), 111–153. https://doi.org/10.1080/10463280240000046

Jost, J. T., & Hunyady, O. (2005). Antecedents and consequences of system-justifying ideologies. *Current Directions in Psychological Science*, *14*(5), 260–265. https://doi.org/10.1111/j.0963-7214.2005.00377.x

Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, *38*(6), 586–602. https://doi.org/10.1016/s0022-1031(02)00505-x

Jost, J. T., & Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among african americans and european americans. *Journal of Experimental Social Psychology*, *36*(3), 209–232. https://doi.org/10.1006/jesp.1999.1403

Jost, J. T., & van der Toorn, J. (2012). System justification theory. In Van Lange, Paul A., A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (p. 537). Thousand Oaks: Sage Publications Ltd.

Jost, J. T., Wakslak, C. J., & Tyler, T. R. (2008). System justification theory and the alleviation of emotional distress: Palliative effects of ideology in an arbitrary social hierarchy and in society. In *Justice* (Vol. 25, pp. 181–211). Emerald Group Publishing Limited. https://doi.org/10.1016/S0882-6145(08)25012-5

Kaas, L., & Manger, C. (2011). Ethnic discrimination in germany's labour market: A field experiment. *German Economic Review*, *13*(1), 1–20. https://doi.org/10.1111/j.1468-0475.2011.00538.x

Kahn, K. B., & Davies, P. G. (2011). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes & Intergroup Relations*, *14*(4), 569–580. https://doi.org/10.1177/1368430210374609

Kahraman, B., & Knoblich, G. (2000). Stechen statt sprechen: Valenz und Aktivierbarkeit von Stereotypen über Türken [Stereotypes about turks: Valence and activation.]. *Zeitschrift Für Sozialpsychologie*, *31*(1), 31–43. https://doi.org/10.1024//0044-3514.31.1.31

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology*, *81*(5), 774–788. https://doi.org/10.1037/0022-3514.81.5.774

Kay, A. C., & Jost, J. T. (2003). Complementary justice: Effects of poor but happy and poor but honest stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology*, *85*(5), 823–837. https://doi.org/10.1037/0022-3514.85.5.823

Kay, A. C., & Jost, J. T. (2014). Theoretical integration in motivational science: System justification as one of many autonomous motivational structures. *Behavioral and Brain Sciences*, *37*(02), 146–147. https://doi.org/10.1017/S0140525X13002057

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, *47*(3), 193–198. https://doi.org/10.1023/A:1021003307511

Kite, M. E., Stockdale, G. D., Whitley, B. E., & Johnson, B. T. (2005). Attitudes toward younger and older adults: An updated meta-analytic review. *Journal of Social Issues*, *61*(2), 241–266. https://doi.org/10.1111/j.1540-4560.2005.00404.x

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The abc of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675. https://doi.org/10.1037/pspa0000046

Krieger, N., Carney, D., Lancaster, K., Waterman, P. D., Kosheleva, A., & Banaji, M. (2010). Combining explicit and implicit measures of racial discrimination in health research. *American Journal of Public Health*, *100*(8), 1485–1492. https://doi.org/10.2105/AJPH.2009.159517

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. *Frontiers in Psychology*, *4*, 1–12. https://doi.org/10.3389/fpsyg.2013.00863

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging*, *30*(1), 180–193. https://doi.org/10.1037/a0038586

Lane, K. A., Mitchell, J. P., & Banaji, M. R. (2005). Me and my group: Cultural status can disrupt cognitive consistency. *Social Cognition*, *23*(4), 353–386. https://doi.org/10.1521/soco.2005.23.4.353

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

Latrofa, M., Vaes, J., Cadinu, M., & Carnaghi, A. (2010). The cognitive representation of self-stereotyping. *Personality and Social Psychology Bulletin*, *36*(7), 911–922. https://doi.org/10.1177/0146167210373907

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, Michael L W, Pennekamp, S. F., Doosje, B., … Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*(1), 144–165. https://doi.org/10.1037/0022-3514.95.1.144

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. Boston, MA: Springer US.

Levy, B. R., & Banaji, M. R. (2002). Implicit ageism. In T. D. Nelson (Ed.), *Ageism* (pp. 49–75). Cambridge, MA: The MIT Press.

Levy, B. R., Hausdorff, J. M., Hencke, R., & Wei, J. Y. (2000). Reducing cardiovascular stress with positive self-stereotypes of aging. *The Journal of Gerontology: Series B*, *55*(4), 205–213. https://doi.org/10.1093/geronb/55.4.p205

Levy, B. R., Zonderman, A. B., Slade, M. D., & Ferrucci, L. (2009). Age stereotypes held earlier in life predict cardiovascular events in later life. *Psychological Science*, *20*(3), 296–298. https://doi.org/10.1111/j.1467-9280.2009.02298.x

Lewin, K. (1948). *Resolving social conflicts: Selected papers on group dynamics*. New York: Harper & Brothers.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology*, *27*(1), 363–385. https://doi.org/10.1146/annurev.soc.27.1.363

Livingston, R. W. (2002). The role of perceived negativity in the moderation of african americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology*, *38*(4), 405–413. https://doi.org/10.1016/S0022-1031(02)00002-1

Loo, R. (2002). A caveat on using single–item versus multiple–item scales. *Journal of Managerial Psychology*, *17*(1), 68–75. https://doi.org/10.1108/02683940210415933

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*(2), 73–79. https://doi.org/10.1027/1614-2241.4.2.73

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, *18*(3), 302–318. https://doi.org/10.1177/0146167292183006

Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology*, *47*(2), 391–396. https://doi.org/10.1016/j.jesp.2010.11.002

Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review*, *8*(4), 383–401. https://doi.org/10.1207/s15327957pspr0804{\textunderscore }4

Major, B., & O'Brien, L. T. (2005). The social psychology of stigma. *Annual Review of Psychology*, *56*, 393–421. https://doi.org/10.1146/annurev.psych.56.091103.070137

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*(2), 209–226. https://doi.org/10.1207/s15516709cog2802{\textunderscore }4

Marini, M., Sriram, N., Schnabel, K., Maliszewski, N., Devos, T., Ekehammar, B., … Nosek, B. A. (2013). Overweight people have low levels of implicit weight bias, but overweight

nations have high levels of implicit weight bias. *PloS ONE*, *8*(12), e83543. https://doi.org/10.1371/journal.pone.0083543

Massey, D. S., & Martin, J. A. (2003). The nis skin color scale. Princeton, NJ: Princeton University Press.

Meeus, J., Mayor, J. P., González, R., Brown, R., & Manzi, J. (2017). Racial phenotypicality bias in educational expectations for both male and female teenagers from different socioeconomic backgrounds. *European Journal of Social Psychology*, *47*(3), 289–303. https://doi.org/10.1002/ejsp.2247

Mentovich, A., & Jost, J. T. (2008). The ideological id? System justification and the unconscious perpetuation of inequality. *Conneticut Law Review*, *40*(4), 1095–1116.

Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, *54*(12), 1053–1060. https://doi.org/10.1037/0003-066X.54.12.1053

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mouzon, D. M., & McLean, J. S. (2017). Internalized racism and mental health among african-americans, us-born caribbean blacks, and foreign-born caribbean blacks. *Ethnicity & Health*, *22*(1), 36–48. https://doi.org/10.1080/13557858.2016.1196652

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103–122. https://doi.org/10.1002/ejsp.2420220202

Mummendey, A., & Otten, S. (1998). Positive–negative asymmetry in social discrimination. *European Review of Social Psychology*, *9*(1), 107–143. https://doi.org/10.1080/14792779843000063

Neugarten, B. L. (1974). Age groups in american society and the rise of the young-old. *The ANNALS of the American Academy of Political and Social Science*, *415*(1), 187–198. https://doi.org/10.1177/000271627441500114

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6(1)*, 101–115. https://doi.org/10.1037/1089-2699.6.1.101

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., … Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88. https://doi.org/10.1080/10463280701489053

Oldmeadow, J. A., & Fiske, S. T. (2010). Social status and the pursuit of positive social identity: Systematic domains of intergroup differentiation and discrimination for high- and low-status groups. *Group Processes & Intergroup Relations*, *13*(4), 425–444. https://doi.org/10.1177/1368430209355650

Olson, M. A. (2009). Measures of prejudice. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 367–386). New York, NY: Psychology Press; US.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the implicit association test: Personalizing the iat. *Journal of Personality and Social Psychology*, *86*(5), 653–667. https://doi.org/10.1037/0022-3514.86.5.653

Olson, M. A., Fazio, R. H., & Han, H. A. (2009). Conceptualizing personal and extrapersonal associations. *Social and Personality Psychology Compass*, *3*(2), 152–170. https://doi.org/10.1111/j.1751-9004.2008.00164.x

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin*, *82*(1), 85–86. https://doi.org/10.1037/h0076158

Oxford Dictionaries. (2019). *Maghreb*. Retrieved from https://en.oxforddictionaries.com/definition/us/Maghreb

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248. https://doi.org/10.1080/1047840X.2017.1335568

Phills, C. E., Hahn, A., & Gawronski, B. (2020). The bidirectional causal relation between implicit stereotypes and implicit prejudice. *Personality and Social Psychology Bulletin*, 1–13. https://doi.org/10.1177/0146167219899234

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology*, *76*(1), 114–128. Retrieved from http://dx.doi.org/10.1037/0022-3514.76.1.114

Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, *17*(1), 271–320. https://doi.org/10.1080/10463280601055772

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741–763. https://doi.org/10.1037/0022-3514.67.4.741

Pyke, K. D. (2010). What is internalized racial oppression and why don't we study it? Acknowledging racism's hidden injuries. *Sociological Perspectives*, *53*(4), 551–572. https://doi.org/10.1525/sop.2010.53.4.551

Rae, J. R., Newheiser, A.-K., & Olson, K. R. (2015). Exposure to racial out-groups and implicit race bias in the united states. *Social Psychological and Personality Science*, *6*(5), 535–543. https://doi.org/10.1177/1948550614567357

Ramos, M. R., Cassidy, C., Reicher, S., & Haslam, S. A. (2012). A longitudinal investigation of the rejection-identification hypothesis. *British Journal of Social Psychology*, *51*(4), 642–660. https://doi.org/10.1111/j.2044-8309.2011.02029.x

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, *106*(6), 897–911. https://doi.org/10.1037/a0036498

Reimer, N. K., Becker, J. C., Benz, A., Christ, O., Dhont, K., Klocke, U., … Hewstone, M. (2016). Intergroup contact and social change: Implications of negative and positive contact for collective action in advantaged and disadvantaged groups. *Personality and Social Psychology Bulletin*, *43*(1), 121–136. https://doi.org/10.1177/0146167216676478

Rivers, A. M., & Hahn, A. (2018). What cognitive mechanisms do people reflect on when they predict iat scores? *Personality and Social Psychology Bulletin*, *45*(6), 878–892. https://doi.org/10.1177/0146167218799307

Rivers, A. M., Rees, H. R., Calanchini, J., & Sherman, J. W. (2017). Implicit bias reflects the personal and the social. *Psychological Inquiry*, *28*(4), 301–305. https://doi.org/10.1080/1047840X.2017.1373549

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351. https://doi.org/10.2307/2087176

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Thousand Oaks, CA, US: Sage Publications, Inc. https://doi.org/10.4135/9781412984997

Rubin, M., Badea, C., & Jetten, J. (2014). Low status groups show in-group favoritism to compensate for their low status and compete for higher status. *Group Processes & Intergroup Relations*, *17*(5), 563–576. https://doi.org/10.1177/1368430213514122

Rudman, L. A., & Ashmore, R. D. (2007). Discrimination and the implicit association test. *Group Processes & Intergroup Relations*, *10*(3), 359–372. https://doi.org/10.1177/1368430207078696

Rudman, L. A., Feinberg, J., & Fairchild, K. (2002). Minority members' implicit attitudes: Automatic ingroup bias as a function of group status. *Social Cognition*, *20*(4), 294–320. https://doi.org/10.1521/soco.20.4.294.19908

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, *94*(1), 18–38. https://doi.org/10.1037/0033-2909.94.1.18

Savelkoul, M., Scheepers, P., van der Veld, W., & Hagendoorn, L. (2012). Comparing levels of anti-muslim attitudes across western countries. *Quality & Quantity*, *46*(5), 1617–1624. https://doi.org/10.1007/s11135-011-9470-9

Schmitt, M., Dalbert, C., Montada, L., Gschwendner, T., Maes, J., Reichle, B., … Braehler, E. (2008). Verteilung des glaubens an eine gerechte welt in der allgemeinbevoelkerung. Normwerte fuer die skala allgemeiner gerechte-welt-glaube. Distribution of the belief in a just world in the general population: Norms for the general-belief-in-a-just-world-scale. *Diagnostica*, *54*(3), 150–163. https://doi.org/10.1026/0012-1924.54.3.150

Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychological Bulletin*, *140*(4), 921–948. https://doi.org/10.1037/a0035754

Schneider, J., Yemane, R., & Weinmann, M. (2014). *Diskriminierung am Ausbildungsmarkt: Ausmaß, Ursachen und Handlungsempfehlungen*. Berlin: SVR-Forschungsbereich.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity*, *1(4)*, 353–376. Retrieved from http://dx.doi.org/10.1080/152988602760328012

Schwartz, M. B., Vartanian, L. R., Nosek, B. A., & Brownell, K. D. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity*, *14*(3), 440–447. https://doi.org/10.1038/oby.2006.58

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515–530. https://doi.org/10.1037/0022-3514.51.3.515

Selvin, H. C. (1958). Durkheim's suicide and problems of empirical research. *American Journal of Sociology*, *63*(6), 607–619. https://doi.org/10.1086/222356

Sherman, J. W. (2009). Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes* (pp. 391–426). New York, NY: Psychology Press.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*(2), 314–335. https://doi.org/10.1037/0033-295X.115.2.314

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY, US: Cambridge University Press.

Sidanius, J., & Pratto, F. (2004). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge [u.a.]: Cambridge University Press.

Sinclair, S., Huntsinger, J., Skorinko, J., & Hardin, C. D. (2005). Social tuning of the self: Consequences for the self-evaluations of stereotype targets. *Journal of Personality and Social Psychology*, *89*(2), 160–175. https://doi.org/10.1037/0022-3514.89.2.160

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, *67*, 415–437. https://doi.org/10.1146/annurev-psych-073115-103235

Spicer, C. V. (1999). *Effects of self-stereotyping and stereotype threat on intellectual performance* (Dissertation). University of Kentucky, Lexington, KY, United States.

Spruyt, B., & van der Noll, J. (2016). The “islamized stranger”: On “chronic” versus “contextual” salience in the measurement of anti-muslim prejudice. *Political Psychology*. https://doi.org/10.1111/pops.12383

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology*, *51*(3), 165–179. https://doi.org/10.1027/1618-3169.51.3.165

Strabac, Z., & Listhaug, O. (2008). Anti-muslim prejudice in europe: A multilevel analysis of survey data from 30 countries. *Social Science Research*, *37*(1), 268–286. https://doi.org/10.1016/j.ssresearch.2007.02.004

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In S. Worchel & W. G. Austin (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey; Calif.: Brooks/Cole.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210. https://doi.org/10.1037/0033-2909.103.2.193

Teige Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to implicit association tests and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117–139). New York: Guilford Press; US.

Thurman, W. (2015). *The blacker the berry: A novel of negro life*. Mansfield Centre, CT: Martino Publishing. (Original work published 1929).

Tiggemann, M., & Rothblum, E. d. (1997). Gender differences in internal beliefs about weight and negative attitudes towards self and others. *Psychology of Women Quarterly*, *21*(4), 581–593. https://doi.org/10.1111/j.1471-6402.1997.tb00132.x

Todd, A. R., & Burgmer, P. (2013). Perspective taking and automatic intergroup evaluation change: Testing an associative self-anchoring account. *Journal of Personality and Social Psychology*, *104*(5), 786–802. https://doi.org/10.1037/a0031999

Turner, J. C., & Reynolds, K. J. (2012). Self-categorization theory. In Van Lange, Paul A., A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 399–417). Thousand Oaks: Sage Publications Ltd.

Uhlmann, E., Dasgupta, N., Elgueta, A., Greenwald, A. G., & Swanson, J. (2002). Subgroup prejudice based on skin color among hispanics in the united states and latin america. *Social Cognition*, *20*(3), 198–226. https://doi.org/10.1521/soco.20.3.198.21104

van Veelen, R., Otten, S., Cadinu, M., & Hansen, N. (2016). An integrative model of social identification: Self-stereotyping and self-anchoring as two cognitive pathways. *Personality and Social Psychology Review*, *20*(1), 3–26. https://doi.org/10.1177/1088868315576642

van Veelen, R., Otten, S., & Hansen, N. (2011). Linking self and ingroup: Self-anchoring as distinctive cognitive route to social identification. *European Journal of Social Psychology*, *41*(5), 628–637. https://doi.org/10.1002/ejsp.792

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, *134*(4), 504–535. https://doi.org/10.1037/0033-2909.134.4.504

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

Viechtbauer, W., Lopez-Lopez, J. A., Sanchez-Meca, J., & Marin-Martinez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods*, *20*(3), 360–374. https://doi.org/10.1037/met0000023

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., … Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, *114*(25), 6521–6526. https://doi.org/10.1073/pnas.1702413114

Weiss, D., & Freund, A. M. (2012). Still young at heart: Negative age-related information motivates distancing from same-aged people. *Psychology and Aging*, *27*(1), 173–180. https://doi.org/10.1037/a0024819

Weiss, D., & Lang, F. R. (2012). They are old but I feel younger: Age-group dissociation as a self-protective strategy in old age. *Psychology and Aging*, *27*(1), 153–163. https://doi.org/10.1037/a0024887

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York: Guilford Press; US.

White, K. R. G., Kinney, D., Danek, R. H., Smith, B., & Harben, C. (2020). The resistance to change-beliefs scale: Validation of a new measure of conservative ideology. *Personality and Social Psychology Bulletin*, *46*(1), 20–35. https://doi.org/10.1177/0146167219841624

Wilkins, C. L., Kaiser, C. R., & Rieck, H. (2010). Detecting racial identification: The role of phenotypic prototypicality. *Journal of Experimental Social Psychology*, *46*(6), 1029–1034. https://doi.org/10.1016/j.jesp.2010.05.017

Wright, S. C., & Lubensky, M. E. (2009). The struggle for social equality: Collective action versus prejudice reduction. In S. Demoulin, J. P. Leyens, & J. F. Dovidio (Eds.), *Intergroup misunderstandings: Impact of divergent social realities* (pp. 291–310). New York: Psychology Press; US.

Xu, F. K., Nosek, B. A., Greenwald, A. G., Ratliff, K. A., Bar-Anan, Y., Umansky, E., … Smith, C. (2017). Project implicit demo website datasets. https://doi.org/10.17605/OSF.IO/Y9HIQ

Xu, F. K., Nosek, B. A., Greenwald, A. G., Ratliff, K. A., Umansky, E., & Lofarno, N. (2018). Project implicit international website datasets. Retrieved from osf.io/kaqi5

Zick, A., Berghan, W., & Mokros, N. (2019). Gruppenbezogene menschenfeindlichkeit in deutschland 2002–2018/19. In A. Zick, B. Küpper, & W. Berghan (Eds.), *Verlorene mitte - feindselige zustände* (pp. 53–116). Bonn: DIETZ.

Zick, A., Küpper, B., & Krause, D. (2016). *Gespaltene Mitte – feindselige Zustände: Rechtextreme Einstellungen in Deutschland 2016*. Friedrich-Ebert-Stiftung.

APPENDIX A

Correlational Analyses of SDO and RWA

Here, we report supplemental correlational analyses using two different measures, which have been interpreted as indexing system-justifying beliefs (Jost & Hunyady, 2005). These correlational analyses are based on a sub-sample of Project Implicit studies, which assessed these construct.

Method

Study Selection. We included only studies which reported measures of SDO and / or RWA. The final number of datasets included in the present study was $n = 24$, yielding a total of $k = 24$ independent effect sizes.

Measures.

Intergroup evaluation. We used the same three dependent measures as reported in the main text.

Right-wing authoritarianism. Right-wing authoritarianism (RWA) was assessed using a 15-item measure, adapted from Zakrisson (2005) (e.g., "Our country needs a powerful leader, in order to destroy the radical and immoral currents prevailing in society today"). Each participant responded to on average four randomly selected items from this measure, using a scale from 1 (*strongly disagree*) to 6 (*strongly agree*).

Social dominance orientation. Social dominace orientation (SDO) was assessed using a 12-item measure, adapted from Pratto, Sidanius, Stallworth, and Malle (1994) (e.g., "Some

people are just inferior to others"). Each participant responded to on average four randomly selected items from this measure, using a scale from 1 (*strongly disagree*) to 6 (*strongly agree*).

Data analysis. R scripts and RMarkdown scripts to reproduce data preparation, analyses, figures, and tables can be found at https://osf.io/cxp9z/.

Results

SDO. To test the relationship between disadvantaged group members' SDO and intergroup evaluation at the individual level, we calculated the correlation between SDO and the three measures of intergroup evaluation within each sample and fitted three separate random-effects models, weighting each correlation coefficient by its corresponding sample size.

IAT D Scores. We observed an average effect of $r = .06$, $z = 7.51$, $p < .001$, 95% *CI* [0.05; 0.08], indicating that the correlation between SDO and IAT *D* Scores at the individual level was very small. The estimated amount of total heterogeneity was $\tau^2 = 0.0005$, $Q(22) = 183.94$, $p < .001$, accounting for a large proportion of the total variability, $I^2 = 71.61\%$.

One-item preference scores. We observed an average effect of $r = .10$, $z = 5.95$, $p < .001$, 95% *CI* [0.07; 0.13], indicating that the correlation between SDO and one-item preference scores at the individual level was small. The estimated amount of total heterogeneity was $\tau^2 = 0.00$, $Q(22) = 467.15$, $p < .001$, accounting for a large proportion of the total variability, $I^2 = 95.96\%$.

Feeling thermometer difference scores. We observed an average effect of $r = .05$, $z = 4.02$, $p < .001$, 95% *CI* [0.02; 0.07], indicating that the correlation between SDO and feeling thermometer difference scores at the individual-level was very small. The estimated amount of

total heterogeneity was τ^2 = 0.00, $Q(21) = 138.56$, $p < .001$, accounting for a large proportion of the total variability, I^2 = 83.08%.

Taken together, we observed positive but small correlations between SDO and intergroup evaluation: Members of disadvantaged groups displayed more favorable evaluations towards an advantaged outgroup relative to their disadvantaged ingroup the higher their self-reported SDO levels.

RWA. To test the relationship between disadvantaged group members' RWA and intergroup evaluation at the individual level, we calculated the correlation between RWA and the three measures of intergroup evaluation within each sample and fitted three separate random-effects models, weighting each correlation coefficient by its corresponding sample size.

IAT D Scores. We observed an average effect of r = -.01, z = -4.06, p = < .001, 95% *CI* [-0.02; -0.01], indicating that RWA and IAT *D* Scores were uncorrelated at the individual level. The estimated amount of total heterogeneity was τ^2 = 0.0000, $Q(22) = 27.16$, p .205, with I^2 = 0.00%.

One-item preference scores. We observed an average effect of r = .00, z = 0.11, p = .915, 95% *CI* [-0.02; 0.03], indicating RWA and one-item preference scores were uncorrelated at the individual level. The estimated amount of total heterogeneity was τ^2 = 0.00, $Q(22) = 86.65$, $p <$ .001, with I^2 = 84.91%.

Feeling thermometer difference scores. We observed an average effect of r = .01, z = 0.67, p = .503, 95% *CI* [-0.01; 0.03], indicating that RWA and feeling thermometer difference

scores were uncorrelated at the individual-level. The estimated amount of total heterogeneity was $\tau^2 = 0.00$, $Q(21) = 58.57$, $p < .001$, with $I^2 = 80.19\%$.

Taken together, on average, correlational analyses with RWA yielded null effects across all three measures of intergroup evaluation: RWA levels among members of disadvantaged groups were unrelated to their intergroup evaluations.

Discussion

We ran a series of individual-level analyses assessing the relationships between SDO, RWA, and intergroup evaluation in disadvantaged group members. Here, we observed small correlations between SDO and outgroup favoritism for all three dependent measures. However, another pattern emerged regarding RWA, where we observed correlations close to zero for all three dependent measures.

Two issues complicate clear interpretations of these findings. First, participants did not complete either the full SDO or RWA scales. Instead, each participant responded to few randomly-selected items from each scale. Consequently, the available individual-level data suffers from both relatively high measurement error and relatively low construct validity—which, in turn, implies a high risk of underestimating the true relationship between system justifying beliefs and outgroup favoritism at the individual level. Second, the SDO and RWA measures were only included in approximately one-third of the datasets examined in the main analyses. This subset of datasets is very selective, with all focusing on race- and ethnicity-related intergroup attitudes. Hence, analyses that relied on this subset of datasets are less suited to

examine the generalizability of SJT's predictions, and thus stand in contrast to our main meta-analytic findings across a wider variety of social groups.

References

Jost, J. T., & Hunyady, O. (2005). Antecedents and consequences of system-justifying ideologies. *Current Directions in Psychological Science*, *14*(5), 260–265. https://doi.org/10.1111/j.0963-7214.2005.00377.x

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741–763. https://doi.org/10.1037/0022-3514.67.4.741

Zakrisson, I. (2005). Construction of a short version of the right-wing authoritarianism (rwa) scale. *Personality and Individual Differences*, *39*(5), 863–872. https://doi.org/10.1016/j.paid.2005.02.026

Exploratory Moderator Analyses of Cultural Differences

To further explore whether societal aspects moderate intergroup evaluations among disadvantaged groups in a more quantitative fashion, we searched for indices of cultural difference that may be related to system-justifying beliefs. Specifically, we used Hofstede's cultural value dimensions (Hofstede, Hofstede, & Minkov, 2010) that describe aggregate country differences, with the most well-established being collectivism-individualism, masculinity-femininity, uncertainty avoidance, and power-distance.

Using Power Distance as Continuous Moderator

Hofstede's power distance index would seem to be the most promising proxy for culture-level metrics of system-justifying beliefs, in that it reflects the degree to which the less powerful members of a society accept and expect that power is distributed unequally. As an index of societal inequality, this cultural value dimension appears to be conceptually most similar to system justification motives (Jost & Hunyady, 2005). Based on these data, we conducted a series of mixed-effects meta-regression analyses, but we did not observe significant moderation effects of the power distance index for IAT *D* Scores, $Q_M(1) = 0.02$, $p = .880$, one-item preference scores, $Q_M(1) = 2.37$, $p = .124$, or feeling thermometer difference scores, $Q_M(1) = 0.03$, $p = .866$.

Using Four Cultural Dimensions Simultaneously

We also ran a series of meta-regression analyses, each time simultaneously using the first four Hofstede dimensions–individualism, power distance, masculinity, and uncertainty avoidance–as continuous moderators. Here, we also did not observe significant overall moderation effects for IAT *D* Scores, $Q_M(4) = 1.35$, $p = .853$, one-item preference scores, $Q_M(4)$

= 8.45, p = .076, or feeling thermometer difference scores, $Q_M(4) = 3.58$, p = .465. These additional analyses suggest that cultural and societal differences–indexed by the Hofstede dimensions–do not seem to be reliably related to intergroup evaluations among disadvantaged groups. However, given the rather low power of these analyses (due to only n = countries), these are not strong tests of the relationship between country-level factors and outgroup favoritism in disadvantaged groups.

References

Jost, J. T., & Hunyady, O. (2005). Antecedents and consequences of system-justifying ideologies. *Current Directions in Psychological Science*, *14*(5), 260–265. https://doi.org/10.1111/j.0963-7214.2005.00377.x

Analyses with Advantaged Groups

Here, we report supplemental analyses using advantaged groups. Specifically, we examined overall effects of intergroup evaluation for all three measures of intergroup evaluation as well as relationships between conservatism and intergroup evaluations.

Method

Participants. The total sample size was N = 4,757,009.

Intergroup Domains. Table S1 provides an overview of advantaged group samples, dependent measures, moderators, and descriptive statistics.

Old vs. Young. From the k = 14 studies focusing on age-related group evaluations, we included only participants with a self-reported age of 54 years and younger (see Kite, Stockdale, Whitley, & Johnson, 2005; Neugarten, 1974).

Arab vs. French. From the k = 1 study focusing on evaluations of Arab people relative to French people, we included only participants who reported being French or White and who self-categorized as Christian (e.g., Protestant; Catholic).

Disabled vs. Abled. From the k = 1 study focusing on evaluations of disability relative to non-disability, we included only participants who indicated that they did not have a disability.

Black vs. White. From the k = 11 studies focusing on evaluations of Black people relative to White people, we included only participants who self-categorized as White.

Religious groups.

From the *k* = 2 studies focusing on evaluations of religious groups, we included only participants who self-categorized as Christian, thus yielding 2 independent samples.

Gay vs. Straight. From the *k* = 15 studies focusing on evaluations of gay people relative to straight people, we included only participants who self-categorized as heterosexual.

Dark-Skinned vs. Light-Skinned. From the *k* = 13 studies focusing on evaluations of dark-skinned people relative to light-skinned people, we included only participants who self-categorized as somewhat somewhat light-skinned, light-skinned, or very light-skinned.

Overweight vs. Normal Weight. From the *k* = 14 studies focusing on evaluations of overweight people relative to normal weight people, we included only participants who self-categorized as being neither underweight nor overweight, slightly, moderately, or very underweight.

Table A1

Overview of advantaged group samples, variables, and descriptive statistics.

		IAT		Preference		Thermometer		
Group	Country	dz	N	dz	N	dz	N	Conservatism
Younger (<55) age	AUS	1.16	16615	0.48	18332	0.19	18807	-0.28
Younger (<55) age	BEL	1.44	1698	0.62	1711	0.38	1841	-0.23
Younger (<55) age	BRA	1.01	4430	0.52	5152	0.26	5290	-0.25
Younger (<55) age	CAN (EN)	1.09	21171	0.41	22560	0.12	23040	-0.42
Younger (<55) age	CAN (FR)	1.27	1794	0.56	1816	0.29	1877	-0.43
Christian	CHN	1.25	7912	0.70	8540	-0.03	9077	-0.45
Younger (<55) age	DEU	1.42	16162	0.58	16145	0.29	17135	-0.51
Younger (<55) age	ESP	1.22	6913	0.52	7205	0.07	7468	-0.53
Younger (<55) age	FRA	1.44	14136	0.62	14506	0.32	15109	-0.35
Younger (<55) age	GBR	1.15	35603	0.39	39795	0.13	40662	0.01

Younger (<55) age	KOR	1.35	5911	1.16	6235	-0.26	7046	-0.33
Younger (<55) age	NLD	1.32	7512	0.62	8095	0.27	8422	-0.56
Younger (<55) age	SWE	1.20	8689	0.47	8270	0.09	9097	-0.04
Younger (<55) age	USA	1.10	711861	0.43	733460	0.09	757832	-0.15
Non-Arab / Non-Muslim	FRA	1.17	10457	0.70	10642	0.54	1106	-0.36
Non-disabled participants	USA	1.16	272975	0.54	285593	0.19	291482	-0.26
White participants	AUS	0.99	8833	0.57	9701	0.40	9755	-0.38
White participants	BEL	0.87	8817	0.64	9576	0.44	9984	-0.29
White participants	BRA	0.99	4990	0.55	6431	0.31	6487	-0.31
White participants	CAN (EN)	0.97	39599	0.50	41850	0.34	42615	-0.47
White participants	CAN (FR)	1.09	2767	0.60	2825	0.43	2919	-0.51
White participants	DEU	0.91	33528	0.52	34818	0.29	36254	-0.57
White participants	ESP	1.05	7929	0.67	8451	0.41	8692	-0.67
White participants	FRA	1.02	20146	0.51	21183	0.32	21790	-0.21

White participants	GBR	0.93	78012	0.57	90927	0.40	92420	-0.33
White participants	SWE	0.75	13898	0.58	13978	0.35	14821	-0.07
White participants	USA	0.96	1597308	0.61	1559488	0.47	1637808	-0.20
Christian	USA	1.17	8354	1.48	8146	0.86	8218	0.06
Christian	USA	1.37	8539	1.73	8321	1.09	8380	0.08
Straight	AUS	0.78	11012	0.62	13460	0.58	14162	-0.34
Straight	BEL	0.77	2330	0.82	2471	0.79	2851	-0.27
Straight	BRA	0.66	6812	0.89	8740	0.83	9143	-0.25
Straight	CAN (EN)	0.74	16684	0.62	18234	0.58	19175	-0.42
Straight	CAN (FR)	0.48	2044	0.62	1999	0.64	2256	-0.50
Christian	CHN	0.45	14518	1.27	17979	0.93	18714	-0.48
Straight	DEU	0.57	23868	0.76	23754	0.75	26778	-0.52
Straight	ESP	0.74	8665	0.74	9614	0.64	10410	-0.64
Straight	FRA	0.52	17878	0.66	18733	0.62	20749	-0.37

Straight	GBR	0.81	30963	0.66	35388	0.64	36920	-0.40
Straight	KOR	0.80	12973	1.05	18003	0.47	18651	-0.52
Straight	NLD	0.93	5861	0.78	6847	0.80	7370	-0.68
Straight	RUS	1.10	2596	1.48	2546	1.07	3071	-0.02
Straight	SWE	0.49	12682	0.61	12603	0.62	14615	-0.06
Straight	USA	0.87	767542	0.71	791703	0.69	824433	-0.25
Light-skinned	BEL	0.79	2560	0.55	2832	0.32	3142	-0.39
Light-skinned	BRA	0.95	4770	0.62	6878	0.41	7087	-0.29
Light-skinned	CAN (EN)	0.92	8951	0.45	9842	0.30	10267	-0.56
Light-skinned	CAN (FR)	1.15	1111	0.48	1147	0.39	1230	-0.64
Christian	CHN	1.28	1832	0.95	2236	0.23	2330	-0.48
Light-skinned	DEU	0.92	12637	0.53	13029	0.26	14216	-0.68
Light-skinned	ESP	1.01	3564	0.63	4180	0.32	4324	-0.65
Light-skinned	FRA	1.01	9239	0.41	9757	0.32	10461	-0.43

Light-skinned	GBR	0.90	13139	0.45	14583	0.30	15327	-0.57
Light-skinned	KOR	1.11	1460	0.97	1920	-0.35	1985	-0.45
Light-skinned	NLD	0.72	6432	0.60	7508	0.32	8104	-0.73
Light-skinned	SWE	0.71	11711	0.57	12552	0.37	13534	-0.12
Light-skinned	USA	0.91	558605	0.44	585443	0.31	315233	-0.31
Normal weight	AUS	1.16	5592	1.11	6847	0.63	6989	-0.35
Normal weight	BEL	1.06	1066	1.14	1240	0.58	1287	-0.28
Normal weight	BRA	0.88	3513	1.09	4563	0.78	4610	-0.28
Normal weight	CAN (EN)	1.08	9049	1.03	10111	0.58	10286	-0.44
Normal weight	CAN (FR)	1.46	1075	1.29	1152	0.91	1196	-0.46
Christian	CHN	0.26	3906	0.77	4798	-0.29	4913	-0.45
Normal weight	DEU	1.44	11310	0.96	12222	0.62	12781	-0.56
Normal weight	ESP	1.10	3527	1.08	4047	0.39	4158	-0.57
Normal weight	FRA	1.32	6714	1.21	7414	0.75	7621	-0.39

Normal weight	GBR	1.02	11518	1.03	13113	0.59	13375	-0.44
Normal weight	KOR	0.80	2640	1.52	3441	-0.20	3550	-0.35
Normal weight	NLD	1.08	4587	1.33	5535	0.66	5735	-0.61
Normal weight	SWE	1.30	6594	1.12	7240	0.69	7611	-0.04
Normal weight	USA	1.13	136920	1.12	139779	0.65	143990	-0.25

Note. Top row of column labels refer to dependent measures: IAT = IAT; Preference = one-item preference measure; Thermometer = feeling thermometer difference score. Second row of column labels refer to Group = advantaged group examined in analyses; dz = Cohen's d_z, with positive scores indicating outgroup favoritism and negative scores indicating ingroup favoritism; N = sample size, Conservatism = sample-average of conservatism. Abbreviations in second column denote country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

Results

Testing the predictions of SJT at the individual level.

Implicit versus explicit measures. We fitted three separate random-effects models, using the three measures of intergroup evaluation as dependent variables.

IAT D Scores. We observed a significant large mean effect of d_z = 1.00, z = 32.49, p < .001, 95% *CI* [0.94; 1.06]. This medium-sized positive effect indicates that, on average, members of advantaged groups displayed ingroup favoritism on the IAT. The estimated amount of total heterogeneity was τ^2 = 0.07, $Q(70)$ = 73,129.55, p < .001, accounting for a large proportion of the total variability. The percentage of the heterogeneity not attributable to sampling error was I^2 = 99.97%.

One-item preference scores. We also observed a large mean effect of d_z = 0.77, z = 20.58, p = < .001, 95% *CI* [0.70; 0.84], indicating that, on average, members of advantaged groups displayed ingroup favoritism on the one-item preference measures. The estimated amount of total heterogeneity was τ^2 = 0.10, $Q(70)$ = 133,992.31, p < .001, accounting for a large proportion of the total variability, and I^2 = 99.97%, indicating high total heterogeneity.

Feeling thermometer difference scores. We observed a medium-sized mean effect of d_z = 0.43, z = 12.27, p < .001, 95% *CI* [0.36; 0.50], indicating that, on average, members of advantaged groups also displayed ingroup favoritism on feeling thermometer difference scores. The estimated amount of total heterogeneity was τ^2 = 0.09, $Q(70)$ = 64,160.26, p < .001, and I^2 = 99.95%, indicating high total heterogeneity.

Taken together, advantaged groups displayed large-to-medium effects of ingroup favoritism on implicit and explicit measures.

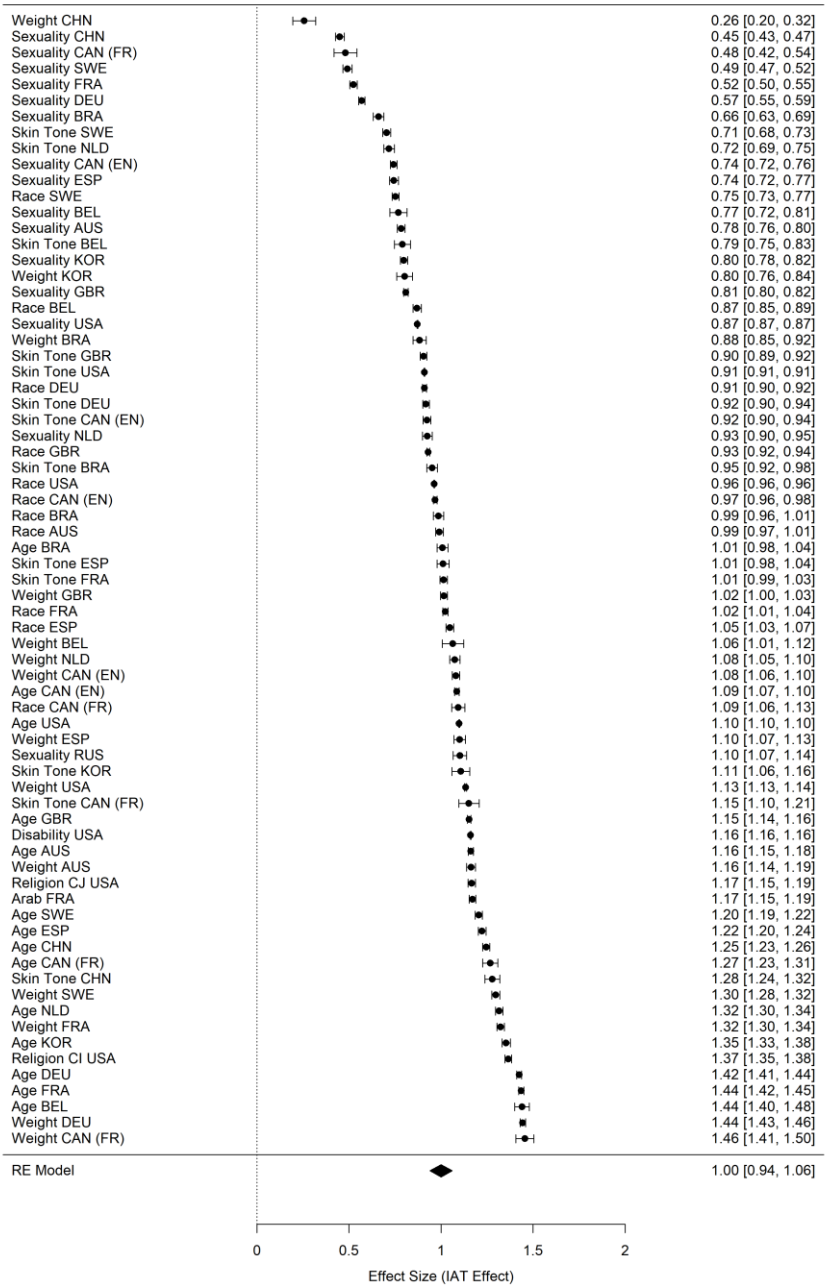

Figure A1. Caterpillar plot of random-effects meta-analysis of IAT effects (IAT *D* Scores) with study effects ordered by effect size. Positive scores indicate ingroup favoritism and negative scores indicate outgroup favoritism from the perspective of the advantaged groups. Error bars depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Dataset labels denote the intergroup domain for each study and the respective country. Abbreviations for religious groups: CH = Christian participants, CJ = Christianity vs. Judaism, CI = Christianity vs. Islam. Country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

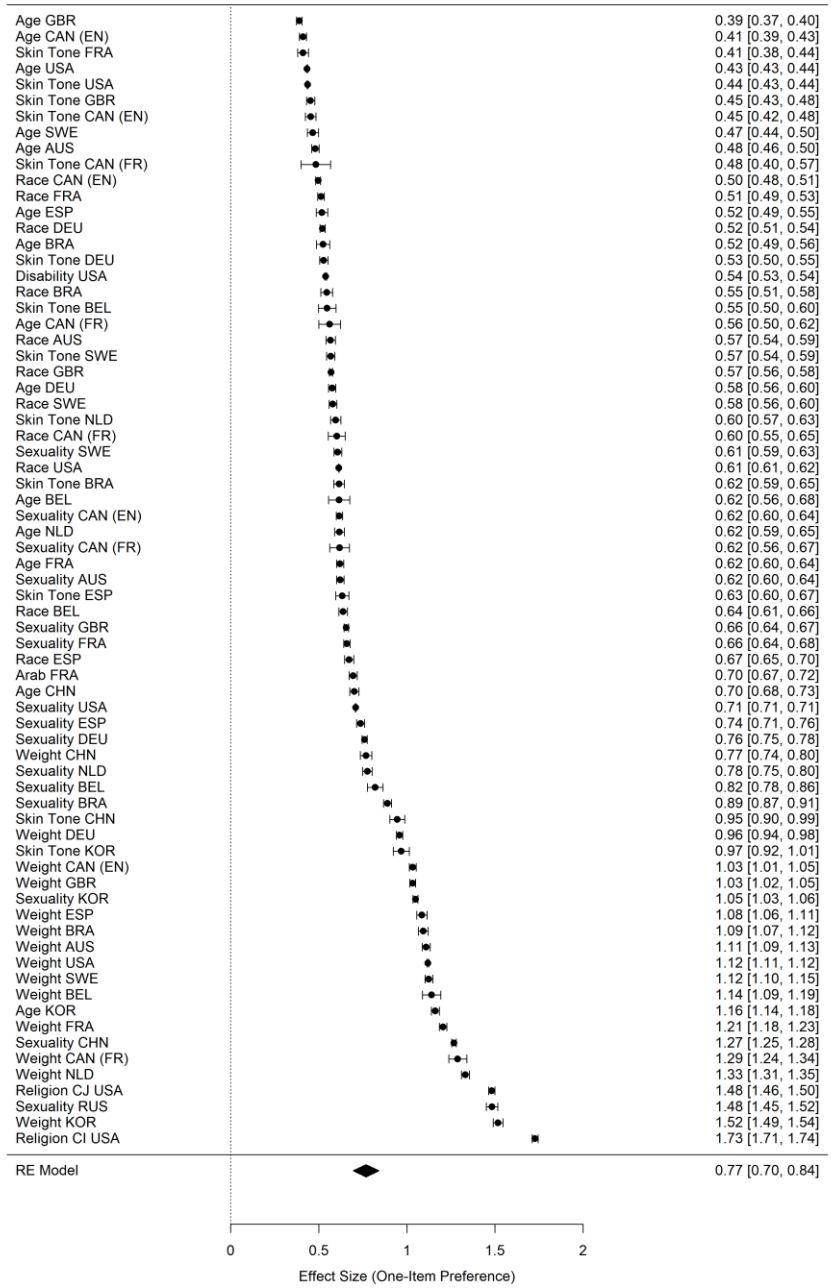

Figure A2. Caterpillar plot of random-effects meta-analysis of one-item preference scores with study effects ordered by effect size. Positive scores indicate ingroup favoritism and negative scores indicate outgroup favoritism from the perspective of the advantaged groups. Error bars depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Dataset labels denote the intergroup domain for each study and the respective country. Abbreviations for religious groups: CJ = Christianity vs. Judaism, CI = Christianity vs. Islam. Country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

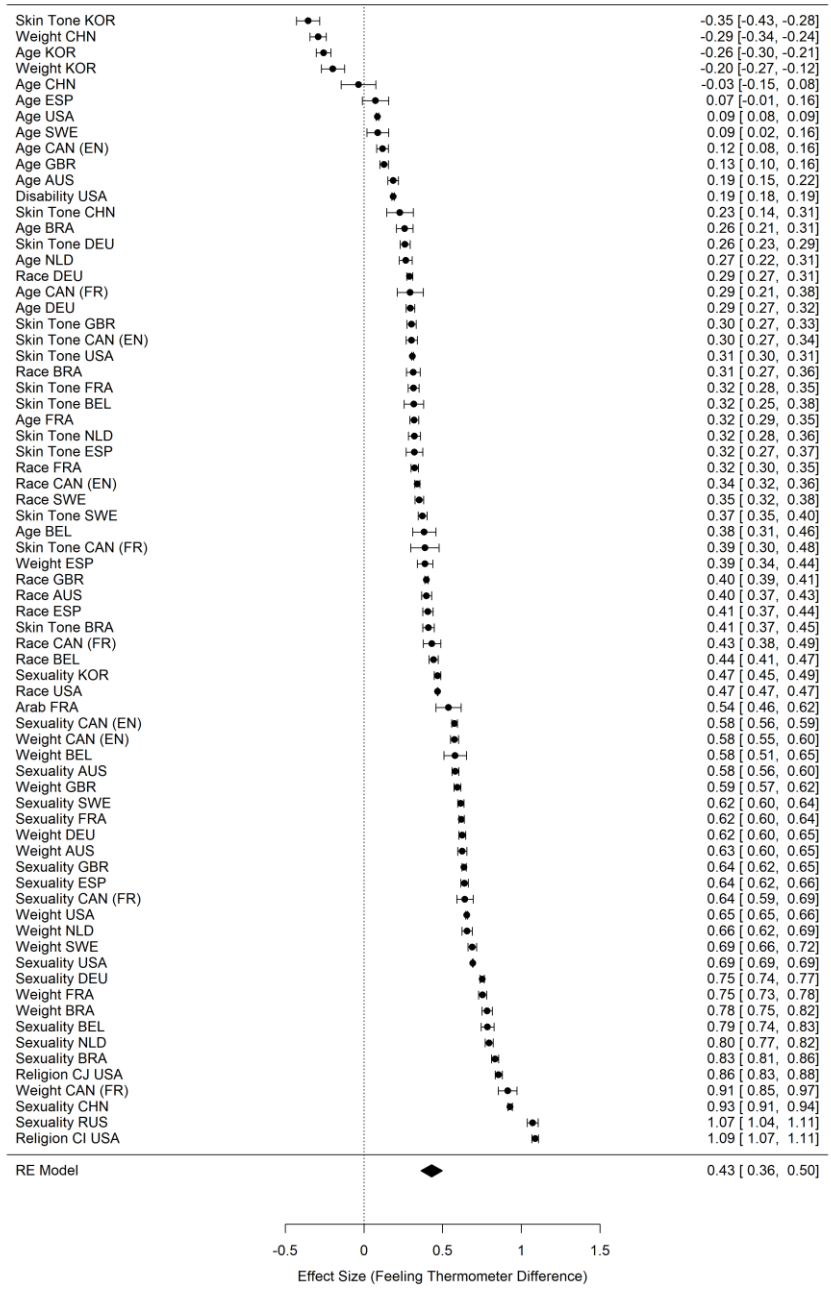

Figure A3. Caterpillar plot of random-effects meta-analysis of feeling thermometer (difference) scores with study effects ordered by effect size. Positive scores indicate ingroup favoritism and negative scores indicate outgroup favoritism from the perspective of the advantaged groups. Error bars depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Dataset labels denote the intergroup domain for each study and the respective country. Abbreviations for religious groups: CJ = Christianity vs. Judaism, CI = Christianity vs. Islam. Country codes: AUS = Australia, BEL = Belgium, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, DEU = Germany, ESP = Spain, FRA = France, GBR = United Kingdom, KOR = Korea, NLD = The Netherlands, RUS = Russia, SWE = Sweden, USA = United States.

Conservatism. To test the relationship between advantaged group members' conservative beliefs and intergroup evaluation at the individual level, we followed the same procedure as detailed in the main text.

IAT D Scores. We observed an average effect of $r = .10$, $z = 13.51$, $p < .001$, 95% *CI* [0.09; 0.12], indicating that the correlation between conservatism and IAT *D* Scores at the individual level was small. The estimated amount of total heterogeneity was $\tau^2 = 0.0039$, $Q(70) = 21{,}013.06$, $p < .001$, accounting for a large proportion of the total variability, $I^2 = 99.53\%$.

One-item preference scores. We observed an average effect of $r = .15$, $z = 9.20$, $p < .001$, 95% *CI* [0.12; 0.19], indicating that the correlation between conservatism and one-item preference scores at the individual level was small. The estimated amount of total heterogeneity was $\tau^2 = 0.02$, $Q(70) = 93{,}162.54$, $p < .001$, accounting for a large proportion of the total variability, $I^2 = 99.91\%$.

Feeling thermometer difference scores. We observed an average effect of $r = .16$, $z = 9.82$, $p < .001$, 95% *CI* [0.13; 0.19], indicating that the correlation between conservatism and feeling thermometer difference scores at the individual-level was small. The estimated amount of total heterogeneity was $\tau^2 = 0.02$, $Q(69) = 103{,}040.45$, $p < .001$, accounting for a large proportion of the total variability, $I^2 = 99.91\%$.

Taken together, we observed positive but small correlations between conservatism and intergroup evaluation: Members of advantaged groups displayed more favorable evaluations towards their ingroup relative to an disadvantaged outgroup the more they described themselves as conservative.

Examining relationships between conservatism and intergroup evaluation at the group level.

To test the relationship between advantaged groups' conservative beliefs and intergroup evaluation at the group level, we fitted three separate mixed-effects meta-regression models with self-reported conservatism aggregated at the sample level as continuous moderators, and using advantaged groups' IAT *D* Scores, one-item preference scores, and feeling thermometer difference scores as dependent variables.

IAT D Scores. First, we fitted a mixed-effects model with group aggregates of self-reported conservatism as a continuous moderator, using advantaged groups' IAT *D* Scores as dependent variable. We did not observe a significant moderating effect, $Q_M(1) = 0.66$, $p = .417$, $R^2 = 0.00\%$. This indicates that group aggregates of self-reported conservatism were unrelated to group aggregates of implicit ingroup favoritism (see Figure C4, Panel A).

One-item preference scores. Again, we fitted a mixed-effects model with group aggregates of self-reported conservatism as a continuous moderator, now using advantaged groups' one-item preference scores as dependent variable. We did not observe a significant moderating effect, $Q_M(1) = 2.55$, $p = .110$, $R^2 = 2.16\%$. This indicates that group aggregates of self-reported conservatism were unrelated to group aggregates of one-item preference scores (see Figure C4, Panel B).

Feeling thermometer difference scores. Lastly, we fitted the same mixed-effects model with feeling thermometer difference scores as the dependent variable and self-reported conservatism as a continuous moderator. Here, we also did not observe a significant moderating effect, $Q_M(1) = 1.87$, $p = .171$, with conservatism accounting for only 1.23% of the heterogeneity

in feeling thermometer difference scores. This indicates that that group aggregates of self-reported conservatism were unrelated to group aggregates of feeling thermometer difference scores (see Figure C4, Panel C).

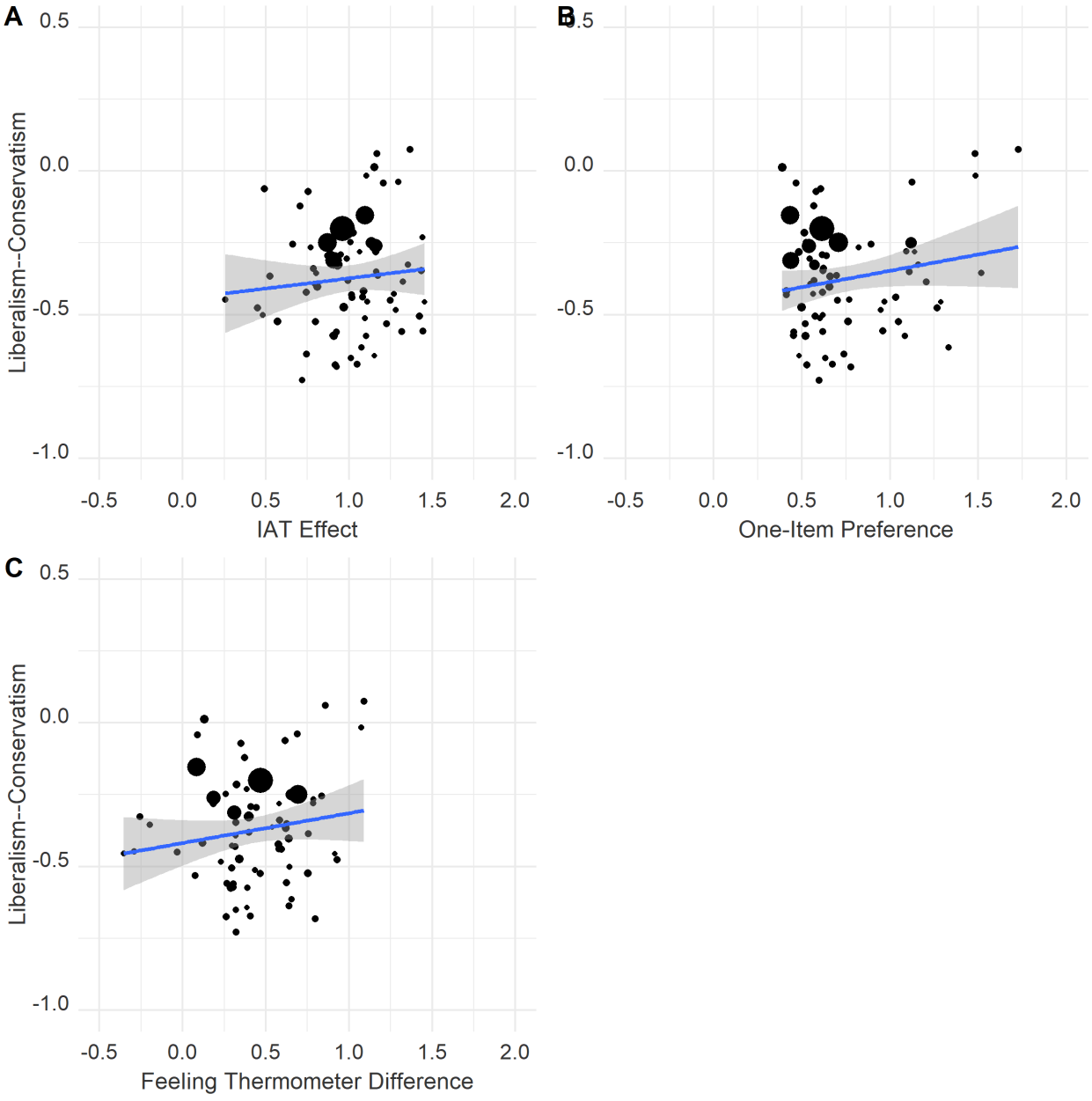

Figure A4. Scatterplot depicting the relationship between conservatism and intergroup evaluations among advantaged groups. Y-axes reflect disadvantaged groups' sample-level mean Cohen's d_z for conservatism, measured on a one-item 6- point and/or 7-point scale. X-axes reflect mean Cohen's d_z for IAT *D* Scores (Panel A), one-item preference scores (Panel B), and feeling thermometer difference scores (Panel C). Each circle corresponds to a different social group, with circle size reflecting sample size. Positive values on the y-axes indicate more conservative attitudes. Positive values on the x-axes indicate more favorable evaluations of the advantaged group relative to the disadvantaged group (i.e., ingroup favoritism for these samples).

Discussion

Taken together, we examined the relationships between conservatism and intergroup evaluation in advantaged groups at both the individual and group levels in a series of additional moderator analyses. At the individual level, we observed small correlations for all three intergroup bias metrics, consistent with the hypothesis that members of advantaged groups display more ingroup favoritism the more they endorse conservative beliefs (Jost, Banaji, & Nosek, 2004). However, these effects were not replicated at the group level. Specifically, we did not observe significant moderation effect of conservatism for IAT *D* Scores, one-item preference scores, or feeling thermometer difference scores. In short, we found that the relationship between conservatism and intergroup evaluations in advantaged groups depended on the level of analysis, with small, but reliable, correlations at the individual level, but non-significant correlations at the group level.

References

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Kite, M. E., Stockdale, G. D., Whitley, B. E., & Johnson, B. T. (2005). Attitudes toward younger and older adults: An updated meta-analytic review. *Journal of Social Issues*, *61*(2), 241–266. https://doi.org/10.1111/j.1540-4560.2005.00404.x

Neugarten, B. L. (1974). Age groups in american society and the rise of the young-old. *The ANNALS of the American Academy of Political and Social Science*, *415*(1), 187–198. https://doi.org/10.1177/000271627441500114

Additional Figures

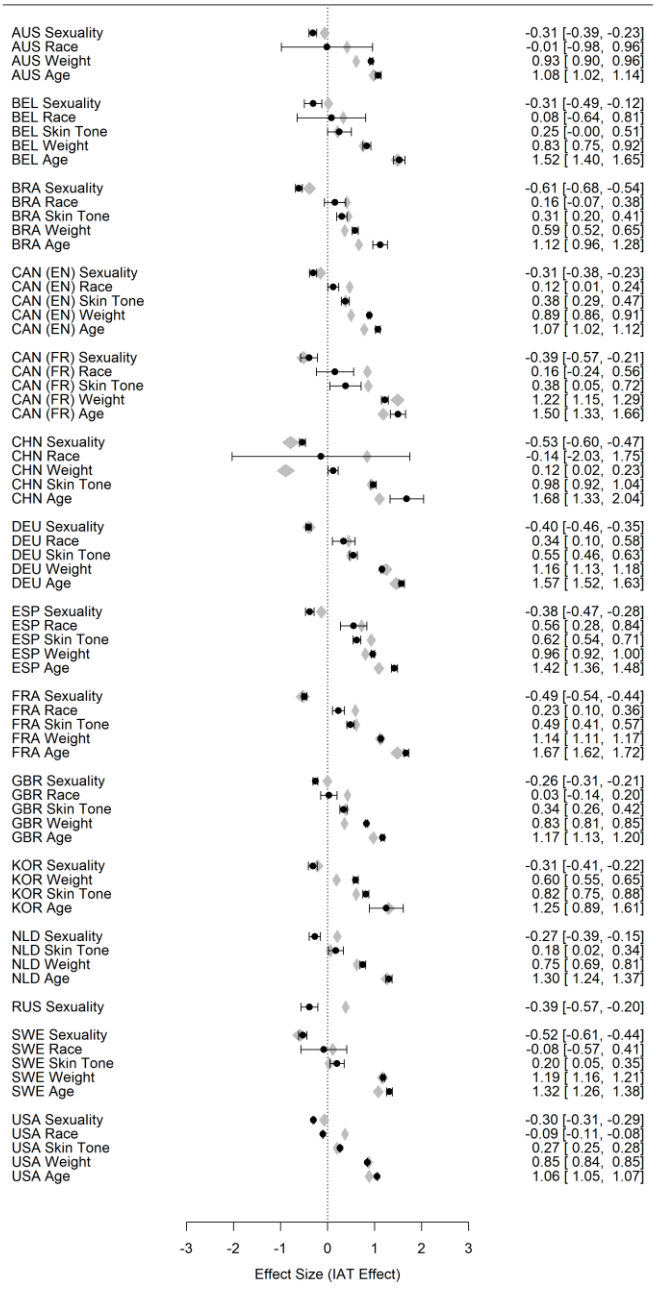

Figure A5. Forest plot of random-effects meta-analysis of disadvantaged groups' IAT *D* Scores (black circles) and IAT *D* Score stigma as continuous moderator (gray polygons). Study effects are ordered by country *and* effect size of IAT *D* Scores. Positive scores of IAT *D* Scores indicate outgroup favoritism and negative scores indicate ingroup favoritism from the perspective of the disadvantaged groups. Positive scores of IAT *D* Score stigma indicate a preference for advantaged relative to disadvatanged groups; negative scores indicate a preference for disadvantaged groups relative to advantaged groups. Error bars for IAT *D* Scores depict

95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Country codes: AUS = Australia, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, FRA = France, DEU = Germany, KOR = Korea, NLD = The Netherlands, RUS = Russia, SPA = Spain, SWE = Sweden, GBR = United Kingdom, USA = United States.

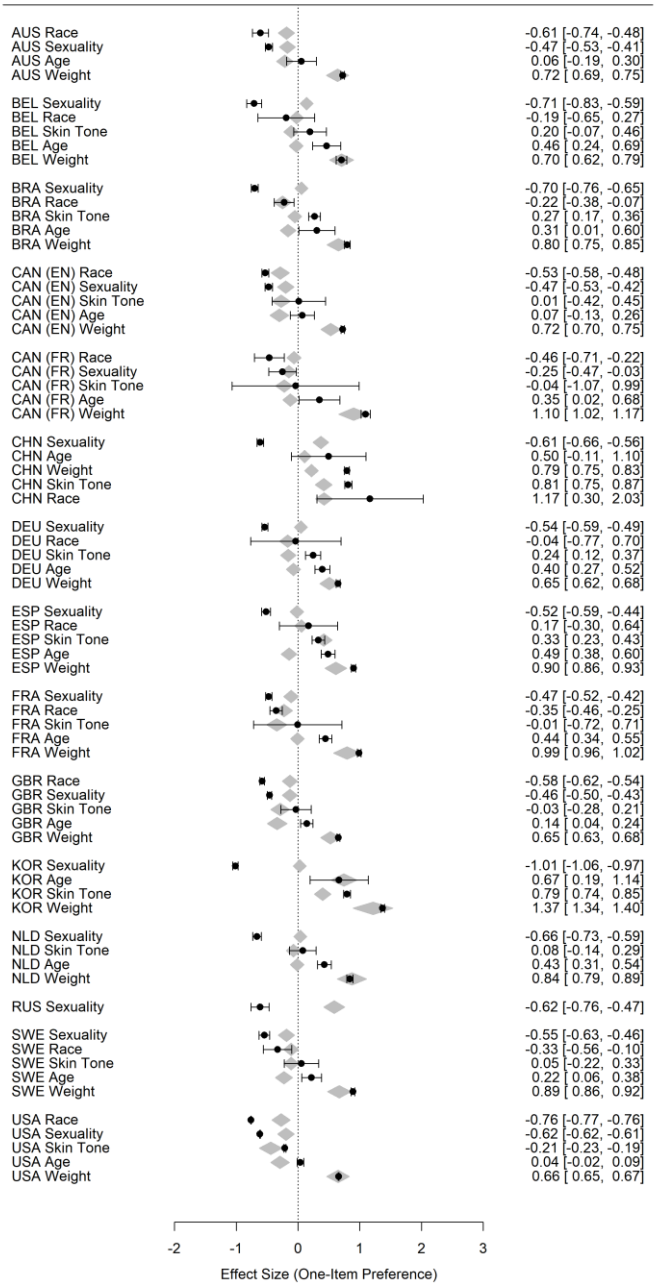

Figure A6. Forest plot of random-effects meta-analysis of disadvantaged groups' one-item preference scores (black circles) and one-item preference score stigma as continuous moderator (gray polygons). Study effects are ordered by country *and* effect size of IAT *D* Scores. Positive scores of one-item preference scores indicate outgroup favoritism and negative scores indicate ingroup favoritism from the perspective of the disadvantaged groups. Positive scores of one-item preference score stigma indicate a preference for advantaged relative to disadvatanged groups; negative scores indicate a preference for disadvantaged groups relative to advantaged groups. Error bars for one-item preference scores depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Country codes: AUS = Australia, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, FRA = France, DEU = Germany, KOR = Korea, NLD = The Netherlands, RUS = Russia, SPA = Spain, SWE = Sweden, GBR = United Kingdom, USA = United States.

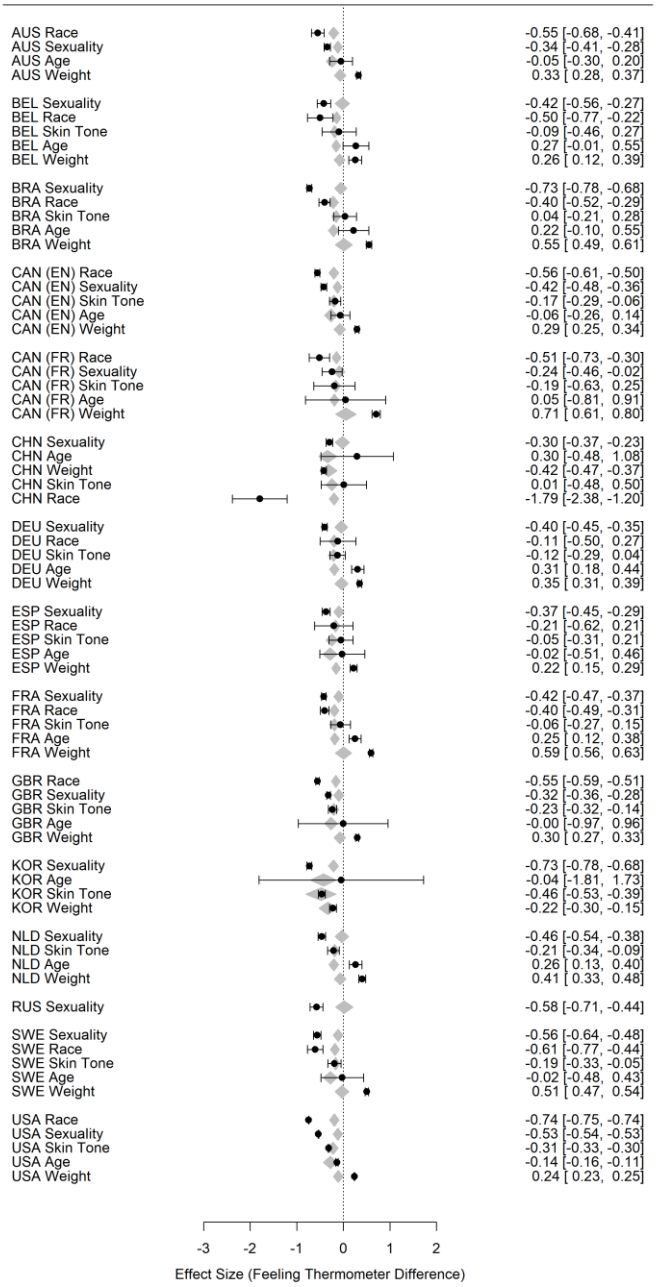

Figure A7. Forest plot of random-effects meta-analysis of disadvantaged groups' feeling thermometer difference scores (black circles) and feeling thermometer difference score stigma as continuous moderator (gray polygons). Study effects are ordered by country *and* effect size of IAT *D* Scores. Positive scores of feeling thermometer difference scores indicate outgroup favoritism and negative scores indicate ingroup favoritism from the perspective of the disadvantaged groups. Positive scores of feeling thermometer difference score stigma indicate a preference for advantaged relative to disadvatanged groups; negative scores indicate a preference for disadvantaged groups relative to advantaged groups. Error bars for feeling thermometer difference scores depict 95% confidence intervals and values in squared brackets indicate lower and upper bounds of confidence intervals. Country codes: AUS = Australia, BRA = Brazil, CAN (EN) = Canada (English), CAN (FR) = Canada (French), CHN = China, FRA = France, DEU = Germany,

KOR = Korea, NLD = The Netherlands, RUS = Russia, SPA = Spain, SWE = Sweden, GBR = United Kingdom, USA = United States.

APPENDIX B

Supplemental Tables (Study 1 and 2A)

Table B1

Zero-order correlations of Turkish-German participants' IAT D scores, feeling thermometer difference scores, overlap of self, ingroup, and outgroup (OSIO) subscales regarding Turks and Germans, collective self-esteem, and their belief in a just world in Study 1.

	1	2	3	4	5	6	7	8	9	10	*M*	*SD*
1. IAT	-										0.20	0.35
2. Turks	.27*	-									21.47	27.78
3. Germans	.06	.51***	-								16.53	28.11
4. SGT	.33*	.10	-.28*	-							3.55	1.04
5. GST	.30*	.18	-.18	.70***	-						3.84	1.24
6. SGG	-.12	.11	-.01	.27*	.12	-					3.29	0.96
7. GSG	-.09	.16	.19	.16	.02	.64***	-				3.26	1.23
8. CSE Private	.04	.11	.05	.03	-.15	.00	.04	-			3.00	0.93
9. CSE Public	.14	.36**	.11	.48***	.34*	.18	.16	.21	-		3.62	0.80

10. CSE Identity	.18	.18	-.20	.27*	.40**	-.02	.17	.22	.31*	-	3.77	0.95
11. BJW	.00	-.01	-.07	.05	.08	.11	.14	.15	-.11	.27*	2.65	0.98

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; IAT = IAT *D* Score; Turks = feeling thermometer score regarding Turks; Germans = feeling thermometer score regarding Germans; SGT = self in group Turkish; GST = group in self Turkish; SGG = self in group German; GSG = group in self German; CSE = collective self-esteem: Private = private collective self-esteem subscale; Public = public collective self-esteem subscale; Identity = importance to identity subscale; BJW = belief in a just world. IAT *D* Scores are coded such that more positive scores indicate more favorable evaluations of Arab and Muslim individuals relative to Whites; feeling thermometer scores are coded such that more positive values indicate more positive evaluations regarding the respective target category.

Table B2

Results of a two (Task Order: Turkish+positive first vs. German+positive first) by two (IAT Version: standard vs. personalized) ANOVA with IAT D Score as dependent variable.

Factor	df	F	p	eta
order	1,49	3.78	0.058	0.07
version	1,49	0.56	0.459	0.01
order:version	1,49	3.27	0.077	0.06

Table B3

Zero-order correlations of Muslim participants' AMP scores, self-reported liking, feeling thermometer scores regarding Muslims and Germans, overlap of self, ingroup, and outgroup (OSIO) subscales regarding Muslims and Germans, and their belief in a just world in Study 2.

	1	2	3	4	5	6	7	8	*M*	*SD*
1. AMP	-								0.58	1.70
2. Liking	.22	-							0.69	1.11
3. Muslims	.06	.05	-						34.80	20.10
4. Germans	.12	-.46**	.32*	-					14.80	21.91
5. SGM	.13	.04	.24	-.01	-				5.65	1.58
6. GSM	.12	.03	.04	.08	.55***	-			6.24	1.32
7. SGG	.01	-.29*	.22	.49***	.18	.15	-		4.24	1.51
8. GSG	.07	-.33*	.25	.26	-.01	-.14	.63***	-	3.07	1.67
9. BJW	-.15	-.14	.02	.02	-.10	-.21	.16	.04	2.97	0.72

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; AMP = AMP score; Liking = self-reported liking score for prime images; Muslims = feeling thermometer score regarding Muslims; Germans = feeling thermometer score regarding Germans; SGM = self in group Muslim; GSM = group in self Muslim; SGG = self in group German; GSG = group in self German; BJW = belief in a just world. AMP scores and self-reported liking scores are coded such that more positive scores indicate more favorable evaluations of Arab and Muslim individuals relative to Whites; feeling

thermometer scores are coded such that more positive values indicate more positive evaluations regarding the respective target category.

Stereotypicality Ratings of Primes (Study 2A)

As a manipulation check, we examined whether Arab exemplars with Muslim headgear were rated as more stereotypically Muslim than other exemplars, and whether Muslim and non-Muslim participants differed in how they rated the stereotypicality of exemplars. We submitted participants' mean exemplar stereotypicality scores to a 2 (Exemplar Ethnicity: Arab vs. White) by 2 (Exemplar Headgear: yes vs. no) by 2 (Participant Group: Muslim vs. non-Muslim) ANOVA with repeated measures on the first two factors. We observed a significant main effect of Exemplar Ethnicity, $F(1,108) = 987.93$, $p < .001$, $\eta_p^2 = .90$, 95% *CI* [0.87; 0.92], indicating that Arab exemplars ($M = 6.18$, $SD = 1.53$) were rated as more stereotypically Muslim than White exemplars ($M = 1.81$, $SD = 0.80$). However, this effect was qualified by a significant two-way interaction of Exemplar Ethnicity and Exemplar Headgear, $F(1,108) = 100.41$, $p < .001$, $\eta_p^2 = .48$, 95% *CI* [0.37; 0.57], and a significant two-way interaction of Exemplar Ethnicity and Participant Group, $F(1,108) = 8.99$, $p = .003$, $\eta_p^2 = .08$, 95% *CI* [0.02; 0.17].[52]

Post hoc comparisons for the two-way interaction of Exemplar Ethnicity and Exemplar Headgear revealed that Arab exemplars with Muslim headgear ($M = 6.74$, $SD = 1.33$) were rated as significantly more stereotypically Muslim than Arab exemplars without headgear ($M = 5.62$,

[52] We also observed a significant main effect of headgear, $F(1,108) = 104.93$, $p < .001$, $\eta_p^2 = .49$, 95% *CI* [0.38; 0.58], which, however, was not of particular interest for the present research.

SD = 1.51), $t(216) = -14.32$, $p < .001$, $d_z = -1.37$, 95% *CI* [-1.59; -1.14][53], than White exemplars without headgear (*M* = 1.82, *SD* = 0.82), $t(143) = 32.97$, $p < .001$, $d_z = 3.14$, 95% *CI* [2.73; 3.55], and than White exemplars with headgear (*M* = 1.81, *SD* = 0.78), $t(145) = 32.85$, $p < .001$, $d_z = 3.13$, 95% *CI* [2.73; 3.54].

Post hoc comparisons for the two-way interaction of Exemplar Ethnicity and Participant Group revealed that Muslim participants rated Arab exemplars (*M* = 5.98, *SD* = 1.65) as significantly more stereotypically Muslim than White exemplars (*M* = 2.10, *SD* = 1.01), $t(108) = 18.64$, $p < .001$, $d_z = 2.75$, 95% *CI* [2.28; 3.21]. Similarly, non-Muslim participants rated Arab exemplars (*M* = 6.32, *SD* = 1.42) as significantly more stereotypically Muslim than White exemplars (*M* = 1.61, *SD* = 0.52), $t(108) = 26.62$, $p < .001$, $d_z = 3.33$, 95% *CI* [2.82; 3.83]. Interestingly, Muslim participants and non-Muslim participants did not significantly differ in the extent to which they rated Arab exemplars as stereotypically Muslim, $t(213) = -1.63$, $p = .627$, $d_s = -0.32$, 95% *CI* [-0.70; 0.07]. Furthermore, Muslim participants and non-Muslim participants also did not significantly differ in the extent to which they rated White exemplars in terms of stereotypicality, $t(213) = 2.36$, $p = .115$, $d_s = 0.46$, 95% *CI* [0.07; 0.84].

---

[53] *CIs* for Cohens's d_z were calculated using code provided by Lakens (2015; http://daniellakens.blogspot.com/2015/05/the-perfect-t-test.html)

Instructions for ICI Ratings (Study 2B)

Online participants read the following instruction:

This is a study on person perception and first impressions. Sometimes people need to make judgments about others from far distances or at night, when visual information is not completely clear. On the following pages you will be presented with a series of blurred faces. As a result of the the blurring process these faces might sometimes look similar to each other, but they are in fact different individuals. Your task is to rate the faces according to how much they seem to possess a given trait. Please do not think too long before making a judgment. Because we are especially interested in your spontaneous reactions, we encourage you to rely on your intuition and your gut feeling while making your judgments.